Investor Sentiment in Blogs

Design of a Classifier and Validation by a Portfolio Simulation

Dissertation

in partial fulfillment of the requirements for the degree of
Doctor of Economics (Dr. oec.)

submitted to the
Faculty of Business, Economics and Social Sciences
of the University of Hohenheim

by

Achim Klein

May 2016

Dean: Prof. Dr. Dirk Hachmeister

Head of the Examination Committee: Prof. Dr. Hans-Peter Burghof
Supervisor and Primary Reviewer of the Doctoral Thesis: Prof. Dr. Stefan Kirn
Secondary Reviewer of the Doctoral Thesis: Prof. Dr. Robert Jung

Dates of the Oral Examinations ("Rigorosum"):

1. Oral Examination on Banking and Finance
   ("Bankwirtschaft und Finanzdienstleistungen / ABWL"): 2016-04-11
   Examiner: Prof. Dr. Hans-Peter Burghof, Committee Member: Prof. Dr. Stefan Kirn
2. Oral Examination on Econometrics ("Ökonometrie / AVWL"): 2016-04-28
   Examiner: Prof. Dr. Robert Jung, Committee Member: Prof. Dr. Stefan Kirn
3. Oral Examination on Information Systems ("Wirtschaftsinformatik"): 2016-05-13
   Examiner: Prof. Dr. Stefan Kirn, Committee Member: Prof. Dr. Hans-Peter Burghof

Acknowledgements

First, I thank the job interviewer who asked me along the lines: “Well, you like to acquire knowledge, don’t you want to write a dissertation then?”. Second, I thank Prof. Dr. Hans-Peter Burghof for attracting me to the field of finance and the University of Hohenheim with its nice botanical gardens.

I thank Prof. Dr. Stefan Kirn for making this thesis possible at the chair of information systems and for supervising, supporting, and reviewing my thesis. I am especially grateful for the large degrees of freedom I experienced throughout the years of my dissertation phase. This freedom allowed me to dive into many interesting research topics related to investor sentiment analysis, text mining, machine learning, econometrics, and investor decision support. Thus, I am grateful for having learned a lot on this journey. During this journey, I enjoyed helpful discussions with my colleagues. I would like to thank Dr. Jörg Leukel and Dr. Marcus Müller for their scientific advice. I also thank my team of the FIRST project, especially Martin Riekert, Olena Altuntas, Velizar Dinev, Lyubomir Kirilov, Lilli Gredel, and Beka Namchevadze. I had initiated the FIRST project (co-funded by the EU in framework programme 7 under contract number 257928, 2010-2013), which served as background and helped to drive my dissertation project. I am grateful for many fruitful discussions with project partners, especially with Prof. Dr. Jan Muntermann, who adviced me to keep things simple.

Finally, I would like to thank my family, Korinna, and Joshua Leonhard, for supporting my dissertation project.

Overview

1 Introduction **1**
1.1 Motivation 1
1.2 Research Approach 3
1.3 Thesis Structure 5

2 Analysis of the State of the Art **7**
2.1 Asset Pricing Theory 7
2.2 Behavioral Finance Theory 17
2.3 Blogs as a Source of Investor Sentiment 27
2.4 Approaches for Textual Investor Sentiment Classification 39
2.5 Effects of Textual Investor Sentiment on Returns 61
2.6 Summary of the Research Gap 83

3 Design of a Classifier for Investor Sentiment in Blogs **89**
3.1 Corpus 89
3.2 Classifier 97
3.3 Evaluation 104

4 Validation by a Portfolio Simulation **111**
4.1 Datasets 111
4.2 Design and Hypotheses 125
4.3 Results 143
4.4 Summary of Hypotheses Testing 188

5 Conclusions **193**
5.1 Research Contributions 193
5.2 Practical Implications 195
5.3 Limitations 196

Bibliography **197**

A Appendix **213**
A.1 Sample Blog Documents 213
A.2 Corpus of Blog Documents 215
A.3 Investor Sentiment Classifier Implementation 226
A.4 Portfolio Simulation Dataset Retrieval 227
A.5 Market Data 232

Contents

List of Tables ..... **ix**

List of Figures ..... **xi**

List of Abbreviations ..... **xiii**

1 Introduction **1**
1.1 Motivation 1
1.2 Research Approach 3
1.3 Thesis Structure 5

2 Analysis of the State of the Art **7**
2.1 Asset Pricing Theory 7
2.1.1 Inefficient Markets and Abnormal Returns 7
2.1.2 Models of Normal and Abnormal Returns 9
2.1.2.1 Capital Asset Pricing Model 9
2.1.2.2 Fama-French Model 11
2.1.2.3 Carhart Model 12
2.1.3 Estimating Unexpected Abnormal Returns 13
2.1.4 Example for Abnormal Return Existence 15
2.2 Behavioral Finance Theory 17
2.2.1 Psychological Biases Affecting Asset Prices 17
2.2.2 Investor Sentiment in Behavioral Finance 20
2.2.3 Measures of Investor Sentiment 24
2.3 Blogs as a Source of Investor Sentiment 27
2.3.1 Blog Characteristics 27
2.3.2 Blog Platforms 31
2.3.3 Examples of Blog Documents 34
2.3.4 Model of Investor Sentiment in Blog Documents 36
2.4 Approaches for Textual Investor Sentiment Classification 39
2.4.1 Dictionary-based Approaches 39
2.4.2 Knowledge-based Approaches 41
2.4.3 Supervised Machine Learning Approaches 45
2.4.3.1 Vector Space Model 45
2.4.3.2 Supervised Machine Learning 47
2.4.3.3 Naive Bayes 52
2.4.3.4 Support Vector Machines 54
2.4.3.5 Assessment with Respect to this Work 60
2.5 Effects of Textual Investor Sentiment on Returns 61
2.5.1 Effects of Investor Sentiment from News 62
2.5.2 Effects of Investor Sentiment from Stock Message Boards 67
2.5.3 Effects of Investor Sentiment from Twitter 70
2.5.4 Effects of Investor Sentiment from Blogs 72
2.6 Summary of the Research Gap 83

3 Design of a Classifier for Investor Sentiment in Blogs **89**
3.1 Corpus 89
3.1.1 Sampling and Annotation Approaches 89
3.1.2 Corpus Description and Analysis 92
3.1.3 Evaluation 96
3.2 Classifier 97

3.2.1 Document-Vector Transformation ... 97
3.2.1.1 Feature Definition ... 98
3.2.1.2 Feature Selection and Feature Extraction ... 99
3.2.1.3 Feature Weighting ... 100
3.2.2 Support Vector Machine Configuration ... 103
3.3 Evaluation ... 104
3.3.1 Cross-Validation Approach ... 104
3.3.2 Experiments and Results ... 105

4 Validation by a Portfolio Simulation ... 111
4.1 Datasets ... 111
4.1.1 Data Acquisition ... 111
4.1.2 Description and Analysis ... 112
4.1.2.1 Seekingalpha ... 112
4.1.2.2 Blogspot ... 119
4.2 Design and Hypotheses ... 125
4.2.1 Portfolio Simulation Design ... 126
4.2.2 Hypotheses ... 130
4.2.3 Testing of the Hypotheses ... 133
4.2.3.1 Alpha-based Tests ... 133
4.2.3.2 Benchmark-based Tests ... 141
4.3 Results ... 143
4.3.1 Hypothesis H1: Effects on a Long & Short Portfolio ... 144
4.3.1.1 Seekingalpha ... 144
4.3.1.2 Blogspot ... 148
4.3.2 Hypothesis H2: Effects on a Long Portfolio ... 151
4.3.2.1 Seekingalpha ... 151
4.3.2.2 Blogspot ... 153
4.3.3 Hypothesis H3: Effects on a Short Portfolio ... 155
4.3.3.1 Seekingalpha ... 155
4.3.3.2 Blogspot ... 158
4.3.4 Hypothesis H4: Effects in Relation to Portfolio Sizes ... 160
4.3.4.1 Seekingalpha ... 160
4.3.4.2 Blogspot ... 166
4.3.5 Hypothesis H5: Effects in Relation to the Datasets ... 171
4.3.5.1 With Transaction Costs ... 171
4.3.5.2 Without Transaction Costs ... 175
4.3.6 Hypothesis H6: Effects in Relation to the Momentum Effect ... 177
4.3.6.1 Existence of Abnormal Momentum Returns ... 178
4.3.6.2 Comparing Sentiment Portfolios to Momentum Portfolios ... 179
4.3.6.2.1 With Transaction Costs ... 179
4.3.6.2.2 Without Transaction Costs ... 185
4.4 Summary of Hypotheses Testing ... 188

5 Conclusions ... 193
5.1 Research Contributions ... 193
5.2 Practical Implications ... 195
5.3 Limitations ... 196

Bibliography ... 197

A Appendix ... 213
A.1 Sample Blog Documents ... 213

A.2 Corpus of Blog Documents 215
A.3 Investor Sentiment Classifier Implementation 226
A.4 Portfolio Simulation Dataset Retrieval 227
A.5 Market Data 232

List of Tables

Table 1: Overview of approaches to study effects of investor sentiment on stock returns .. 84
Table 2: Overview and descriptive analysis of the Seekingalpha blog document corpus 94
Table 3: Time period analysis of the blog documents' publication dates in Corpus A 95
Table 4: Body length analysis of blog documents in Corpus A 96
Table 5: Evaluation of annotations of stock-specific investor sentiment 97
Table 6: Summary of parameter configurations for the classifier 97
Table 7: Word n-grams for an example sentence 98
Table 8: Effects of documents annotated with one vs. multiple sentiment orientations 106
Table 9: Accuracy of the classifier using a specific text representation..... 108
Table 10: Accuracy of the classifier using a specific *C*-parameter 109
Table 11: Number of investor sentiments in the Seekingalpha dataset 113
Table 12: Mean investor sentiment index in the Seekingalpha dataset 115
Table 13: Standard deviations of monthly investor sentiment index (Seekingalpha) 117
Table 14: Number of investor sentiments in the Blogspot dataset 119
Table 15: Mean investor sentiment index in the Blogspot dataset 122
Table 16: Standard deviations of monthly investor sentiment index (Blogspot) 123
Table 17: Portfolio simulation design overview..... 130
Table 18: Overview of hypotheses and sub-hypotheses..... 133
Table 19: H1.1 *t*-test results for Seekingalpha regarding the long-short portfolio 145
Table 20: Excess returns and performance measures (Seekingalpha long-short portfolio) 147
Table 21: Test results for positive excess returns (Seekingalpha long-short portfolio) 148
Table 22: H1.1 *t*-test results for Blogspot regarding the long-short portfolio 149
Table 23: Excess returns and performance measures (Blogspot long-short portfolio) 150
Table 24: Test results for positive excess returns (Blogspot long-short portfolio) 150
Table 25: H2.1 *t*-test results for Seekingalpha regarding the long portfolio 151
Table 26: Excess returns and performance measures (Seekingalpha long portfolio)..... 152
Table 27: Test results for positive excess returns (Seekingalpha long portfolio)..... 153
Table 28: H2.1 *t*-test results for Blogspot regarding the long portfolio. 153
Table 29: Excess returns and performance measures (Blogspot long portfolio)..... 154
Table 30: Test results for positive excess returns (Blogspot long portfolio)..... 155
Table 31: H3.1 *t*-test results for Seekingalpha regarding the short portfolio 156
Table 32: Excess returns and performance measures (Seekingalpha short portfolio)..... 157
Table 33: Test results for positive excess returns (Seekingalpha short portfolio)..... 157
Table 34: H3.1 *t*-test results for Blogspot regarding the short portfolio 158
Table 35: Excess returns and performance measures (Blogspot short portfolio)..... 159
Table 36: Test results for positive excess returns (Blogspot short portfolio)..... 160
Table 37: H1.1 *t*-test results for the Seekingalpha long-short portfolio with *N* stocks 161
Table 38: H1.2 test results for the Seekingalpha long-short portfolio with *N* stocks..... 162
Table 39: H4.1 t-test results for the Seekingalpha long-short portfolio 165
Table 40: H4.2 test results for the Seekingalpha long-short portfolio..... 166
Table 41: H1.1 *t*-test results for the Blogspot long-short portfolio with *N* stocks..... 167
Table 42: H1.2 test results for the Blogspot long-short portfolio with *N* stocks 168
Table 43: H4.1 t-test results for the Blogspot long-short portfolio 169
Table 44: H4.2 test results for the Blogspot long-short portfolio..... 171
Table 45: H5.1 t-test results regarding Seekingalpha vs. Blogspot long-short portfolios.. 173
Table 46: H5.2 test results regarding Seekingalpha vs. Blogspot long-short portfolios 174
Table 47: Test results for positive alpha of the momentum long-short portfolio 178
Table 48: Test results for positive median excess returns of the momentum portfolio..... 179
Table 49: H6.1 t-test results for Seekingalpha vs. momentum long-short portfolios..... 181

Table 50: H6.1 t-test results for Blogspot vs. momentum long-short portfolios 181
Table 51: Test results for higher mean excess returns of Blogspot vs. momentum 183
Table 52: H6.2 test results for Seekingalpha vs. momentum long-short portfolios 184
Table 53: H6.2 test results for Blogspot vs. momentum long-short portfolios 184
Table 54: Overview of test results for H1, H2, and H3 without transaction costs 189
Table 55: Overview of test results for H1, H2, and H3 with transaction costs 189
Table 56: Overview of test results for H4, H5, and H6 190
Table 57: Seekingalpha blog documents annotated with one sentiment orientation 215
Table 58: Seekingalpha blog documents annotated with multiple sentiment orientations. 224
Table 59: Stocks referenced by investor sentiment annotations in the entire corpus 225
Table 60: Natural language processing resources used 226
Table 61: DJIA stocks and terms used to search for and retrieve blog documents 228

List of Figures

Figure 1: Apple stock price time series and related rumors published in blogs 2
Figure 2: Research design 4
Figure 3: Built-up of a mispricing 8
Figure 4: Daily cumulative unexpected abnormal returns for the eToys stock 16
Figure 5: Elements of an example investment blog and blog document 28
Figure 6: The Seekingalpha investment blog platform website 33
Figure 7: Excerpt of an investment blog document based on fundamental analysis 34
Figure 8: Excerpt of an investment blog document based on technical analysis 35
Figure 9: Supervised machine learning of an investor sentiment classifier 48
Figure 10: Training and test error as a function of classifier complexity 50
Figure 11: An example of an optimal hyperplane 55
Figure 12: Mean cumulative unexpected abnormal return for events of long recommendations in Seekingalpha blog documents 75
Figure 13: Mean cumulative unexpected abnormal return for events of short recommendations in Seekingalpha blog documents 76
Figure 14: Box plots of Seekingalpha investor sentiment indexes 114
Figure 15: Seekingalpha investor sentiment index market vs. cumulative DJIA returns 118
Figure 16: Scatter plot of DJIA log returns and Seekingalpha investor sentiment index market 118
Figure 17: Box plots of Blogspot investor sentiment indexes 121
Figure 18: Blogspot investor sentiment index market vs. cumulative DJIA returns 124
Figure 19: Scatter plot of DJIA log returns and Blogspot investor sentiment index market 125
Figure 20: Probability density function of the *t*-distribution 139
Figure 21: Plot of the returns of the Seekingalpha long-short portfolio 146
Figure 22: Plot of the returns of the Blogspot long-short portfolio 149
Figure 23: Plot of the returns of the Seekingalpha long portfolio 152
Figure 24: Plot of the returns of the Blogspot long portfolio 154
Figure 25: Plot of the returns of the Seekingalpha short portfolio 156
Figure 26: Plot of the returns of the Blogspot short portfolio 159
Figure 27: Plot of alpha and mean excess return for the Seekingalpha long-short portfolio, ignoring transaction costs 163
Figure 28: Plot of alpha and mean excess return for the Seekingalpha long-short portfolio, adjusting for transaction costs 164
Figure 29: Plot of alpha and mean excess return for the Blogspot long-short portfolio, ignoring transaction costs 169
Figure 30: Plot of alpha and mean excess return for the Blogspot long-short portfolio, adjusting for transaction costs 170
Figure 31: Plot of alpha for the long-short Blogspot & Seekingalpha portfolios, adjusting for transaction costs 172
Figure 32: Plot of the mean excess returns of the long-short Blogspot & Seekingalpha portfolios, adjusting for transaction costs 173
Figure 33: Plot of the monthly Sharpe ratio for the long-short Blogspot & Seekingalpha portfolios, adjusting for transaction costs 175
Figure 34: Plot of alpha for the long-short Blogspot & Seekingalpha portfolios, ignoring transaction costs 176
Figure 35: Plot of the mean excess returns of the long-short Blogspot & Seekingalpha portfolios, ignoring transaction costs 176

Figure 36: Plot of the monthly Sharpe ratio for the long-short Blogspot & Seekingalpha portfolios, ignoring transaction costs 177
Figure 37: Plot of alpha for the long-short Blogspot & Seekingalpha portfolios and the momentum simulation's portfolio, adjusting for transaction costs 180
Figure 38: Plot of the mean excess returns for the long-short Blogspot & Seekingalpha portfolios and the momentum simulation's portfolio, adjusting for transaction costs. 182
Figure 39: Plot of the monthly Sharpe ratio for the long-short Blogspot & Seekingalpha portfolios and the momentum simulation's portfolio, adjusting for transaction costs. 185
Figure 40: Plot of alpha for the long-short Blogspot & Seekingalpha portfolios and the momentum simulation's portfolio, ignoring transaction costs..... 186
Figure 41: Plot of the mean excess returns for the long-short Blogspot & Seekingalpha portfolios and the momentum simulation's portfolio, ignoring transaction costs 187
Figure 42: Plot of the monthly Sharpe ratio for the long-short Blogspot & Seekingalpha portfolios and the momentum simulation's portfolio, ignoring transaction costs. 187
Figure 43: Excerpt of a Blogspot document about an iPod with cell phone capabilities ... 213
Figure 44: Excerpt of a Blogspot document about Apple as a cell phone provider 213
Figure 45: Excerpt of a Seekingalpha document about Apple's upcoming iPhone release 214
Figure 46: Excerpt of a Seekingalpha document about an anticipated iPhone announcement 214
Figure 47: Excerpt of the official press release of Apple's iPhone announcement 214

List of Abbreviations

AD	Anderson-Darling
AMEX	American Stock Exchange
API	Application Programming Interface
APT	Arbitrage Pricing Theory
BaH	Buy-and-Hold
BPS	Basis Points
CAPM	Capital Asset Pricing Model
CI	Confidence Interval
CRSP	Center for Research in Security Prices
DJIA	Dow Jones Industrial Average
DJICI	Dow Jones Internet Commerce Index
DJNS	Dow Jones News Service
FF	Feature Frequency
HTML	Hypertext Markup Language
IDF	Inverse Document Frequency
IR	Information Retrieval
JB	Jarque-Bera
KB	Kilo Bytes
MACD	Moving Average Convergence / Divergence
ML	Machine Learning
MSE	Mean Squared Error
MSH	Morgan Stanley High-Tech Index
NASDAQ	National Association of Securities Dealers Automated Quotations
NLP	Natural Language Processing
NW	Newey-West
NYSE	New York Stock Exchange
OLS	Ordinary Least Squares
POS	Part of Speech
RSS	Really Simple Syndication
SRT	Signed Rank Test
S&P	Standard & Poor's
SRM	Structural Risk Minimization
SVM	Support Vector Machine
SVR	Support Vector Regression
TF	Term Frequency
URI	Uniform Resource Identifier
WSJ	Wall Street Journal

1 Introduction

1.1 Motivation

Investor sentiments are real-world investors' beliefs that drive their demand for stocks (Shleifer, 2000). These beliefs are often subject to systematic biases that can induce abnormal returns (Hirshleifer, 2001). As earning abnormal returns is a specific goal of investors, measuring investor sentiment and studying its effects on abnormal returns has been subject to research in behavioral finance (e.g., (Shleifer, 2000)). In recent years, the World Wide Web has become a new and ever growing source of investor sentiments. With the advent of Web 2.0 technology (Murugesan, 2007), everyone having access to the web can publish their investor sentiment directly online.

So called weblogs (blogs for short) are personal online journals that have become popular in the Web 2.0. The number of blogs has increased from more than 35 million in 2006 to 173 million in 2011 (NM Incite, 2012). Investor sentiment is abundant in investment blogs where investors, for instance, present and discuss personal assessments of investment ideas – possibly before actually implementing trades. These investor sentiments may concern any type of financial instrument. Investor sentiment may also manifest as rumors about companies and their products. Thus, investor sentiment from blogs may provide new information that has not been priced yet, and thus may offer a potential profit opportunity. A recent study suggests that using sentiment from (generic topic) blog documents yields higher returns than using sentiments from news, Twitter, and LiveJournal (Zhang & Skiena, 2010).

To illustrate the importance of blogs as a source of investor sentiments, consider the titles of the blog documents shown in Figure 1 (full account of the documents is provided in Appendix A.1). The investor sentiments in the blog documents of Figure 1 provide rumors and indications for the actual news event several months in advance. The official press release of Apple, revealing the introduction of their first cell phone on 2007-01-09 sparked a substantial stock price increase (see Figure 1). But also well before the press release, the stock price was rising at least since July 2006 along with the publication of the blog documents. By means of analyzing the investor sentiment in these blog documents, other investors could have taken advantage by increasing their holdings of the Apple stock in their portfolio.

Figure 1: Apple stock price time series and related rumors published in blogs. The prices are close prices in US $. All rumors were published prior to the press release on 2007-01-09.

The understanding of the effect of investor sentiment from blog documents on abnormal returns is, however, still limited. Research has concentrated on other sources such as investor sentiment from surveys (e.g., (Brown & Cliff, 2005)) or market data proxy variables (e.g., (Baker & Wurgler, 2006)). The main disadvantages of surveys are as follows: they can be only conducted at low frequency (usually monthly) regarding the overall stock market (not individual stocks, due to the effort required) and only a small number of people contribute to the surveys. The main disadvantages of market data proxy variables (e.g., the put-call-ratio) are as follows: the market data proxy variables are elicited indirectly and only information after-the-fact that investors have put orders on the capital markets is included. Research regarding investor sentiment from textual sources has focused on news (e.g., (Leinweber & Sisk, 2011; Schumaker et al., 2012)) and generic topic blogs (e.g., (Gilbert & Karahalios, 2010; Zhang & Skiena, 2010)).

To study the effect of investor sentiment from blog documents on abnormal returns, this thesis makes the following contributions: First, investor sentiments must be made available by means of a time-efficient, consistent, and automatic classifier. The literature lacks (evaluated and reproducible) classifiers of investor sentiment. The major reason is that there is no publicly available standard corpus for evaluating a classifier. This research contributes a classifier of investor sentiment in blogs, which was rigorously evaluated using a novel corpus. Second, unlike prior research, this thesis studies investment-specific blog platforms

such as *Seekingalpha* and compares observed effects with those related to the large blog platform *Blogspot* based on the classifier proposed in this thesis. Seekingalpha has been subject of a recent study (Chen et al., 2014), which, however, did not explicitly relate to investor sentiment and did not explicitly use a classifier of the sentiment orientation. Furthermore, in contrast to most prior studies, this study focuses on monthly aggregates of investor sentiment, mean monthly effects on abnormal returns, and also considers transaction costs. Overall, this research enhances the understanding of the effect of investor sentiment from blog documents on abnormal returns of portfolios of stocks on the monthly level.

1.2 Research Approach

In this thesis, a perspective rooted in behavioral finance theory is used to study the research object, i.e., investor sentiment in blog documents and its relation to abnormal returns. Behavioral finance is an emerging theory aiming to explain possibly non-rational human behavior in stock markets and to predict effects on stock price outcomes (Shleifer, 2000). This thesis studies effects of investor sentiment on abnormal returns. Abnormal returns are in excess of normal returns according an asset pricing model (e.g., (Kothari & Warner, 2007, p.9; Mitchell & Stafford, 2000, pp.308–309)).

Behavioral finance theory has emerged in opposition to the efficient market theory (Shiller, 2003). The semi-strong form of the efficient market hypothesis is the relevant form for this thesis (Fama, 1970). Under this form of the hypothesis, a trading strategy that yields higher expected returns than normal returns would be impossible using only publicly available information (Fama 1970, p.385), including the one from blogs. An informationally efficient market would always fully reflect such information in prices (Fama 1970, p.383). Thus, investor sentiment from blog documents would be useless for investors. However, there is considerable evidence for the existence of inefficiencies in financial markets (e.g., surveyed by (Subrahmanyam, 2007)).

Behavioral finance theory explains mispricings by embracing human failures in decision making under risk due to psychological biases (e.g., (Hirshleifer, 2001; Shleifer, 2000; Tversky & Kahneman, 1974)). For instance, discovering naïve patterns in past price movements is a widespread anomaly used by investors in their decision making (De Bondt, 1998). Behavioral finance theory rests on two major foundations: (1) investor sentiment, and (2) limited arbitrage (Shleifer 2000, p.24).

First, investor sentiment is "[…] the theory of how real-world investors actually form their beliefs and valuations, and more generally their demand for securities." (Shleifer 2000, p.24). Classical examples for investor sentiment are "pseudo-signals" (Shleifer & Summers 1990, p.23) such as advice of brokers, financial gurus, and signals from trend-chasing strategies (Shleifer & Summers 1990, p.23).

Second, due to limits of arbitrage that would counter mispricings (Shleifer & Vishny, 1997) and trades of investors being often correlated (e.g., (Shleifer & Summers 1990, pp.23-

24), investor sentiment can persist and have long term effects on prices. Empirical evidence has been found, for instance, for investor sentiment derived from market newsletters over one to three years (Brown & Cliff, 2005) and cross-sectional effects of market data-based investor sentiment on a monthly basis (Baker & Wurgler, 2006).

The research interest of this thesis is on whether investor sentiment from *blog documents of investment blogs* has an effect on abnormal stock returns on the portfolio level. Based on investor sentiment in multiple blog documents from a certain time interval, an aggregate numeric measure of investor sentiment was constructed. Monthly aggregates were used and monthly effects were studied, corresponding to the typical frequency of fund evaluation (e.g., (Kothari & Warner, 2001)). With respect to the aggregated investor sentiment, the central research question is as follows:

Research question: *What is the effect of investor sentiment from blog documents on abnormal returns of a portfolio of stocks?*

This research contributes to bridging the gap in the behavioral finance literature on understanding the effect of investor sentiment extracted from *blog documents* on abnormal stock returns. To study the effect, sentiment classified from blog documents using text classification methods is being related to the concept of investor sentiment from behavioral finance theory and was used for stock selection in a portfolio simulation.

The research design is illustrated in Figure 2.

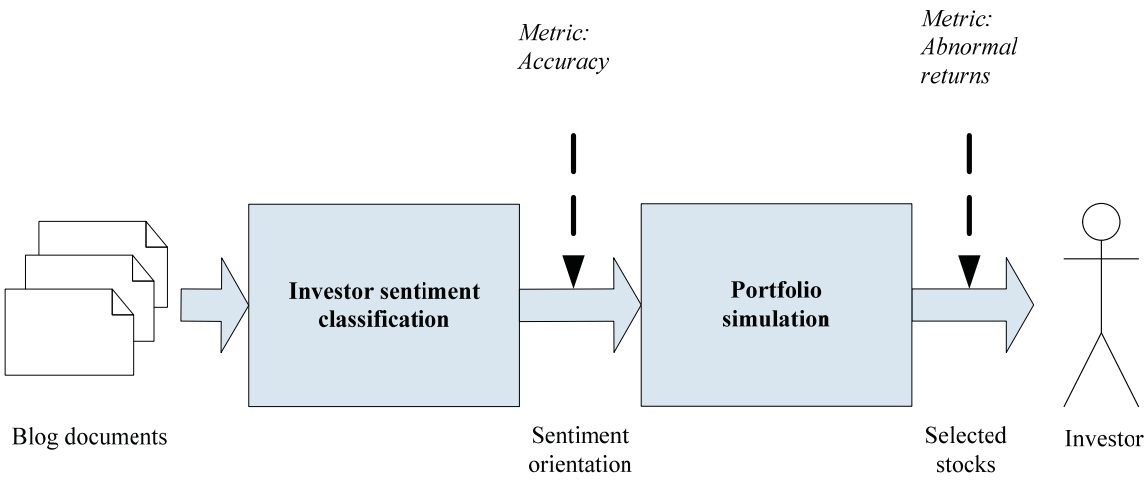


Figure 2: Research design.

A necessary prerequisite for the study was to design and configure an investor sentiment classifier for blog documents. The problem of designing an investor sentiment classifier for the text of blog documents can be attributed to the field of opinion mining and sentiment analysis (Pang & Lee, 2008). The design was guided and evaluated by the accuracy of the approach, i.e., a text classification metric (e.g., (Sokolova & Lapalme, 2009)). Due to ambiguity in financial web documents (Das & Chen, 2007) and subtlety in sentiment expressions in general (Pang & Lee, 2008, p.19), achieving high accuracies is difficult. The

relevant parameters and the best configuration that determine the accuracy were identified, based on findings in the literature and experiments. The design of the classifier focused on the textual level of individual English language blog documents, neglecting pictures, comments, other elements, and relationships between different textual parts. For evaluating the accuracy of the classifier, a novel corpus was created and evaluated.

The study was guided by hypotheses derived from behavioral finance theory. The hypotheses were tested by means of a portfolio simulation focusing on the mean or median abnormal returns on the monthly level of a portfolio of stocks. The portfolio simulation used the level of aggregated monthly investor sentiment from blog documents for selecting the stocks. For practical relevance, the study tested whether the mean or median abnormal returns exceed typical transaction costs.

1.3 Thesis Structure

This thesis is organized into five sections.

Section 1 defines the research objectives.

Section 2 presents the results of the analysis of the state of the art with respect to predictions of behavioral finance theory regarding the effect of investor sentiment on (abnormal) returns of (portfolios of) stocks. Furthermore, investment blogs from two platforms as a source of investor sentiment are described. Approaches for the classification of the sentiment orientation of investor sentiment from these blog documents are discussed. Findings of prior studies of effects of investor sentiment from textual sources on (abnormal) returns are analyzed. Finally, the research gap is lined out.

Section 3 describes the design of a classifier for the sentiment orientation of investor sentiment from blog documents, based on findings in the literature and experiments for determining parameter settings. The classifier was created using a machine learning approach and a novel corpus of manually classified blog documents. Finally, the classifier's accuracy is evaluated using the corpus.

Section 4 presents hypotheses on effects of (aggregated) investor sentiment from blog documents from two blog platforms on abnormal returns of a stock portfolio. To test the hypotheses, five year datasets of blog documents referring to specific stocks from the two blog platforms were used. The aggregated sentiment orientation of the investor sentiments, classified from the blog documents, was used in a portfolio simulation for stock selection. By means of statistical tests, the existence of abnormal returns on the portfolio level or relative differences of abnormal returns among different configurations were tested.

Section 5 concludes by summarizing the research contributions, practical implications, and limitations of the study.

2 Analysis of the State of the Art

This section analyzes the state of the art regarding effects of investor sentiment in blog documents on abnormal returns by (1) discussing concepts from asset pricing theory and behavioral finance to define investor sentiment and model abnormal returns, (2) introducing blogs as a new source of investor sentiment, (3) examining approaches for automatically classifying the sentiment orientation of investor sentiment in blog documents, (4) discussing prior studies on the effects of textual investor sentiment on abnormal returns, and (5) defining the research gap.

2.1 Asset Pricing Theory

To identify *abnormal* returns related to investor sentiment, one needs to understand how assets should *normally* be priced in theory. This section presents the relevant concepts of asset pricing theory and efficient market theory. Several asset pricing models to estimate normal and abnormal returns, which were also used in previous studies, are presented. Finally, some example evidence on the existence of abnormal returns is provided.

2.1.1 Inefficient Markets and Abnormal Returns

In asset pricing theory, the **fair price** should equal the expected value of discounted future asset payoffs (Cochrane, 2005, pp.4–6; Fisher, 1974). For stocks, future payoffs are the future price plus dividends (e.g., (Cochrane, 2005, p.4)). The fair price of a stock is equivalent in meaning to the terms "fundamental value" (e.g., in (Fama, 1991, p.1577)) or "intrinsic value" (e.g., in (Ou & Penman, 1989, p.296)).

In an efficient market, the current price reflects a very good estimate of the fair price (Fama, 1965a, p.90). "A market in which prices always 'fully reflect' available information is called 'efficient.'" (Fama 1970, p.383). The **semi-strong form** of the **efficient market hypothesis** (EMH) is the relevant form of efficiency for this research: the information includes all obviously publicly available information (Fama, 1970, p.404). That is, also information in textual form, such as in blog documents, is included. When the fair price changes (due to new information), the price will adjust "instantaneously" (Fama, 1965a, pp.36,94).

A deviation of the observed price from the fair price is a **mispricing** (e.g., (Sharpe et al., 1998, p.921), referring to the fair price as "investment value"). Such mispricings can be induced by irrational (or not fully rational) behavior of investors (e.g. (Shiller, 2000)), which can relate to investor sentiment (Shleifer & Summers, 1990). When the observable price of a stock is below the estimated fair price, it is considered to be underpriced (e.g., (Bodie et al.,

2009, p.589)). Complementarily, when the observed price is higher than the estimated fair price, the stock is considered to be overpriced (e.g., (Shleifer, 2000, p.3)).

Abnormal returns (i.e., abnormal relative price changes) can be realized by investors, when the mispricing is built up, or when it disappears (e.g. (Penman, 2013, pp.661–663)). Abnormal returns represent empirical deviations from expected returns (e.g., (Mitchell & Stafford, 2000, pp.308–309)). Theoretically expected returns are termed **normal returns** in this thesis (like in, e.g., (Fama, 1998, p.285)). Normal returns compensate an investor for bearing the risk of variable future returns, i.e., the risk premium (e.g., (Penman, 2013, pp.643–644)). Adding a time premium (proxied by the risk free interest rate) as a compensation for the unavailability of the invested capital (e.g., (Grinold & Kahn 2000, p.91) to the risk premium yields the **required return** (e.g., (Penman, 2013, pp.643–644)).

Figure 3 depicts the relationship between the creation of a mispricing and the generation of abnormal returns, which exceed the actually required returns (comprising the normal returns for the given level of risk and the risk free rate).

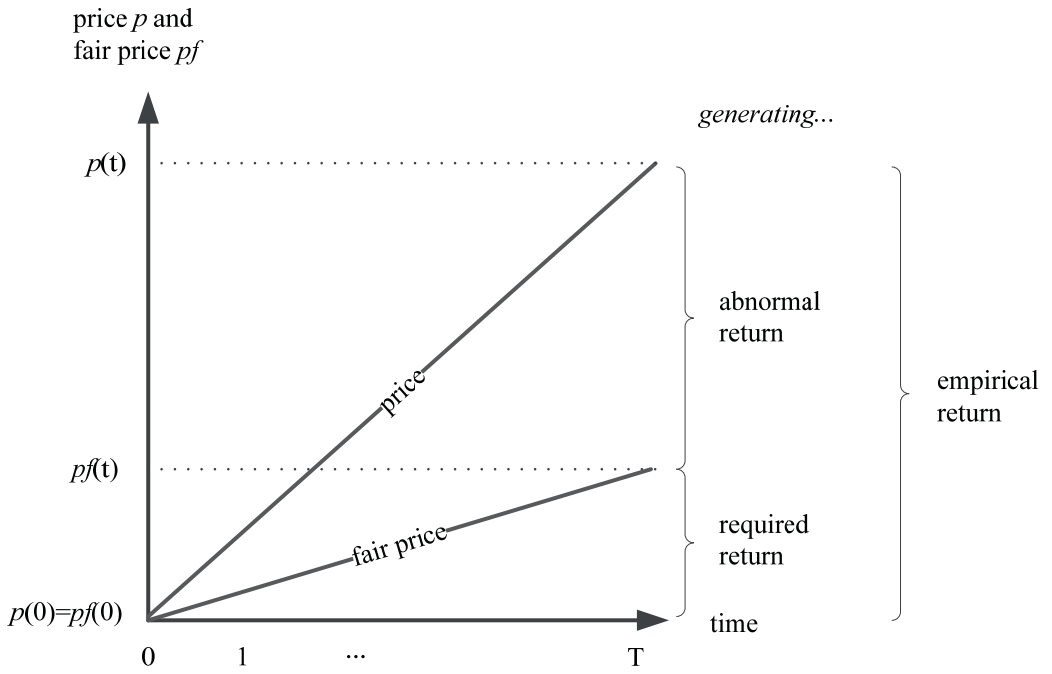


Figure 3: Built-up of a mispricing, inducing abnormal return (based on (Penman, 2013, p.662)).

Empirical tests for abnormal returns typically require a model of normal returns. Models of normal returns are discussed in the next section. Note the problem of testing jointly two hypotheses (e.g., noted in (Fama, 1991, pp.1575–1576) with respect to testing market efficiency) of (1) the existence of abnormal returns, and (2) the model of normal returns is correct. That is, identifying abnormal returns depends on the correctness of the model and its risk factors.

The more inefficient the market, the more abnormal returns will exist and persist (when assuming normal returns in efficient markets, see (Fama, 1970, pp.384–385)). However, in

efficient market theory (EMT), the price will be driven back to the fair price (Fama, 1965a, p.38). Reversion to the fair price is due to arbitrage, which is an essential concept in finance (e.g., (Poitras, 2010) for a review). Below, a textbook definition is provided.

Definition: Arbitrage: "Arbitrage is the process of earning riskless profits by taking advantage of differential pricing for the same physical asset or security" (Sharpe et al., 1998, p.284). Arbitrage "[…] typically entails the sale of a security at a relatively high price and the simultaneous purchase of the same security (or its functional equivalent) at a relatively low price." (Sharpe et al., 1998, p.284).

Arbitrage can be applied to buying (selling or selling short) an under- (over-) priced security and selling (buying) a similar security at a price closer to the fair price (Shleifer, 2000, pp.3–4). This assumes that such a similar security exists and that the fair price can be determined correctly. The effect of arbitrage is to move asset prices to their fair price and thus to counter mispricings and make markets more efficient (e.g., (Shleifer & Vishny, 1997, p.35)). Note that this thesis does not test for market (in)efficiency. It rather studies the potential effect of investor sentiment from blogs on abnormal returns, based on assuming a state of the art model of normal returns.

2.1.2 Models of Normal and Abnormal Returns

Rational asset pricing models serve to estimate theoretically expected – or normal –returns and can be used to identify abnormal returns. Two forms of abnormal returns are distinguished in this thesis: (1) abnormal returns that occur regularly, and (2) transitory abnormal returns in the course of and following an unexpected event. The latter are termed **unexpected abnormal returns** in this thesis. Different models of normal returns take different risk factors into account for which normal returns are assumed to compensate by these models. According to Cochrane (2005, p.152), the most famous asset pricing model is the single risk factor Capital Asset Pricing Model (CAPM). The CAPM is discussed in the following subsection. There are also multi-factor models such as the Fama-French model, which is discussed subsequently. Finally, Carhart's extending model (Carhart, 1997) is presented. Note that it can be subject to debate from a theoretical perspective whether the additional factors of these models represent (rationally) priced risk factors or (irrational) mispricings (e.g., (Chordia & Shivakumar, 2002; Jegadeesh & Titman, 2001)). However, both, the Fama-French and the Carhart model, have been empirically successful in explaining returns (Carhart, 1997, pp.61–62; Fama & French, 1993, 1996).

2.1.2.1 Capital Asset Pricing Model

This section discusses the CAPM by Sharpe and Lintner (Lintner, 1965; Sharpe, 1964). The CAPM only considers a single risk factor, i.e., systematic risk (Sharpe, 1964, p.436) or market risk (Mossin, 1966). However, "market risk accounts for most of the risk of a well-diversified portfolio" (Brealey et al., 2011, p.203). The CAPM defines a linear relationship between normal returns of an asset (in excess of the risk free rate) and systematic risk in

market equilibrium (e.g., (Sharpe, 1964, pp.436–442)). The amount of systematic risk of an asset is given by its sensitivity β to the expected return of the market portfolio in excess of the risk free rate (e.g., (Jensen, 1968, pp.390–391) and (Brealey et al., 2011, p.231)). To estimate empirical betas for arbitrary portfolios and for individual assets and to also estimate abnormal returns in excess of the CAPM normal returns, the following time series regression model can be used.

Definition: CAPM Regression Model & Jensen's Alpha (adapted from (Jensen, 1968, pp.390–394), and (Grinold & Kahn, 2000, pp.13–15)):

$$r_P(t) - r_F(t) = \alpha_P + ß_P\big(r_I(t) - r_F(t)\big) + \varepsilon_P(t) \tag{2.1}$$

where

r_P: Return of portfolio P (e.g., (Jensen, 1968, pp.390–394)).

r_F: Risk free interest rate (e.g., (Jensen, 1968, pp.390–394)).

r_I: Return of an index that proxies the market portfolio, e.g., a stock index (e.g., (Grinold & Kahn, 2000, p.13)).

β_P: Measure of systematic risk (e.g., (Jensen, 1968, pp.390–394)).

α_P: Jensen's (1968) alpha – the constantly occurring abnormal return (see below).

ε_P: The error term (e.g., (Jensen, 1968, pp.393–394)), which can be interpreted as unexpected abnormal return (see, e.g., (Brown & Warner, 1980, p.208) and below).

t: Discrete time.

The normal return part (representing the risk premium) of the CAPM regression model can be defined as follows:

Definition: CAPM Normal Return (adapted from (Jensen, 1968, pp.390–394), and (Grinold & Kahn, 2000, pp.13-15):

$$ß_P\big(r_I(t) - r_F(t)\big) \tag{2.2}$$

where variables are defined as in Definition (2.1).

The logarithmic **return** can be defined as a function that determines the relative price change in a time period:

$$r(t) = \ln\big(p(t)\big) - ln\big(p(t-1)\big) \tag{2.3}$$

where

r: Return of a portfolio of assets.

p: Price of a portfolio of assets.

ln: Logarithm to the base of *e*.

t: Discrete time.

Exceeding the normal return, the alpha variable (i.e., α) in the CAPM regression model in Definition (2.1) represents the *average* return per time period (Jensen, 1968, pp.393–394) that can be defined as **abnormal return** (e.g., (Elton & Gruber, 1997, p.1753)). That is, alpha does not depend on time and represents the abnormal return that can be observed constantly on average in every period. In efficient markets, abnormal returns should be zero (when assuming normal returns in efficient markets, see (Fama, 1970, pp.384–385)). Empirical positive alpha represents a risk-adjusted return and is a performance measure that usually is used to measure the *long-term* (i.e., monthly or yearly) abnormal returns of mutual funds, i.e., *portfolios* of multiple assets (e.g., (Jensen, 1968; Kothari & Warner, 2001)).

The *short-term* (i.e., several days) abnormal returns of *individual* assets that do occur *transitory*, related to a specific, event are measured differently. That is, beside normal returns and alpha, the CAPM regression model leaves the error term $\varepsilon(t)$. The error term ε can be interpreted as the unexpected *or* abnormal return of an individual security (e.g., (Brown & Warner, 1980, p.208)). This thesis uses the term **unexpected abnormal return** to distinguish it from the mean abnormal returns in terms of alpha. The *expected value* of the unexpected abnormal return should be zero (i.e., $E(\varepsilon)=0$) in an efficient market (Brown & Warner, 1980, p.209).

2.1.2.2 Fama-French Model

The Fama-French model is a three-factor model of normal returns (representing the risk premium), which extends the original CAPM with two priced risk factors based on observed pricing anomalies with respect to CAPM predictions (e.g., (Fama & French, 2004) for a review) and the claim that stock risks are multidimensional (Fama & French, 1992, 1993). The two additional risk factors have been found empirically as: (1) average returns of small stocks are higher than the ones of large stocks (the size factor), and (2) average returns of stocks with high book value compared to market value are higher than for stocks with a low ratio (the book to market equity factor) (Fama & French, 1992). Including the two factors, the Fama-French model absorbs most of the pricing anomalies of the CAPM (Fama & French, 1996, p.56) and well explains average returns across stocks (Fama & French, 1993, 1996). Due to the empirical success, Fama and French argue that the two factors are priced

risk factors (e.g., (Fama & French, 1993, p.5, 1996, pp.56–57)). The time series regression form of the Fama-French model for a portfolio of stocks or individual stocks is defined as:

Definition: Fama-French Model (adapted from (Fama & French, 1993, 1996, pp.55-56)):

$$r_P(t) - r_F(t) = \alpha + ß_1\big(r_I(t) - r_F(t)\big) + ß_2 \cdot SMB(t) + ß_3 \cdot HML(t) + \varepsilon(t) \quad (2.4)$$

where

r_P: Return of portfolio P.

r_F: Risk free interest rate.

r_I: Return of an index that proxies the market portfolio, e.g., a stock index.

$ß_1$: Sensitivity of the portfolio's return with respect to the return of the index.

SMB: Difference of return of a portfolio of small stocks and a portfolio of large stocks (Fama & French, 1993, 1996, pp.55-56).

$ß_2$: Sensitivity of the portfolio's return with respect to size factor.

HML: Difference of return of a portfolio of stocks with a high ratio of book to market value and return of a portfolio of stocks with a low ratio (Fama & French, 1993, 1996, pp.55-56).

$ß_3$: Sensitivity of the portfolio's return with respect to the book to market equity factor.

α_i: The mean abnormal return similarly to Jensen's (1968) alpha.

ε: The error term, which can be interpreted as unexpected abnormal return (see, e.g., (Brown & Warner, 1980, p.208) and Section 2.1.3).

t: Discrete time.

2.1.2.3 Carhart Model

The model of Carhart for normal returns adds the momentum factor to the Fama-French three-factor model to account for the momentum effect (Carhart, 1997). The momentum effect relates to the observation of positive returns on buying (selling) stocks that have highest (lowest) returns in the last months (Jegadeesh & Titman, 1993). This effect is not explained by the Fama-French model ((Fama & French, 1996) cited in (Carhart, 1997, p.61)). Carhart found his model to perform better in terms of lower average pricings errors than the CAPM and the Fama-French model (Carhart, 1997, p.62). The time series regression form of the Carhart model for a portfolio of stocks or individual stocks is defined as:

Definition: Carhart Model (adapted from (Carhart, 1997, p.67)):

$$r_P(t) - r_F(t) = \alpha + ß_1\big(r_I(t) - r_F(t)\big) + ß_2 SMB(t) + ß_3 HML(t) + ß_4 MOM(t) + \varepsilon(t) \quad (2.5)$$

where

r_P: Return of the portfolio P.

r_F: Risk free interest rate.

r_I: Return of an index that proxies the market portfolio, e.g., a stock index.

$ß_1$: Sensitivity of the portfolio's return with respect to the return of the index.

SMB: Return for the size factor, see (Fama & French, 1993).

$ß_2$: Sensitivity of the portfolio's return with respect to size factor.

HML: Return for the book to market equity factor, see (Fama & French, 1993).

$ß_3$: Sensitivity of the portfolio's return with respect to the book to market equity factor.

MOM: Return of the portfolio that mimics the momentum factor, computed "[…] as the equal-weight average of firms with the highest 30 percent eleven-month returns lagged one month minus the equal-weight average of firms with the lowest 30 percent eleven-month returns lagged one month." (Carhart,1997,p.61).

$ß_4$: Sensitivity of the portfolio's return with respect to the momentum factor.

α: The mean abnormal return similarly to Jensen's (1968) alpha.

ε: The error term, which can be interpreted as unexpected abnormal return (see, e.g., (Brown & Warner, 1980, p.208) and Section 2.1.3).

t: Discrete time.

2.1.3 Estimating Unexpected Abnormal Returns

To identify and quantify the effect of a certain *event* related to investor sentiment on prices of *individual stocks* at typically daily frequency, unexpected abnormal returns can be estimated on the event day and in the following days. Unexpected abnormal returns can be generically defined as realized returns (conditional on an event) in excess of normal returns (unconditional on the event) according to a specific model (Kothari & Warner, 2007, p.9):

Definition: Unexpected Abnormal Return (adapted from (Kothari & Warner, 2007, p.9)):

$$ra_i(t) = r_i(t) - rn_i(t) \quad (2.6)$$

where

ra_i: Unexpected abnormal return of asset i.

r_i: Realized and observed return of asset i.

rn_i: Normal return of asset i with respect to an (unexpected) event.

t: Discrete time.

The central variable in the definition above is the normal return. Note that the normal return regarding an (unexpected) event is defined differently (below) compared to Section 2.1.2. To estimate normal returns, there are basically economic and statistical models in the literature (Campbell et al., 1997a, p.153). The economic models require restricting economic assumptions (Campbell et al., 1997a, p.154). Economic models have been presented in Section 2.1.2 in Definitions (2.1), (2.4), and (2.5). The normal return (to be used here) comprises the respective model's right hand side with the unexpected abnormal return term ε subtracted (see Section 2.1.2). That is, the normal return includes also the alpha coefficient (as defined by, e.g., (Campbell et al., 1997, pp.158–159)). The normal return in the event time period can be estimated by these models using *coefficients* (i.e., α and βs) of these models *estimated* in a time period *prior* to the event time period (Campbell et al., 1997, p.152).

Statistical models have only statistical assumptions (which are required also for economic models in practice to be used for estimation) (Campbell et al., 1997a, pp.153-154). Usually, statistical models assume returns to be "[…] jointly multivariate normal and independently and identically distributed through time." (Campbell et al., 1997a, p.154). A common statistical model is the market model with one factor that relates the return of an asset to the market portfolio's return (Campbell et al., 1997, pp.151,155). Because the market portfolio is unobservable, for estimating the model with respect to the stock market, a broad-based stock index such as the S&P 500 index can be used (Campbell et al., 1997, p.155). An early form of the market model was proposed by Sharpe (Sharpe, 1963, p.281). However, Sharpe's version is defined as one-period model. The market model has essentially the same form as the CAPM regression model (see Section 2.1.2.1). However, there are no economic restricting assumptions in the market model, which can be defined as follows:

Definition: Market Model (adapted from (Campbell et al., 1997b, p.155)):

$$r_i(t) = \alpha_i + ß_i r_I(t) + \varepsilon_i(t) \tag{2.7}$$

where

r_i: Return of asset i.

r_I: Return of an index, e.g., a stock index.

β_i: Sensitivity of asset i's return with respect to the return of the index.

α_i: Residual return of asset i.

ε_i: Unexpected abnormal return.

t: Discrete time.

The **normal return** with respect to an (unexpected) event **of the market model** is given by: $rn_i(t) = \alpha_i + ß_i r_I(t)$ (e.g., (Brown & Warner, 1985, p.7)).

2.1.4 Example for Abnormal Return Existence

Abnormal returns can be observed when a mispricing of a stock (i.e., an overpricing or underpricing relative to an estimated fair price) develops or ceases (see Section 2.1.1). The mispricing observation assumes a correct estimate of the fair price. Exemplarily, this section considers a historic time period that observed many overpriced stocks to provide an example of the generation of abnormal returns.

An example of a period of overpricing in the stock market has been "the Internet or Dot.com bubble of the 1990s" with large stock price increases, which could be hardly explained by rational valuation models (Baker & Wurgler, 2007, p.129). For instance, an estimate of low expected future cash flows of a company might lead to a low fair price of its stock compared to the observed price and thus to an overpricing of the stock. Fundamental analysts could now assume future prices to converge to the fair price (e.g., (Bodie et al., 2009, p.595) who refer to intrinsic value). Under this assumption, the overpricing would be expected to decrease. However, in reality this might be not the case. In contrast to fundamental analysis, technical analysis (e.g., (Murphy, 1999)) does not consider the economic prospects of a company. Rather, technical analysts engage, for instance, in price trend following (also termed positive feedback trading) by assuming increasing (or decreasing) prices to persist (e.g., (De Long et al., 1990a)). Under this assumption, an overpriced stock's price might still increase despite its overpricing. During the built-up and continuation of the overpricing, positive abnormal returns might be observed.

An example of an overpriced stock during the end of the 1990s is the eToys stock ((Edgecliffe, 1999) cited in (Shiller, 2000, p.176)). eToys' business model was to sell toys over the internet ((Edgecliffe, 1999) cited in (Shiller, 2000, p.176)). In contrast to the traditional over-the-counter toy retailer Toys "R" Us, eToys' sales (profits) were $30 million (-$28.6 million) compared to $11.2 billion ($376 million) in 1998 ((Edgecliffe, 1999) cited in (Shiller, 2000, p.176)). However, eToys was valued $8 billion compared to $6 billion of Toys "R" Us in 1999 ((Edgecliffe, 1999) cited in (Shiller, 2000, p.176)). Regarding the overall U.S. stock market, the increase in prices was not matched in real earnings growth during the 1990s (Shiller, 2000, p.6). In January 2000, the price-earnings ratio had risen to an unprecedented level above 44 (Shiller, 2000, p.8). Considering the stock price, the all-time-high was on 1999-10-11 at about $84 and had risen from about $45 at the beginning of September 1999 (source: Datastream). This time period is exemplarily considered in the following to test for abnormal returns.

To estimate the distinct day-by-day change of the abnormal return relative to the previous level of alpha, the Fama-French model (see Section 2.1.2.2) was estimated in the preceding estimation window (1999-05-24 until 1999-08-31), which starts shortly after the first stock trading day. The estimated coefficients of alpha and betas were used to calculate the unexpected abnormal return in the period 1999-09-07 until 1999-10-11 by resolving the formula of Definition (2.4) for unexpected abnormal return as described in Section 2.1.3.

Figure 4 shows that substantial positive unexpected abnormal returns existed over the period, amounting cumulatively to more than 50% in the end.

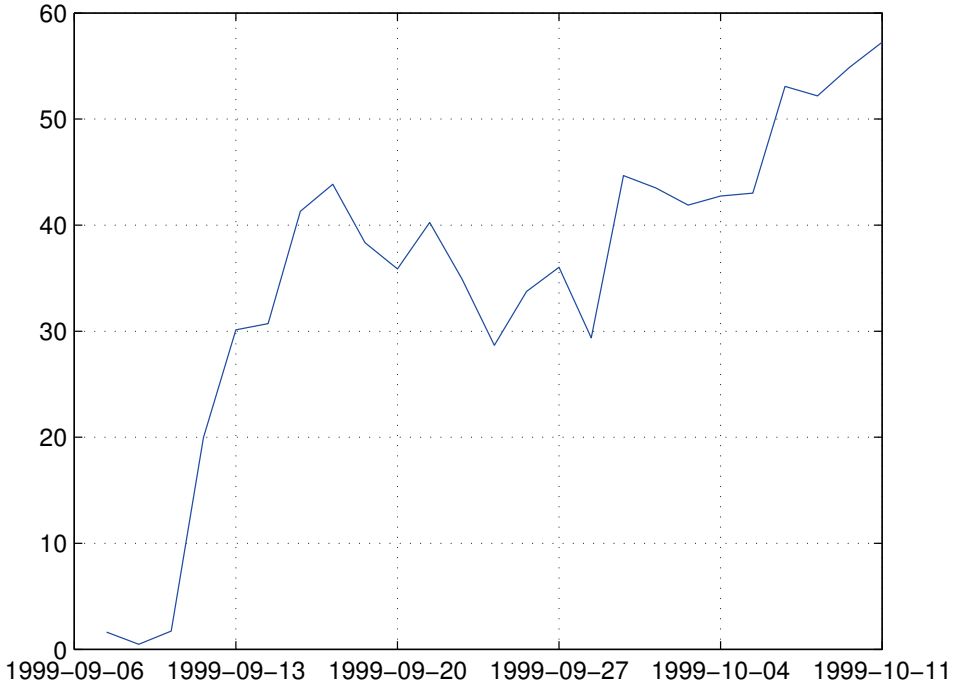


Figure 4: Daily cumulative unexpected abnormal returns for the eToys stock. Abnormal returns are in percent and were estimated by the Fama-French model using market data described in Appendix A.5.

To estimate the *mean* abnormal return of the eToys stock over the outlined time period, the alpha of the Fama-French model (see Definition (2.4)) was estimated in the period 1999-09-07 until 1999-10-11 using market data described in Appendix A.5. The *daily* alpha was estimated to be 3.1% and is statistically significantly different from zero at the 5% level (with a p-value=0.0379 and t-statistic=2.2162).

Attempting to explain the stock market price bubble during which the eToys example was observed, the 2000 book "Irrational Exuberance" makes reference to the famous 1996 speech of Alan Greenspan who had used this phrase to describe investors' behavior (Shiller, 2000, p.3). The book describes structural, cultural, and psychological factors that had led to the stock market price bubble (Shiller, 2000). One factor has been the adoption of internet technology in the 1990s by existing companies and startups (Shiller, 2000, pp.19–21). While the impact of such a technology on the valuation of companies can be questioned (based on economic growth models, e.g., (Barro & Sala-i-Martin, 1995), cited in (Shiller, 2000, p.21)), the public impression actually matters (Shiller, 2000, pp.19–21). In contrast to many other technologies, everyone can use the internet and create a website, thus making it plausible for many people that it would matter economically (Shiller, 2000, pp.19–21). Furthermore, like in Ponzi schemes, speculative feedback loops can amplify a price bubble (Shiller, 2000, pp.64–68). That is, when prices go up, investors are rewarded, thus triggering further buying (De Long et al., 1990a; Shiller, 2000, pp.64–68). In Section 2.2.1, psychological biases are

described that are considered as major drivers for the creation of mispricings (and thus, abnormal returns) in behavioral finance theory (e.g., (Shleifer, 2000)).

2.2 Behavioral Finance Theory

Behavioral finance theory guides the study of relationships of investor sentiment from blog documents to abnormal returns in this thesis and it informed hypothesis formulation about potential effects of investor sentiment. First, the theory of investor sentiment relates to psychological biases that systematically affect investors' decision making (e.g., (Hirshleifer, 2001; Shleifer, 2000; Tversky & Kahneman, 1974). Second, noise trader theory explains and predicts aggregate market effects of investor sentiment, which can induce mispricings and abnormal returns (e.g., (De Long et al., 1990b; Shleifer & Summers, 1990). Third, an overview of different types of measures of investor sentiment is provided.

2.2.1 Psychological Biases Affecting Asset Prices

The examples, provided in Section 2.1.4, for the existence of abnormal returns (in the form of alpha or unexpected abnormal returns) violate predictions of efficient market theory (see Section 2.1.1). This section argues that these violations result from systematic and predictable errors in investors' decision making that are due to psychological biases.

In his "portrait of the individual investor", De Bondt (1998) surveys evidence in the literature for widespread individual investor decision making anomalies and errors in four categories: (1) identification of price patterns, (2) asset valuation, (3) diversification, and (4) trading practices. Regarding the first category, individual investors engage in trend following and technical analysis of stock prices (De Bondt, 1998, pp.833–834). Thus, based on the self-fulfilling prophecy of trend following, positive autocorrelation and thus return predictability is induced into price and return time series (e.g., (De Long et al., 1990a; Moskowitz et al., 2012)). This effect violates the (weak form) EMH (Fama, 1970). Second, individual investors usually do not use valuation models from the literature (De Bondt, 1998, pp.834–835). Rather, they judge stocks, e.g., by recent company reputation and buy stocks of glamorous companies or companies with a good reputation (De Bondt, 1998, pp.834–835). Third, in violation of portfolio theory (Markowitz, 1952, 1959), individual investors underdiversify, own few assets in total and large amounts of riskless assets (De Bondt, 1998, pp.835–837). Fourth, many individual investors trade stocks based on impulses and random tips, instead of pursuing a plan (De Bondt, 1998, p.837). Furthermore, they sell winning stocks but favor not to realize losses (De Bondt, 1998, p.837). De Bondt provides further survey anomaly evidence and concludes that the picture of the individual investor is "sorry" (De Bondt, 1998, p.832).

At the root of investors' decision making errors, cognitive resource constraints make human decision makers use heuristic simplifications (Hirshleifer, 2001, p.1540). Heuristics simplify the assessment of probabilities and prediction tasks (Tversky & Kahneman, 1974,

p.1124). The heuristics lead to systematic and predictable errors (Tversky & Kahneman 1974, p.1131). Therefore, they are important for this thesis because investors' heuristics might induce possibly persistent abnormal returns.

Relying on heuristics leads to biases in judgment under uncertainty (Tversky & Kahneman, 1974). This thesis specifically focuses on biases of investors regarding (the estimation of) normal returns, which can be suspected to lead to abnormal returns. Hirshleifer categorizes sources for judgment and decision biases in (1) heuristic simplifications, (2) self-deception, and (3) emotional loss of control (Hirshleifer, 2001). Examples for simplifying heuristics are anchoring, representativeness, and conservatism (Hirshleifer, 2001, pp.1541–1548). Important examples of self-deception are overconfidence, overoptimism, and self-attribution (Hirshleifer, 2001, pp.1548–1550). Finally, Hirshleifer surveys evidence on emotions, e.g., moods affecting judgments and risky choices (Hirshleifer, 2001, pp.1550–1551). The following elaborates on each of the mentioned sources for biases and relates them to abnormal returns.

Anchoring refers to basing estimations of a value on an initial value (that may have been suggested) (Tversky & Kahneman, 1974, p.1128). As an example, regarding asset price prediction, De Bondt suggests individual (small) investors to use two anchors: (1) past price *changes*, and (2) representative past price *levels* (De Bondt, 1993, p.357). De Bondt (1993) found survey and experimental evidence of investors basing their predictions of price changes on past price changes, but to adjust their prediction of the price-level-range towards representative past price levels. For instance, in a continuous price uptrend, investors would predict the uptrend to continue (i.e., prices to rise), but the lower end of the predicted price-level-range would be biased towards a lower than today's price because past prices have been lower and are considered representative (De Bondt, 1993, p.357). De Bondt hypothesizes the former continuation prediction to be oriented short-term and the latter reverting prediction to be oriented long-term (De Bondt, 1993, p.368). This hypothesis is consistent with evidence for positive autocorrelation in stock returns on time horizons of up to one year (e.g., (Poterba & Summers, 1988)) and strong negative autocorrelations in stock returns on time horizons greater than one year (e.g., (Fama & French, 1988)). Poterba and Summers (1988) interpret this as a potential effect of deviation of prices from their fundamental values (i.e., fair prices) and reversion in the long-run. Consistent with this interpretation, a positive feedback trading (i.e., price trend following) model also suggests the emerging positive short-run return autocorrelation and negative long-term return autocorrelation to be associated with deviation of prices from fundamental values, i.e., fair prices (De Long et al., 1990a). Assuming this mispricing, one can expect abnormal returns (in terms of alpha and unexpected abnormal returns) to be realizable during its built-up (or fading) (see Section 2.1.1).

The **representativeness** heuristic refers to the assessment of the probability that some object A is related to another object B by the degree A is representative of (i.e., is similar to) B (Tversky & Kahneman, 1974, p.1124). This heuristic neglects, for instance, prior probabilities of A and B and leads to errors (Tversky & Kahneman, 1974, p.1124). That is,

Bayes' theorem on conditional probabilities is violated. For instance, a description of a person that is similar to a stereotype of a profession can lead to overestimate the probability that this person pursues this profession by not taking into account the (possibly very low) base-rate (i.e., the prior probability) of the profession (Tversky & Kahneman, 1974, p.1124). The representative heuristic can lead to price trends (Hirshleifer, 2001, p.1545). The price trends may be due to assuming that future price changes will be directed similar to recent past price changes. If a price trend refers to a mispriced stock, one can suspect representativeness to increase the mispricing and to induce abnormal returns (in terms of alpha and unexpected abnormal returns).

Conservatism is another relevant decision bias ((Edwards, 1968) cited in, e.g., (Barberis et al., 1998, p.315; Hirshleifer, 2001, p.1546; Tversky & Kahneman, 1974, p.1125)). Conservatism means to slowly change a belief given new evidence ((Barberis et al., 1998, p.315) with implicitly referring to the bias published by (Edwards, 1968)). That is, regarding rational Bayesian belief updating of the posterior probability, the update is too small in magnitude ((Edwards, 1968) cited in (Barberis et al., 1998, p.315)). The more useful the evidence, the larger the difference to the rational outcome becomes ((Edwards, 1968) cited in (Hirshleifer, 2001, p.1546)). Thus, conservatism can lead investors to not fully account for information of, e.g., an earnings announcement (Barberis et al., 1998, p.315). Bernard and Thomas are some of the researchers who present evidence on investors responding gradually on earnings announcements, and thus creating unexpected abnormal return over 60 trading days after an earnings announcement (Bernard & Thomas, 1989).

Overconfidence refers to overestimation of, e.g., the ability to do well on tasks, of own contributions to positive outcomes, and of occurrences of future events (see (Odean, 1998) for a survey). Odean models overconfidence "[…] as a belief that a trader's information is more precise than it actually is." (Odean, 1998, p.1893). Thus, on Bayesian belief updating regarding the posterior probability, they overweight their information (Odean, 1998, p.1893). Under certain circumstances (e.g., if information is interpreted differently) this can increase price deviations from the fundamental value (Odean, 1998, pp.1911–1912). Whereas under other circumstances the deviation might be decreased due to overconfidence, in total market efficiency should decrease in effect of overconfidence (Odean, 1998, p.1912). Assuming market efficiency to decrease and mispricings to increase, abnormal returns (in terms of alpha and unexpected abnormal returns) would be expected regarding the affected stocks (see Section 2.1.1).

Overconfidence implies **overoptimism** (Hirshleifer, 2001, p.1548) and has been evidenced in many settings ((Miller & Ross, 1975) cited in (Hirshleifer, 2001, p.1548)). For instance, analysts are overoptimistic about earnings of initial public offerings (IPO) of which many are underpriced (Rajan & Servaes, 1997, p.508). Thus, one can expect overoptimism to generate positive abnormal returns (in terms of alpha and unexpected abnormal returns) – at least in the short run.

Biased self-attribution refers to attributing favorable outcomes to one's abilities and unfavorable outcomes to external causes ((Langer & Roth, 1975; Miller & Ross, 1975) cited in (Hirshleifer, 2001, pp.1548–1549)). Self-attribution helps people learn to be overconfident (Hirshleifer, 2001, p.1549). For instance, a trade based on private information that is later confirmed by public information can lead to attributing the success to the own private information and in effect increase confidence (Daniel et al., 1998, p.1842).

The described errors in investors' decision making are widespread because potentially "[…] all investors may be imperfectly rational" (Hirshleifer, 2001, p.1536) and almost any investor is affected by some biases (Hirshleifer, 2001, p.1537). Even researchers who were thinking intuitively have been found to be affected (Tversky & Kahneman, 1974, p.1130). Furthermore, people usually do not detect their biases and also do not learn statistical rules from data (Tversky & Kahneman, 1974, p.1130). Thus, asset prices are affected *long-term* by fundamental psychological effects (Hirshleifer, 2001, p.1538). In effect, even though abnormal patterns in returns are well known, some do not disappear for many years (Hirshleifer, 2001, p.1539). For instance, the momentum effect was published in the early 1990s (Jegadeesh & Titman, 1993) and in 2011 the authors found the effect to still exist (Jegadeesh & Titman, 2011). The prevalence can result from uncertainty whether others have been already exploiting the effect or from information that is ignored or misused by everyone (Hirshleifer, 2001, p.1539).

2.2.2 Investor Sentiment in Behavioral Finance

Inefficiencies emerging as abnormal returns have been suggested to be caused by psychological biases in the literature reviewed in the previous section. These biases would affect investors' decision making under risk and in turn asset prices. The suggested explanations have served as a starting point for behavioral finance theory, in which investor sentiment is a central concept for explaining and predicting inefficient market outcomes (e.g., (Shiller, 2003; Shleifer, 2000)). Robert J. Shiller is one of the founders and major contributors of behavioral finance theory and was awarded the Nobel Prize in Economics in 2013 in this regard. He introduces the field as follows: "Behavioral finance—that is, finance from a broader social science perspective including psychology and sociology—is now one of the most vital research programs, and it stands in sharp contradiction to much of efficient markets theory." (Shiller, 2003, p.83). One fundamental reason for the "impossibility of informationally efficient markets" is that information is costly in reality and inefficiencies are required as compensation for acquiring the information (Grossman & Stiglitz, 1980).

The most important piece under the umbrella of behavioral finance with respect to this thesis is noise trader theory. Noise trader theory provides an alternative to efficient market theory which is claimed to be more plausible and more accurate (Shleifer & Summers, 1990). Thus, noise trader theory can explain the long-term existence of substantial mispricings (e.g., (De Long et al., 1990b)).

Noise trader theory is based on the following assumptions:

(1) Investor sentiment affects the demand for risky assets of some not fully rational investors (Shleifer & Summers, 1990, p.19).

(2) Arbitrage is limited because it entails risks and arbitrageurs have short time horizons (Shleifer & Summers, 1990, pp.19-20).

In the following, arguments are provided for these assumptions and also a definition of investor sentiment in relation to its effects on abnormal returns.

Limited Arbitrage

Arbitrageurs have rational expectations about returns, they are not subject to investor sentiment, and drive prices towards fair prices (Shleifer & Summers, 1990, p.20). Reconsidering the definition of arbitrage (see Section 2.1.1), it is conventionally assumed to be a riskless speculation that aims at buying a security at a low price and simultaneously selling a similar security at a high price. Arbitrage requires a substitute security, which may not exist in reality, e.g., for stock portfolios (e.g., (Shleifer & Summers, 1990, pp.20–21)). Thus, considering underpriced (overpriced) stocks, buying (selling) the stock without a hedging position would be an option (e.g., (Shleifer & Summers, 1990, p.21)). Trading mispriced stocks requires an estimate of the fair price (Shleifer & Summers, 1990, p.22). Assuming the price to move in direction of the fair price in the future, one could set up an exploiting position (e.g., (Penman, 2013, pp.661–664)).

However, arbitrage is assumed to be limited in noise trader theory because of the following types of risk: (1) fundamental risk, i.e., the fair price estimate or prediction may have been wrong (e.g., due to unforeseen fundamental news), and (2) future resale price risk, i.e., the future price might deviate even more from the fair price (De Long et al., 1990b; Shleifer & Summers, 1990, p.21). The deviation of prices from fair prices can be caused by noise trading in noise trading theory (De Long et al., 1990b, p.735). Noise trading can be defined as trading based on beliefs about expected returns that can deviate from normal returns (De Long et al., 1990b). The deviating beliefs of noise traders can get larger over time to even “extreme” levels and can persist for a long time (De Long et al., 1990b, p.705). The associated risk of arbitrageurs (who bet against the deviations) is termed “noise trader risk” (De Long et al., 1990b, p.705). Due to noise trader risk, arbitrageurs are forced to have short time horizons (De Long et al., 1990b, p.705). Otherwise, they might face large (and increasing) potential losses (De Long et al., 1990b, p.705).

The described risks limit the position sizes of arbitrageurs, and thus also their potential for driving prices towards fair prices (Shleifer & Summers, 1990, p.21). Another reason for the limits of arbitrage is that real-world arbitrage is costly because it involves transaction costs and holding costs (Pontiff, 1996). Holding costs are, e.g., interest for borrowing cash (for buying long positions) or fees for borrowing stocks (for selling short) (e.g., (Pontiff, 1996, p.1138; Shleifer & Summers, 1990, p.21)). Furthermore, real-world arbitrage requires

knowledge and information about specific assets that only some specialized professional investors possess (Shleifer & Vishny, 1997, p.36). To be effective, they require large positions, and thus external sources of capital (Shleifer & Vishny, 1997, p.36). If they lose money due to the described noise trader risk in the short run, the capital might be withdrawn from them (Shleifer & Vishny, 1997, p.37). Thus, noise trader risk limits the effectiveness of arbitrage (Shleifer & Vishny, 1997, p.54).

Due to the limits of arbitrage, it seems plausible that noise trading can affect stock prices. Shleifer also concludes "More than news seems to move stock prices." (Shleifer, 2000, p.20) on reviewing relevant literature (Shleifer, 2000, pp.16–23). Noise trading is linked directly to investor sentiment in the following.

Investor Sentiment

Investor sentiment is "[…] the theory of how real-world investors actually form their beliefs and valuations, and more generally their demand for securities." (Shleifer 2000, p.24). Investor sentiment can be regarded "noise" according to noise trader theory (Shleifer & Summers, 1990) because the demand of noise traders for risky assets "[…] is affected by their beliefs or sentiments that are not fully justified by fundamental news." (Shleifer & Summers, 1990, p.19). That is, noise in financial markets is in contrast to information (Black, 1986), which should be actually relevant for asset pricing.

Examples for noise are "pseudo-signals" (Shleifer & Summers, 1990, p.23). Providers of such signals are for instance "[…] technical analysts, stockbrokers, or economic consultants […]" (De Long et al., 1990b, p.706) and include for instance volume and price patterns, and forecasts of gurus (De Long et al., 1990b, p.735). Furthermore, trading strategies not based on fundamentals, such as trend following, can create uninformed demand changes (Shleifer & Summers, 1990, p.23).

Kyle was one of the first to use the term "noise traders" for investors with uninformed demand for a risky asset (Kyle, 1985). Thus, irrationality is attributed to noise traders (e.g., (De Long et al., 1990b, p.703)). Noise traders' demand that is not driven by fundamental information can be due to investor sentiment or constraints such as institutional restrictions on holdings (e.g., (Shleifer & Vishny, 1997, p.52)). In this thesis, only investor sentiment is considered as a driver of noise trading. Although being no fundamental information, noise traders trade "[…] on noise as if it were information." (Black, 1986, p.531). Thus, they can introduce noise into prices (Black, 1986, p.532).

Uninformed demand of noise traders (based on their investor sentiment) will only have an effect on asset prices if is correlated across noise traders (e.g., (Shleifer & Summers, 1990, p.23)). Thus, correlated misperceptions and correlated trading are essential assumptions in noise trader models (e.g., (De Long et al., 1991)). There is substantial evidence that supports the assumptions: Evidence from buying and selling orders of individual investors at a discount broker and retail broker suggests that the monthly correlation is about 75% and that

orders cumulate over time (Barber et al., 2009, pp.549,567,568). Furthermore, the trading decisions correlate with observed past returns: investors tend to buy (sell) stocks that have performed well in the past 1-3 years (1-3 quarters) (Barber et al., 2009, p.566). Barber et al. (2009) suggest this correlation to be due to psychological biases. Other U.S. retail investors trading data also shows correlated buying and selling and supports the influence of investor sentiment on stock returns and predictions of noise trader models (Kumar & Lee, 2006).

Noise trading makes financial markets inefficient to some degree (Black, 1986) and introduces mispricings (e.g., (De Long et al., 1990b; Shleifer & Summers, 1990; Shleifer & Vishny, 1997)). A major reason is that noise traders "[…] may be subject to systematic biases." (Shleifer & Summers, 1990, p.20). Such biases of investors have been described above (see Section 2.2.1). The resulting creation of mispricings can manifest in abnormal returns (see Section 2.1.1).

Noise traders' beliefs that can induce mispricings are also termed investor sentiment (e.g., (Shleifer & Summers, 1990)). Definitions of "investor sentiment" are diverse: Investor sentiment can be defined simply in terms of investors being either optimistic or pessimistic forecasters (e.g., (Siegel, 1992)). Also Brown and Cliff think of investor sentiment as "excessive optimism or pessimism" and a "bias" (Brown & Cliff, 2004, p.4). Several models of investor sentiment relate investor sentiment to different specific psychological biases that affect investors' belief formation required for asset valuation (e.g., (Barberis et al., 1998; Daniel et al., 1998)). Thus, the belief can be erroneous regarding the value (i.e., the fair price) of an asset. For valuation, cash flows can be used, thus investor sentiment can be defined as "[…] a belief about future cash flows and investment risks that is not justified by the facts at hand." (Baker & Wurgler, 2007, p.129). Investing based on erroneous valuations affects prices – if many investors act in the same way. Therefore, "Investor sentiment reflects the common judgment errors made by a substantial number of investors […]" (Shleifer, 2000, p.12). Regarding its effect on prices, investor sentiment can be regarded an error on the level of individual investors or the market, which on the aggregate level leads to inefficient prices (Shefrin, 2008, pp.6,9). Therefore, investor sentiment can be defined relative to normal returns: "Intuitively, sentiment represents the expectations of market participants relative to a norm: a bullish (bearish) investor expects returns to be above (below) average, whatever 'average' may be." (Brown & Cliff, 2004, p.2). The norm can be interpreted as the required return (Brown & Cliff, 2004, p.4), which equals the normal return plus the risk free rate (see Section 2.1.1). In this line, one can also assume positive (negative) levels of investor sentiment to relate to positive (negative) abnormal returns (in excess of normal returns).

This thesis' definition of investor sentiment focuses on the relative level of investor sentiment:

Definition: Investor Sentiment (adapted from (Brown & Cliff, 2004, p.2)): A higher (lower) level of investor sentiment is expectations of market participants regarding higher (lower) future abnormal stock returns.

As discussed in Section 2.1, abnormal returns could be measured based on alpha (i.e., the mean abnormal return per period of a portfolio of stocks) or unexpected abnormal returns (i.e., the short-term abnormal return of an individual stock). However, this thesis focuses on abnormal returns in terms of alpha.

Investor sentiment can have long term effects on prices because of the limits of arbitrage (Shleifer & Summers, 1990) and because of fundamental psychological effects (Hirshleifer, 2001, p.1538). In a noise trader model, it has been shown that noise traders as a group can survive long term in the market (De Long et al., 1991). Related to the stock market price bubble during the late 1990s, Shiller notes: "Absurd prices sometimes last a long time." (Shiller, 2000, p.176). Furthermore, noise traders as a group can earn higher returns than rational investors (being attributed to higher risk taking) (De Long et al., 1991).

2.2.3 Measures of Investor Sentiment

The previous section has discussed the theory about investor sentiment, attributing it an effect on stock prices. To be able to test the theory and quantify the effect of investor sentiment, investor sentiment needs to be measured. However, "Investor sentiment is not straightforward to measure […]" (Baker & Wurgler, 2007, p.135). Basically, there are three types of measures for investor sentiment: (1) surveys of investors, (2) numerical market data proxies, and (3) textual sources such as news and web documents. In the empirical financial literature, often survey-based or market data-based measures of investor sentiment are used (e.g., (Baker & Wurgler, 2007; Brown & Cliff, 2004, 2005)). Because this thesis focuses on textual measures of investor sentiment, which are discussed in detail in Section 2.5, this section provides only a brief overview of the *types* of measures and their principal shortcomings and advantages. Thus, only some exemplary measures for each type of measure are provided.

Survey-based Measures of Investor Sentiment

Survey-based measures of investor sentiment represent a direct form of measure. A well-known survey of investor sentiment in the U.S. is conducted by the American Association of Individual Investors, which measures investor sentiment by polling its members weekly about whether they are bullish, bearish, or neutral regarding the stock market over the next six months (American Association of Individual Investors, 2014). Survey-based measures are typically available at weekly or monthly frequency and relate to the overall U.S. stock market. Survey data of higher frequency that relates to individual stocks or other financial instruments seems to be infeasible because of the high costs involved. The variable costs of

a survey not only limit the polling frequency but also the number of people polled. Thus, each survey is based only on a limited number of participants. As with every manual survey of humans, this comes with many problems such as measurement errors (arising from interviewers, respondents, the questionnaire, the way the data is collected), sampling errors, and people to be interviewed not responding (Groves, 2004). Thus, the quality of survey-based measures of investor sentiment could be impaired and could be also time-varying.

Market Data-based Measures of Investor Sentiment

Market data-based measures of investor sentiment are derived indirectly from (1) market performance, (2) type of trading activity, (3) derivatives variables, and (4) other measures (Brown & Cliff, 2004, pp.7–13). A large number of possible proxies for measuring investor sentiment based on various market variables have been proposed. For instance, Lee et al. use monthly discounts on closed-end funds to construct a value-weighted-index of these discounts across a set of funds and use this index as an investor sentiment measure (Lee et al., 1991). Measures based on derivatives variables include put/call-ratios of option trading volume and a ratio of expected volatility to current volatility (Brown & Cliff, 2004, pp.11–12). High trading volumes in puts and high expected volatility would be interpreted as negative investor sentiment (Brown & Cliff, 2004, pp.11–12). Investor sentiment measures derived from market data are available at high frequency (e.g., daily) and at low cost. The main disadvantage of market data-based investor sentiment is that it is derived indirectly, and thus requires a theory to relate the market data to investor sentiment (e.g., (Burghardt, 2010, p.39)). Furthermore, the market data (e.g., transactions, volume) represents information after the fact investors have made a decision to trade. That is, the information that they used in the preceding decision making process is not considered, and thus not represented in the measure of investor sentiment. Such information could be in textual form (e.g., newspapers or web information) and is considered by the next type of measure of investor sentiment.

Textual Measures of Investor Sentiment

Textual measures of investor sentiment are investor sentiments available in natural language text, which are extracted either manually or automatically. Many textual outlets exist that contain investor sentiment, e.g., in the form of opinionated assessments of stocks or the stock market. Textual sources could be, e.g., newsletters, news, or web information. Informed by behavioral finance theory, such assessments can be used to infer a measure of investor sentiment. This measure could well represent the investor sentiment of investors in the early stages of the process of investment decision making, influencing also other investors, and thus has a potential impact on stock prices. However, automatic processing of these texts comes with challenges related to the properties of natural language text, such as ambiguity (e.g., (Das & Chen, 2007)). Note that the state of the art regarding approaches for automatically classifying the sentiment orientation of investor sentiment is provided in Section 2.4. Findings regarding the effect of textual measures of investor sentiment on (abnormal) stock returns are discussed in Section 2.5.

Some commercial services provide measures of investor sentiment from text. One prominent example is Investors intelligence. Investors Intelligence's investor sentiment measure is released weekly and is based on classifications of more than 130 independent market newsletters of professional advisors in the U.S. into the categories bullish, bearish, correction, and neutral, suggesting that these advisors are prone to the same errors like individual investors (Investors Intelligence, 2014a, 2014b). Fisher and Statman use the percentage of bullish newsletters classified by Investors Intelligence in the last week of a month as an investor sentiment measure (Fisher & Statman, 2000, pp.16–17). Brown and Cliff (2005) also use Investors Intelligence's elicitation of the number of bullish and bearish market newsletters to construct a monthly bull-bear spread measure, which they use as investor sentiment measure. Another example of a commercial provider of a textual investor sentiment measure is Thomson Reuters' NewsScope Sentiment Engine, which is created from news and is used in the study of Leinweber & Sisk (2011), see Section 2.5.1. The drawbacks of investor sentiment measures of commercial providers in a scientific context are that the methodology used by commercial providers of textual investor sentiment measures and its accuracy are mostly a black box, and the services are costly.

Some measures are based on human classifications, thus they cannot be scaled to analyze web information in a continuous manner. Regarding blogs, Fotak *manually* classified a set of investment blog documents in long and short recommendations (Fotak, 2007). However, he does not relate this explicitly to investor sentiment.

Beside manually extracted investor sentiment, *automatically* extracted investor sentiment from textual sources exists. For instance, a measure of the fraction of negative words from a dictionary in a text has been proposed and applied to news from the Wall Street Journal and the Dow Jones News Wire service (Tetlock et al., 2008). However, this is also not explicitly related to measuring investor sentiment by the authors.

Web information provides a new textual source of investor sentiment. Automatic extraction potentially provides a vast amount of rather unexplored opinionated information that all users, i.e., investors, can create themselves. Thus, the investor sentiment of a broad range of investors can be collected, representing potentially information used in the investment decision making process of these investors and other investors who might consider it. As an example, messages from stock message boards have been classified automatically into bullish, bearish, or hold using a machine learning classifier (e.g., (Antweiler & Frank, 2004; Das & Chen, 2007)). Regarding blogs, mostly general-topic blogs have been studied. General-topic blogs (e.g., from LiveJournal) have been used to create a positive/negative investor sentiment measure using a dictionary-based approach (e.g., (Zhang, 2008)) or an "anxiety measure" using a machine learning approach (Gilbert & Karahalios, 2010). In contrast to these previous studies, this thesis specifically focuses on documents of *investment* blogs as a source of investor sentiment. To this respect, a measure of negative words in Seekingalpha blog documents has been studied (Chen et al., 2014).

However, the authors did not explicitly relate their measure to the concept of investor sentiment and also did not estimate the accuracy of their measure (see Section 2.5.4).

2.3 Blogs as a Source of Investor Sentiment

This section describes investment blogs and provides arguments for investment blogs as an extensive source of investor sentiment that can have an impact on stock prices, potentially generating abnormal returns. This section also describes the model of investor sentiment in blog documents, which is used in the scope of this thesis for classification and aggregation.

2.3.1 Blog Characteristics

Investment blogs can provide subjective expectations, opinions, and analyses on stocks, decision supporting recommendations for investors, rumors and news on products and services of companies, stock markets, and economic development in general. This thesis follows Schmidt (2007) on defining a blog:

Definition: Blog (Schmidt, 2007): "Weblogs, or 'blogs,' are frequently updated websites where content (text, pictures, sound files, etc.) is posted on a regular basis and displayed in reverse chronological order."

Blogs allow internet users to easily create, publish online, and comment on content (e.g., (Kaplan & Haenlein, 2010) and (Agarwal & Liu, 2008)). Blogs allow "[…] people to express their thoughts, voice their opinions, and share their experiences and ideas." (Agarwal & Liu, 2008, p.18). The author of the content of a blog is termed "blogger". Usually, a blog is maintained by an individual blogger (Kaplan & Haenlein, 2010, p.63). The process of creating and publishing content is termed "blogging". The relevant elements of a blog that are studied in this thesis are described subsequently.

Relevant Elements of a Blog

Each blog entry is termed **blog document** in the scope of this thesis. Synonyms for the term blog document include "blog post" (e.g., (Agarwal & Liu, 2008, p.18)) or "blog article" (e.g., (O'Hare et al., 2009, p.9)). Blog documents have a **title**, a **body**, and a **publication date** (e.g., (Mishne, 2007, p.10), see Figure 5). The body usually consists mostly of text but can also contain content of other types, such as pictures or videos (Kaplan & Haenlein, 2010, p.62; Mishne, 2007, p.9). Only the text of blog documents is studied in this thesis. Because the focus in this thesis is on the text of the blog documents of the original authors, and also for simplicity, comments to blog documents are not studied in this thesis.

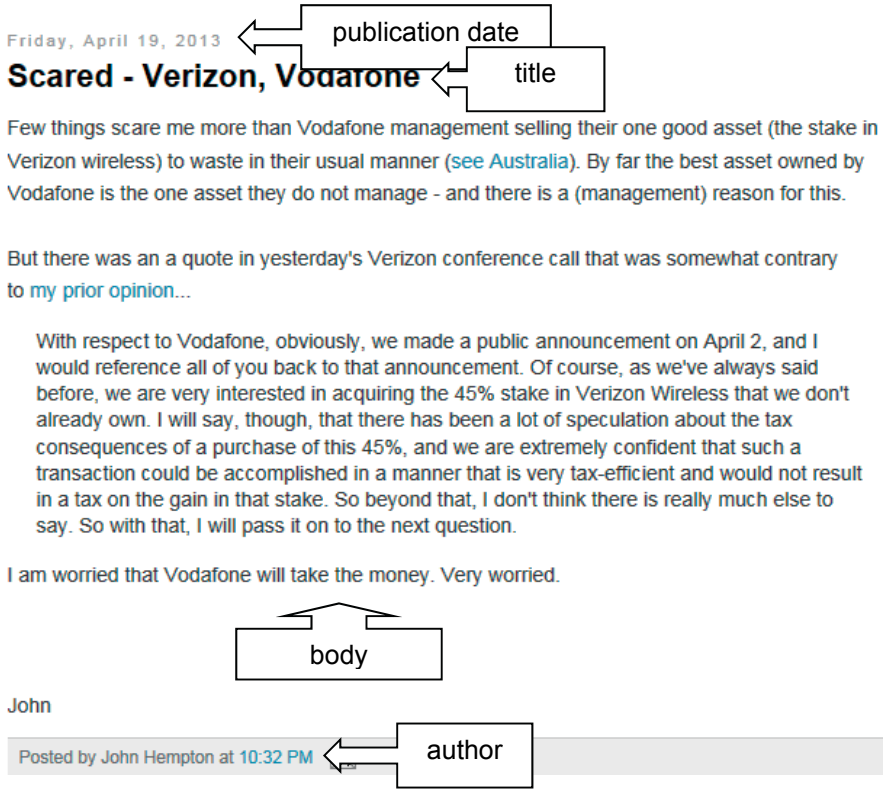


Figure 5: Elements of an example investment blog and blog document. Publication date: 2013-04-19, author: John Hempton, URI: http://brontecapital.blogspot.de/2013/04/scared-verizon-vodafone.html, retrieved 2015-02-04.

Categories of Blogs

Blogs can be categorized along the dimensions (1) personal vs. topics, and (2) individual vs. community ((Krishnamurthy, 2002) cited in (Herring et al., 2004, p.3)). The content of blog documents is "[…] diverse, ranging from journals of daily activities to serious commentaries on important issues." (Nardi et al., 2004, p.46). This thesis focuses on topical blogs that comment on investor sentiment in the form of opinions, and ideas regarding investment topics, i.e., the analysis of stocks and recommendations of buying, selling, or holding these stocks in a portfolio.

Regarding the second dimension of blog categories, blogs are usually maintained by one person (Kaplan & Haenlein, 2010, p.63). This thesis does not differentiate whether these blogs are created by an individual or a community. However, corporate blogs, which are "officially or semi-officially maintained by a company" (Mishne, 2007, p.20), are omitted. A reason is that a corporate blog, e.g., the blog of a company whose stocks are listed on the stock exchange, are likely to be biased with positive investor sentiment (provided by the employees with respect to their company's stock).

Regarding financial blogs, one can differentiate blogs that are (1) integrated in the websites of traditional paper-based media, and (2) independent blogs for financial topics (Hohlfeld & Dörsam, 2008). This thesis focuses on independent blogs as a potential source of investor sentiment of any internet user. Thus, independent blogs should allow for a broader scope of investor sentiment that has been edited and biased by traditional media people to a

smaller degree. Despite being independent from paper-based media, independent blogs can be hosted on a single blog platform.

General Characteristics of Blog Documents

The most often covered topics by independent financial blogs are market reports, indices, analyses of individual financial instruments, and consumer topics (Hohlfeld & Dörsam, 2008, pp.103–104). Thus, the covered topics of independent blogs are found to be similar to professional investment publications and to cover also concrete investment advice (Hohlfeld & Dörsam, 2008, pp.100,108). Therefore, independent financial blogs might be a good source of investor sentiment.

The average length of blog documents of independent financial blogs is 994 characters (Hohlfeld & Dörsam, 2008, p.102). The average length of an investment blog document in this thesis' corpus for evaluating an investor sentiment classifier is 664 words or 3972 characters (see Section 3.1.2 and Table 4). That is, the blog documents can potentially contain a lot of arguments and discussions related to stocks.

Blogs are considered to be more dynamic than webpages (Nardi et al., 2004, p.43). That is, blog documents can be published ad-hoc (Nardi et al., 2004, p.42). The majority of bloggers update their blogs two to three times per week (Technorati, 2011a). However, many professional full time bloggers update their blog several times a day (Technorati, 2011a).

Most blogs are "written by ordinary people" who use it as "a form of personal communication" (Nardi et al., 2004, p.41). The professional background of bloggers is: 61% are hobbyists, 18% are professional part- and full-time bloggers who receive compensation, 8% are corporate bloggers, and 13% are entrepreneurs, blogging for their own company according to a survey by Technorati (Technorati, 2011b). Despite the large amount of hobbyists, more than 50% of all bloggers describe their blogging style as "expert" (Technorati, 2011c).

Information in blogs is spread fast because investors can subscribe to the RSS feed(s) of an arbitrary number of blogs and receive instant notifications regarding the publication of new blog documents. Because most blogs' content is free of charge, a large number of people can potentially make use of it. More than 50% of professional blogs receive more than 10,000 page views per month and more than 30% of professional blogs have more than 10,000 unique visitors per month (Technorati, 2011d).

Number of Relevant Blogs

The number of blogs has rapidly and constantly grown to 173 million in 2011 (NM Incite, 2012). Thus, there is a large and increasing amount of potentially relevant information available for decision makers. However, the number of active blogs that are continuously updated is estimated to be much lower (e.g., (Mishne, 2007, p.22), taking into account information from Technorati and LiveJournal).

This thesis focuses on English-language blogs in the financial domain. 49% of all bloggers are located in the USA and 29% in the EU (Khalid, 2011). Thus, a considerable part of blog documents should be available in English. In 2007, 33% of blog documents were in English (Sifry, 2007).

Considering the financial domain, according a survey (Technorati, 2011c), a little less than 10% of all bloggers write about financial topics (among professional bloggers only, the percentage was slightly above 10%). Considering the number of total blogs, still more than 10 million blogs worldwide presumably write about financial topics.

Investor Sentiment

The authors of investment blog documents might be subject to the behavioral biases discussed in Section 2.2.1, on which the concept of investor sentiment grounds. An indication for this claim is given by the amount of opinions present in blog documents. For instance, Nardi et al. found expressing opinions to be a motivation of bloggers (Nardi et al., 2004, p.43). A general definition of opinions is: "Opinions are usually subjective expressions that describe people's sentiments, appraisals or feelings toward entities, events and their properties." (Liu, 2010, p.627). Similarly to Liu (2010), this thesis regards "opinion" and "sentiment" as synonyms. Regarding investment blogs, authors of a blog document express their opinion on stocks and other financial instruments. O'Hare et al. (2009, p.9) consider financial blogs to be more likely to be opinionated and contain predictive information about stocks than traditional media.

Basically two theories prevail on which an investor sentiment could be based: (1) technical analysis, and (2) fundamental analysis (e.g., (Fama, 1965b, p.55)). Technical analysis assumes future price changes to depend on past price changes (e.g., (Fama, 1965b, p.55)). A branch of technical analysis is trend following, which assumes past trends in price changes to continue (e.g., (Covel, 2004)). Trend following is related to the momentum trading approach (Jegadeesh & Titman, 1993; Moskowitz et al., 2012). Furthermore, charts of prices are being analyzed for price patterns and technical indicators are used to support price prediction (e.g., (Murphy, 1999)). In contrast, fundamental analysis estimates the fair price of a stock (e.g., (Fama, 1965b, p.55) with the fair price termed intrinsic value). Assuming the price to converge to the fair price in the future, fundamental analysis predicts stock prices (e.g., (Fama, 1965b, p.55). For estimating the fair price, e.g., earnings of a company can be used (Fama, 1965a, p.36). For forecasting changes in earnings, financial statement information (e.g., effective tax rate) and macroeconomic variables (e.g., inflation) can be analyzed (Abarbanell & Bushee, 1997).

Whereas opinions formulated in investment blog documents could be based on fundamental or technical analysis outcomes, they can be subjective because the interpretation of fundamental or technical indicators and charts is often subjective and can be subject to behavioral biases. Thus, investment blog documents are regarded to be a source of investor sentiment in this thesis.

Novelty of Information

The novelty of provided information of blogs varies. This thesis focuses on studying blogs that are independent from traditional media, and thus may provide some distinct and novel content. However, bloggers are influenced to a large degree by other blogs, conversations with friends, social media, and also traditional media (Technorati, 2011a). Thus, blogs are likely to contain some redundancy or duplicates. Of the novel information in blogs, at least some seem to be rumors (L. Mitra & G. Mitra 2011, p.4). Rumors potentially contain information before being published as official news. Thus, this thesis considers blogs to be a potentially good and rather novel source for investor sentiment.

Reliability of Information

People generally have high trust in other peoples' opinions as a form of word-of-mouth advertisement (Nielsen Company, 2007). From an advertising perspective, blogs can be regarded a form of consumer generated content (Nielsen Company, 2007). Other consumers have a high level of trust in this type of advertisement (Nielsen Company, 2007). 66% of North Americans (61% worldwide) consider consumer opinions posted on the internet a reliable source of information according to a survey (Nielsen Company, 2007). The perceived reliability of consumer opinions might also refer to investor sentiments in blog documents. Thus, readers of these investor sentiments might also consider them in their own trading decisions.

2.3.2 Blog Platforms

Blogs are structured in communities (Mishne, 2007, p.33). Therefore, a set of blogs was identified for this thesis' study that well represents the investment community. This set of blogs was used to test this thesis' hypotheses regarding investor sentiment from blog documents. Blogs are usually hosted on major blog platforms that allow basically everyone to set up a blog for free and without technical hassle. Such platforms may represent a more homogeneous set of blogs than a random sample of blogs. A technical advantage for retrieving blog documents from blog platforms is the single entry point. The two blog platforms studied in this thesis are discussed in the following.

Seekingalpha

A large platform for investment-specific blogs is Seekingalpha (see Figure 6). The phrase "seeking alpha" presumably refers to Jensen's (1968) alpha. Seekingalpha was founded in 2004 (Seekingalpha, 2014a). Seekingalpha hosts 539,960 blog documents as of 2014-12-03 (Seekingalpha, 2014b) and is "[…] the largest collection of financial blog posts in the world." (McIntyre & Allen, 2009). Seekingalpha has 10,277 contributing authors as of 2014-12-03 (Seekingalpha, 2014c).

Blog documents of Seekingalpha may be a good source of investor sentiment because blog documents are not just news but opinionated content and analysis results of the authors

(Seekingalpha, 2014d, 2014e). Seekingalpha is also regarded to be a source of high quality investor sentiment regarding U.S. stocks in this thesis because publishing documents on Seekingalpha is subject to an editorial process with the following guidelines: (1) high quality opinion and analysis that is convincing, well-presented, and actionable for U.S. investors, (2) market-oriented topics, e.g., on stocks, (3) informed opinions based on rigorous fundamental analysis, (4) original content, (5) compelling title, (6) no stocks with prices <$1, (7) no grammatical or spelling errors, (8) no promotional content, and (9) discretionary selection of documents based on timeliness, uniqueness, and other factors (Seekingalpha, 2014d). Another potential factor for high quality content is that 34% of authors are institutional financial service providers in a 2006 sample (Fotak, 2007, pp.12,13,31). Thus, a large amount of blog documents on Seekingalpha should be on a proficient level.

Regarding the novelty of information in blog documents on Seekingalpha, Fotak found the blog documents to provide some genuine information and no attempt to manipulate markets (Fotak, 2007, p.26). Specifically, he found 41.6% of blog documents to contain novel information in the sense that on days –1, 0, 1 relative to the blog document publication (on day 0) no news or other blog documents have been published (Fotak, 2007, p.13). Fotak hypothesizes this due to the fact that most bloggers are not anonymous and aim to "spread genuine and reliable information" to attract more readers (Fotak, 2007, p.3). However, Fotak found also indications for many blog documents not to contain novel information because almost 45% of blog documents in his 2006 sample are contemporaneous to news from DowJonesNewsWires that feature the same stock and 19% of blog documents are contemporaneous to other blogs documents (Fotak, 2007, pp.14,26).

Many readers of Seekingalpha are investment decision makers or influencers who use it for research, 13.9% of readers are finance professionals, and many readers buy and own stocks ((Seekingalpha, 2014e), with a reference to "Nielsen" for which no details are provided). The number of monthly visitors has been estimated to be more than 2 million in May 2013 (Quantcast, 2013). Furthermore, Seekingalpha is connected with content partners such as Bloomberg, CNBC, and Marketwatch (Seekingalpha, 2014f). Thus, the spread of investor sentiment contained in blog documents from Seekingalpha is large and the investor sentiments potentially have an impact on stock prices. On analyzing the empirical effects, evidence was found that investment blog documents are related to stock price changes that are not reverted in the following days (Fotak, 2007). Fotak found long (expecting a positive future price development) and short (expecting a negative future price development) stock recommendations in blog documents posted on Seekingalpha on average to be accurate (Fotak, 2007, p.22). That is, on average, a long (short) stock recommendation is accompanied by a positive (negative) price reaction (Fotak, 2007, pp.37,38).

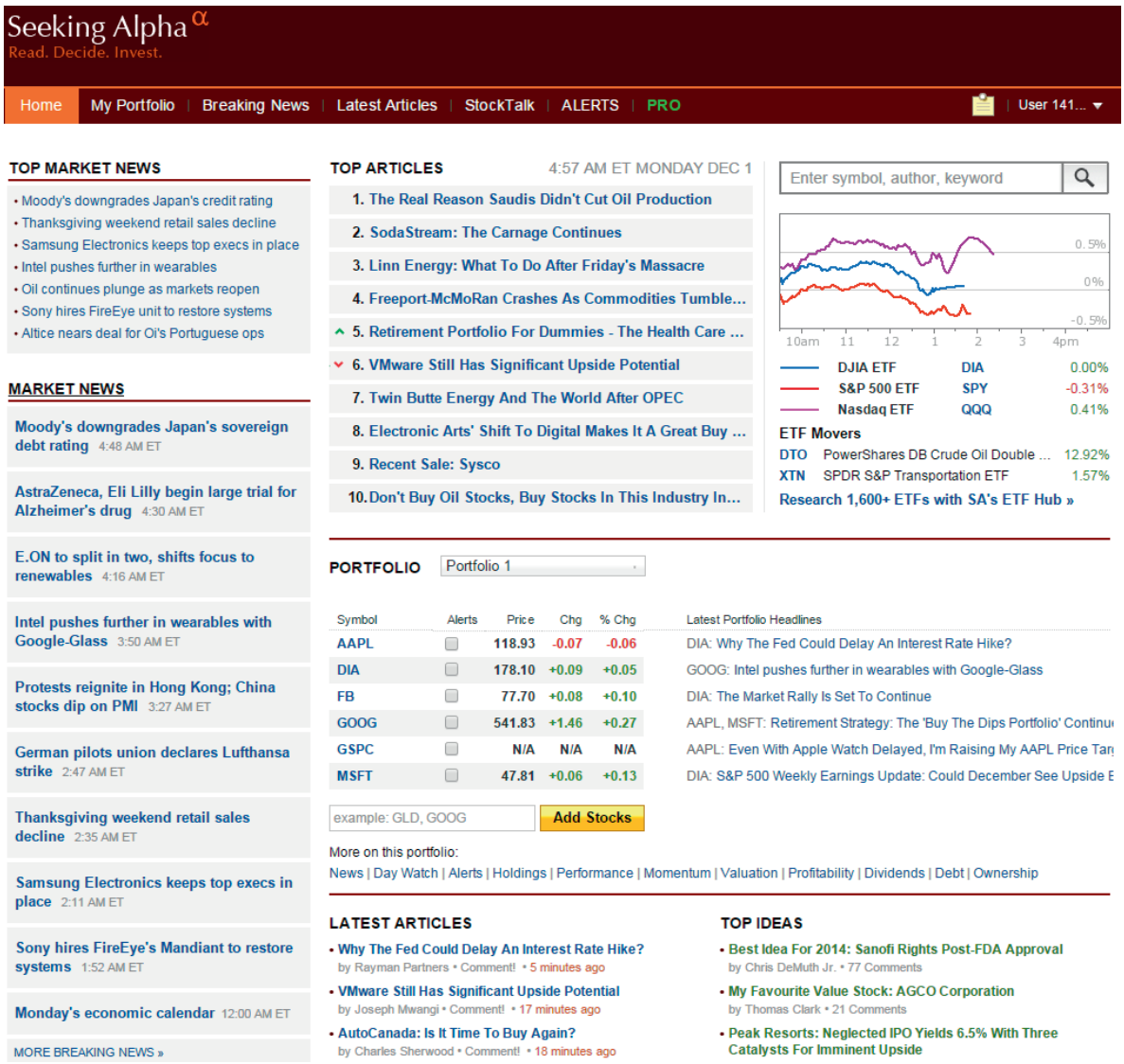


Figure 6: The Seekingalpha investment blog platform website (http://seekingalpha.com, retrieved 2014-12-01).

Blogspot Blog Platform

Many bloggers choose to create a blog with a royalty-free blog platform provider that allows basically anyone to set up a blog. A huge hosting platform for blogs is provided by Google under the URIs http://www.blogspot.com and http://www.blogger.com (linking to the same service as of 2013-07-22 that is provided by Google). More than 30% of bloggers use this blog platform (Technorati, 2011d). In the following, this thesis refers to this blog platform as "Blogspot". The Blogspot blog platform is not specific to investment analysis topics. A blog on Blogspot is provided with a subdomain of the name of the blog. An example of an investment analysis blog hosted on Blogspot is Kevin's Market Blog for which an example blog document is provided in Section 2.3.3.

In comparison to Seekingalpha, blog documents on Blogspot are not constrained regarding the content. That is, there is no editor. Thus, the quality of the blog documents is potentially lower. Furthermore, the addressed financial instruments are neither constrained

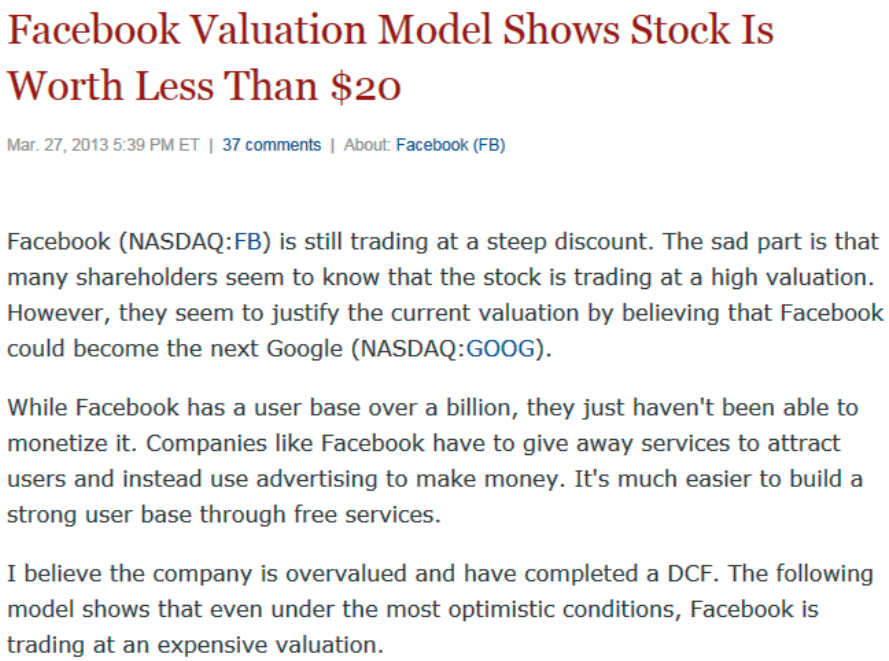
to instruments from the U.S., nor to stocks. Because there is no rule that favors fundamental analysis of stocks, presumably more investment blog documents use technical analysis. The fewer number of constraints and rules may also result in higher heterogeneity of the content. Because of less focus on specific topics and target groups, the number of professional contributors is presumably lower, as well as the potential impact on prices. Finally, the difficulty in retrieving investor sentiment regarding a specific stock might be higher because there are many non-investment topic blogs that might make references to a stock or company in a non-investment context or that might use the typical identifiers of a stock with different meaning.

2.3.3 Examples of Blog Documents

This section provides examples of investment blog documents from the two blog platforms presented. The examples point out the investor sentiment conveyed by these blog documents. The investor sentiment in the examples is based on either fundamental or technical analysis.

Fundamental Analysis Example

Figure 7 provides an example of a blog document from Seekingalpha. The blog document provides an estimation of the fair price of the stock of the Facebook company using fundamental indicators. Assuming the price (i.e., the close price of $26.09 on 2013-03-27 according to Yahoo! Finance) on the date of publication of the blog document to fall to the estimated fair price of $20.18, the author's investor sentiment is negative.



Facebook Valuation Model Shows Stock Is Worth Less Than $20

Mar. 27, 2013 5:39 PM ET | 37 comments | About: Facebook (FB)

Facebook (NASDAQ:FB) is still trading at a steep discount. The sad part is that many shareholders seem to know that the stock is trading at a high valuation. However, they seem to justify the current valuation by believing that Facebook could become the next Google (NASDAQ:GOOG).

While Facebook has a user base over a billion, they just haven't been able to monetize it. Companies like Facebook have to give away services to attract users and instead use advertising to make money. It's much easier to build a strong user base through free services.

I believe the company is overvalued and have completed a DCF. The following model shows that even under the most optimistic conditions, Facebook is trading at an expensive valuation.

Figure 7: Excerpt of an investment blog document based on fundamental analysis. Publication date: 2013-03-27, author: "Kraken", URI: http://seekingalpha.com/article/1305521-facebook-valuation-model-shows-stock-is-worth-less-than-20, retrieved 2015-02-04.

Some opinionated exemplary excerpts of the blog document in Figure 7 are:

> "While Facebook has a user base over a billion, they just haven't been able to monetize it."
>
> "I believe the company is overvalued […]"
>
> "I honestly believe that the stock will decline to $20 over the year and possibly fall even further in 2014, when investors realize speculation has surpassed earnings."

Technical Analysis Example

Figure 8 provides an example of a blog document hosted on Blogspot. The document uses chart analysis and technical indicators for a forecast on U.S. stock prices. The document has a positive investor sentiment regarding the short term future.

Figure 8: Excerpt of an investment blog document based on technical analysis. Publication date: 2009-03-13, author: "Kevin", URI: http://kevinsmarketblog.blogspot.com/2009/03/technical-picture-in-stocks-becoming.html, retrieved 2014-12-01.

Some opinionated exemplary excerpts of the blog document in Figure 8 are:

> "I found this week's price action to be very interesting as the technical picture seems to be favoring the bulls over the short-term."

> "In the above chart, I now have a bullish engulfment pattern COMBINED with a MACD bullish divergence which in my opinion increases the odds of the market moving higher over the short-term."[1]

2.3.4 Model of Investor Sentiment in Blog Documents

This section provides the model of investor sentiment in blog documents. The model serves to quantify and aggregate investor sentiment from individual blog documents to be able to test hypotheses and draw inferences about its effect on abnormal returns of (portfolios of) stocks.

The model of investor sentiment in individual blog documents has been inspired by a generic sentiment model (Liu, 2010), which has been adapted to the field of finance by specifying financial instruments to be the objects of the sentiment (Klein et al., 2011). In this thesis, the focus is on stocks as financial instruments. A positive (negative) investor sentiment orientation expressed by an author of an individual blog document is assumed to refer to an expectation of positive (negative) future abnormal returns of a stock (see Section 2.2.2).

In contrast to Liu (2010), the neutral sentiment orientation is deliberately omitted because of the following reasons (following partially (Klein et al., 2013)): (1) Decreasing the number of sentiment orientations should make the classification problem easier for both humans (i.e., inter-annotator agreement increases) and an automatic classifier (i.e., higher accuracy) ((O'Hare et al., 2009, pp.12,14,15), also cited in (Klein et al., 2013, p.696)). Presumably, this is due to less ambiguity because a decrease in ambiguity of messages of stock message boards was found to increase classification accuracy (Das & Chen, 2007). (2) Regarding financial text classification, some researchers use three classes (e.g., buy/sell/hold or positive/negative/neutral) for classifying the sentiment orientation of a single document (e.g., (Antweiler & Frank, 2004; Das & Chen, 2007; Schumaker et al., 2012)). However, on aggregating these classifications to a measure by which to study relationships or effects on financial market variables (e.g., returns, volatility, trading volume), the hold or neutral classification is often discarded (e.g., (Antweiler & Frank, 2004, p.1267; Das & Chen, 2007, p.1384) – with Antweiler & Frank (2004) also cited in (Klein et al., 2013, p.696)) or the neutral document classification does not provide advantages for investor decision support (e.g., in predicting price direction or in generating higher returns in a trading simulation (Schumaker et al., 2012, pp.17–18), also cited in (Klein et al., 2013, p.696))). For these reasons, only two sentiment orientations are specified in the following definition of investor sentiment in a blog document using a three-tuple:

[1] MACD means moving average convergence/divergence (e.g., (Murphy, 1999, p.252)). The MACD indicator of Gerald Appel is based on the difference of two exponentially smoothed moving averages of different length applied to a stock's price time series (e.g., (Murphy, 1999, pp.252–254)).

Definition: Investor Sentiment Document (inspired by (Klein et al., 2011; Liu, 2010)):

$$sd = (fn, so, t) \tag{2.8}$$

where

$sd \in SD$: Investor sentiment document.

$fn \in FN$: Financial instrument, e.g., a stock, being the object of the investor sentiment document.

$so \in$ {positive, negative}: Sentiment orientation.

t: Publication day of the blog document.

Along with this definition, this work assumes the sentiment orientation of the investor sentiment document to be the same for all stocks that are discussed in the same document. This assumption is corroborated by the analysis of this thesis' corpus of blog documents for the evaluation of this thesis' investor sentiment classifier, which reveals only 7.6% of all blog documents of a 527 set of blog documents to be annotated with diverging sentiment orientations (see Section 3.1.2 and Table 2). Furthermore, all textual parts of a blog document are assumed to be important for determining the investor sentiment document. Thus, all textual parts are included in the analysis.

This thesis follows Das & Chen (2007) on aggregating the investor sentiment document to an investor sentiment index. According to Das and Chen, sentiment is "[…] the net of positive and negative opinion expressed about a stock on its message board." (Das & Chen, 2007, p.1375). For aggregation, the sentiment orientation is represented by a real-valued score:

Definition: Investor Sentiment Document Score (e.g., (Das & Chen, 2007, p.1380; Klein et al., 2011, p.6; Lerman et al., 2009, p.515)):

$$s_{bd,fn,t} = \begin{cases} 1 & if\ so = positive \\ -1 & if\ so = negative \end{cases} \tag{2.9}$$

where

$s \in \{-1,1\}$: Investor sentiment document score.

$fn \in FN$: Financial instrument.

$bd \in BD$: Blog document, i.e., the textual part of it.

$so \in$ {positive, negative}: Sentiment orientation.

t: Publication day of the blog document.

Next, the investor sentiment index is defined as a linear combination of all investor sentiment document score s of all blog documents bd that refer to a given financial instrument fn from a given time period t – which is then normalized to the range [–1,1]:

Definition: Investor Sentiment Index (e.g., (Antweiler & Frank, 2004, p.1266; Klein et al., 2011, p.6; Kumar & Lee, 2006, p.2458; Lerman et al., 2009, p.516)):

$$si_{fn,t} = \frac{\sum_{bd \epsilon BD_{fn,t}} s_{bd,fn,t}}{\sum_{bd \epsilon BD_{fn,t}} |s_{bd,fn,t}|} \tag{2.10}$$

where

$si \in [-1,1]$: Investor sentiment index.

$bd \in BD_{fn,t}$: Set BD of blog documents bd that refer to the financial instrument fn and were published in the time period t.

$s_{bd,fn,t} \in \{-1,1\}$: Investor sentiment document score from a blog document bd that was published in the time period t and refers to the financial instrument fn.

$fn \in FN$: Financial instrument.

t: Discrete time period.

Equal weights are assigned to all sources and authors of different blog documents, assuming each blog document to be equally important. The investor sentiment index expresses direction and intensity of the aggregate sentiment orientation (inspired by (Lerman et al., 2009; Turney & Littman, 2003)): The direction is positive (negative) for $si>0$ (for $si \leq 0$) and the intensity is defined as $|si|$.

Whereas the investor sentiment index refers to an individual financial instrument, a higher aggregate could aggregate all investor sentiment document scores of all stocks of a stock market index. For the stock market level the following index of investor sentiment is defined in the same way like the above index for individual financial instruments:

Definition: Investor Sentiment Index Market (similarly to, e.g., (Antweiler & Frank, 2004, p.1266; Klein et al., 2011, p.6; Kumar & Lee, 2006, p.2458; Lerman et al., 2009, p.516)):

$$sim_t = \frac{\sum_{fn \in FN} \sum_{bd \epsilon BD_{fn,t}} s_{bd,fn,t}}{\sum_{fn \in FN} \sum_{bd \epsilon BD_{fn,t}} |s_{bd,fn,t}|} \tag{2.11}$$

where

$sim \in [-1,1]$: Investor sentiment index market.

$bd \in BD_{fn,t}$: Set BD of blog documents bd that refer to the financial instrument fn and were published in the time period t.

$s_{bd,fn,t} \in \{-1,1\}$: Investor sentiment document score from a blog document bd that was published in the time period t and refers to the financial instrument fn.

$fn \in FN$: Set FN of all financial instruments fn that make up the index that proxies the "market", e.g., the stock market.

t: Discrete time period.

That is, the investor sentiment index market is a bottom-up aggregate (i.e., a micro average) of all investor sentiment document scores from all blog documents referring to all financial instruments of a market index. For simplicity, equal weights are assumed for each financial instrument's investor sentiment document scores.

2.4 Approaches for Textual Investor Sentiment Classification

This section reviews and discusses the state of the art of approaches for accurately classifying the positive/negative sentiment orientation of investor sentiment in blog documents. The focus is on approaches for document level classification of the text of English language investment blog documents. The review differentiates the following types of approaches: (1) dictionary-based approaches, (2) knowledge-based approaches, and (3) supervised machine learning approaches.

2.4.1 Dictionary-based Approaches

A class of approaches for classifying the (sentiment orientation of) investor sentiment in the text of documents is based on dictionaries. Such dictionaries typically comprise sets of positive and negative words (e.g., (Loughran & McDonald, 2011)). A simple approach for positive/negative sentiment classification basically determines whether the number of positive or negative words is higher (used by, e.g., (Hu & Liu, 2004) on the sentence level, also taking into account negations). Such kind of approaches are widespread also for document level sentiment classification (Missen et al., 2013). Regarding investor sentiment classification at the document level, the simple approach of counting the net of positive and negative words and taking negations into account is also used as a baseline (e.g., (Das & Chen, 2007)).

Sentiment Dictionaries

Already existing and popular dictionaries of words with classified sentiment orientation include the following ones:

(1) **General Inquirer's Harvard-IV-4 dictionary** contains English word lists for categories positive and negative (General Inquirer, 2014). This dictionary has been created manually, is not domain specific (Missen et al., 2013), and dates back to ((Stone et al., 1966) cited in, e.g., (Tang et al., 2009, p.10769)). The dictionary has been used widely to classify sentiment (Missen et al., 2013), also in the financial domain (e.g., (Tetlock, 2007)).

(2) **SentiWordNet** is a sentiment word dictionary for sentiment classification extending on WordNet (Baccianella et al., 2010; Esuli & Sebastiani, 2006; SentiWordNet, 2014). WordNet is a large English language lexical resource organized in sets of synonyms that are linked by semantic and lexical relations (Fellbaum (ed.), 1998; WordNet, 2014). SentiWordNet version 1.0 comprises annotations of all sets of synonyms in WordNet

with continuous numerical scores representing positivity, negativity, and objectivity (Esuli & Sebastiani, 2006). A word can be classified positive and negative at the same time, both to a certain degree (Esuli & Sebastiani, 2006, p.418). Thus, obtaining a crisp positive or negative classification on the document level requires setting thresholds for the scores. SentiWordNet has been created automatically using machine learning word classifiers based on training sets that were expanded from manually created seed lists of positive and negative words using synonym and antonym relations of WordNet (supposedly yielding words with the same/opposite orientation) (Esuli & Sebastiani, 2006, pp.418–419). The newer SentiWordNet version 3.0 is based on a newer WordNet and has higher accuracy (Baccianella et al., 2010). SentiWordNet is popular and used by many researchers (Baccianella et al., 2010, p.2200; Tsytsarau & Palpanas, 2012, p.486).

(3) Other non-domain-specific sentiment dictionaries that are also based on WordNet include **fuzzy sentiment** annotations of WordNet words (Andreevskaia & Bergler, 2006).

(4) **Loughran and McDonald**'s sentiment dictionaries of positive and negative words are finance specific (Loughran & McDonald, 2011, 2014). They argue that their dictionaries capture the tone in financial texts and business terminology better than the Harvard-IV-4 dictionary (Loughran & McDonald, 2011). However, Loughran and McDonald focus on official corporate financial reports and not on opinionated web documents.

Investor Sentiment Classification in Blog Documents

Several approaches that use these sentiment dictionaries for classifying the sentiment orientation of investor sentiment in the financial domain exist and are discussed in Section 2.5. However, these approaches (e.g., (Tetlock, 2007; Tetlock et al., 2008)) typically do not address investment blog documents. Furthermore, to be able to compare the accuracy of different approaches for the classification of the sentiment orientation of investor sentiment in investment blog documents, they need to be applied to the same corpus of investment blog documents. To obtain a baseline accuracy of the dictionary-based approach, the General Inquirer Harvard-IV-4 dictionaries (described above) of positive and negative words were used. Using the dictionary, the number of positive words p and the number of negative words n were identified in each blog document of Corpus A (described in Section 3.1.2). The investor sentiment document score s is obtained by $s = \frac{p-n}{p+n}$ (derived from formula (2.10)), where $s \in [-1,1]$. The sentiment orientation of the investor sentiment on the document level is positive if $s>0$, otherwise it is negative. Using 10-fold cross-validation (described in Section 3.3.1), the accuracy is 63.2%. Despite the fact that investor sentiment classification is a hard problem (e.g., (Das & Chen, 2007)), this accuracy is rather low.

Assessment with Respect to this Work

Simple dictionary-based approaches typically use existing sentiment dictionaries for classifying the sentiment orientation on the word level. Thus, they are straightforward to apply in their simplest form. A potential problem with these dictionaries is that they are not

context-specific and often not domain-specific (e.g., (Tsytsarau & Palpanas, 2012, p.487)). That is, they do not take into account influencers of the sentiment orientation such as negations, various senses of a word, intensifiers or diminishers, and the topic or domain ((Wilson, 2008) cited in (Missen et al., 2013)).

To achieve higher context-specificity and domain-specificity, approaches exist for constructing a sentiment (or other) dictionary specific to a corpus (e.g. (Tsytsarau & Palpanas, 2012, p.487)). Such approaches have been reviewed for instance in (Tang et al., 2009, pp.10763–10765; Tsytsarau & Palpanas, 2012, pp.487–489). The basic idea is, starting with some seed words with invariant and known polarity, to infer the context-dependent and domain-dependent polarities of these words from statistical dependencies and co-occurrence relationships with other words in a large corpus (e.g., (Missen et al., 2013; Tang et al., 2009, pp.10763–10765; Tsytsarau & Palpanas, 2012, pp.487–489)). For instance, conjunctions such as "and" or "but" can be used to infer the orientation of adjectives (Hatzivassiloglou & McKeown, 1997). However, in the financial domain, typically simpler dictionary-based approaches for investor sentiment analysis prevail that do not automatically construct corpus-specific dictionaries (e.g., (Das & Chen, 2007; Tetlock, 2007)).

Another, potentially sophisticated, approach to obtain context-specificity, topic-specificity or domain-specificity is using patterns or rules (e.g., (Yi et al., 2003)). These rules represent formalized linguistic and/or domain knowledge. Thus, such approaches are termed "knowledge-based" in the following. The next section reports on such approaches.

Finally, the accuracy of a simple dictionary-based or lexical approach for (document level) sentiment classification is rather low and to be surpassed by a machine learning approach (e.g., (Pang et al., 2002)). An inherent reason might be that sentiment in general can be expressed in a subtle way and is hard to identify or classify by single words (Pang & Lee, 2008, p.19; Pang et al., 2002, p.79). Superiority of machine learning was also found for the problem of document level *investor* sentiment classification in the financial domain (Das & Chen, 2007). For a discussion of the machine learning-based approach, see Section 2.4.3.

2.4.2 Knowledge-based Approaches

This section presents and discusses some principle knowledge-based approaches for the classification of sentiment in general or investor sentiment. Typically, these approaches classify (investor) sentiment on the sub-document or sub-sentence level, which can be aggregated to the document level. Knowledge-based approaches typically build on sentiment dictionaries and use formalized knowledge (in the form of, e.g., ontologies or rules) to obtain (1) a higher specificity of the sentiment classification with respect to a domain or topic, and (2) a higher accuracy compared to simple dictionary-based approaches.

General Sentiment Classification Approaches

The approaches reviewed here are not specific to the financial domain. However, they well represent the principal approach to knowledge-based sentiment classification. They all use

linguistic knowledge in some form of rules (although some authors term them differently). One approach reviewed even uses formalized common sense knowledge. Furthermore, all approaches use information obtained by a parser about the structure of sentences and word dependencies. Thus, the sentiment analysis can be fine-grained and can differentiate topics and even features of a topic to which a sentiment can refer to.

Yi et al. propose an approach for extracting sentiment regarding specific features of a topic on the phrase level (Yi et al., 2003). For extracting these fine-grained sentiments, they used about 120 pre-defined extraction "patterns" that relate a sentiment to a target (Yi et al., 2003, pp.431–432). Sentiment orientations of words were obtained from General Inquirer and another dictionary, which were enriched with synonyms from WordNet (Yi et al., 2003, p.431). The association of the sentiment to a target exploits grammatical structures of a sentence with the help of a syntactic parser and also takes negations into account (Yi et al., 2003). Targets in the paper are, e.g., features of a photo camera such as flash capabilities, the zoom, or the picture quality (Yi et al., 2003, pp.431–432). The approach was evaluated on a camera/music product review corpus (85.6% accuracy) and a general web document and news corpus (up to 93% accuracy) (Yi et al., 2003, pp.430–433). Whereas the accuracy appears to be high, the details of the content, sources of the content, size of the web document corpus, and quality of the manual labelling remain unclear, and both corpora seem to be not publicly available. Regarding this work, it is of special interest that the approach of Yi et al. was not fitted to the financial domain and was not evaluated on a respective corpus. Furthermore, Yi et al. do not enumerate all their sentiment extraction patterns. Thus, their approach is not directly replicable.

Nasukawa and Yi introduced a fine-grained positive/negative/neutral sentiment classification approach regarding a subject on the sub-document level (Nasukawa & Yi, 2003). Their approach relates sentiment terms from a manually created dictionary to a subject (e.g., an organization or a consumer product) in a window of up to 50 words before and after its mention and takes negations into account (Nasukawa & Yi, 2003, pp.72–73). The relationship is identified by local phrase dependencies (i.e., subjects, objects) with the help of a syntactic parser (Nasukawa & Yi, 2003, pp.72–73). Evaluating the approach regarding sentiment classification of camera reviews in web documents, the approach achieved 94.5% accuracy, when counting only the cases found (Nasukawa & Yi, 2003, p.74). However, the approach only extracted 255 of 2000 cases (Nasukawa & Yi, 2003, p.74). In the light of this number, the accuracy does not look that good. Furthermore, regarding long and complex sentences in news, the precision decreases (Nasukawa & Yi, 2003, p.76). For new domains, the sentiment dictionary must be adapted (Nasukawa & Yi, 2003, pp.76–77), but an adaptation to the financial domain seems to be not available. Regarding replicability, the paper only provides some examples of the dictionary and is thus not directly replicable.

Shaikh et al. present an approach for context-specific sentiment classification on the sentence level using rules (Shaikh et al., 2008). The approach comprises the following steps: (1) parse a document and analyze dependencies of words to basically identify subject-verb-

object triplets and dependencies to other triplets (Shaikh et al., 2008, pp.559,564-565,573,576), (2) assign a context-free sentiment score to known words from a manually crafted dictionary (Shaikh et al., 2008, pp.566–568) and to unknown words using (a) relations in WordNet (Shaikh et al., 2008, pp.566–568), (b) semantic relationships retrieved from the ConceptNet common sense ontology ((Liu & Singh, 2004) cited in (Shaikh et al., 2008, p.565)) to concepts of which the representing words were already scored (Shaikh et al., 2008, pp.568–570), and (c) web search engine results (Shaikh et al., 2008, pp.570–572), (3) obtain context-specific sentiment scores for each element of a triplet using rules that take into account multiple word senses (Shaikh et al., 2008, pp.572–575), (4) assess the sentiment score of each triplet in a sentence, taking into account negations and conditionality (Shaikh et al., 2008, pp.572-573-576), and (5) assess the sentiment score for a whole sentence using rules based on dependencies of the triplets (Shaikh et al., 2008, pp.576–578).

The approach of Shaikh et al. (2008) was evaluated on both, the sentence level and the document level. The document level sentiment classification was obtained by averaging the sentence level scores (Shaikh et al., 2008, p.581). The document level evaluation used product reviews and the movie review sentiment corpus by Pang and Lee ((Pang & Lee, 2004) cited in (Shaikh et al., 2008, pp.581–582)). On this corpus, the accuracy was 85.5%, which is better than the accuracy of many machine learning approaches with the exception of one SVM-variant (Shaikh et al., 2008, pp.585,590). For a description of machine learning and SVM, see Section 2.4.3.

Concluding, the approach of Shaikh et al. (2008) achieves high accuracies that are comparable to the ones of machine learning approaches (Shaikh et al., 2008, pp.585,590) despite some rules being naïve and simple heuristics (Shaikh et al., 2008, p.575). However, the approach has not been evaluated with respect to *investor* sentiment classification. Its adaptation to this domain might be hindered by the level of sophistication of the approach (e.g., (Shaikh et al., 2008, p.559)). Regarding replicability of the approach and evaluation results, the approach was evaluated using a publicly available corpus of Pang and Lee (2004), pseudo-code is provided for the contextual sentiment assessment (Shaikh et al., 2008, pp.597–601), and many rules are provided throughout the paper. However, apparently, not all rules were published because the authors only provide "[…] some example rules to compute contextual valence […]" (Shaikh et al., 2008, p.573). Furthermore, the dictionary of manually classified words seems to be not publicly available. Thus, the approach would not be fully replicable.

Investor Sentiment Classification in Blog Documents

Several knowledge-based approaches have been proposed for sentiment classification but not adapted to the financial domain for the classification of investor sentiment. That is, the rules or patterns need to be adapted to the domain and specific financial knowledge needs to be formalized and incorporated into the sentiment classification approach. Replication of published knowledge-based approaches is not straightforward because often some of the

following pieces are not fully publicly available: the rules, extraction patterns, or dictionaries. However, a knowledge-based approach for *investor* sentiment classification in blogs that can be considered at least a baseline has been proposed by (Klein et al., 2011). To be able to compare such a knowledge-based approach for investor sentiment classification with other approaches regarding the classification of investment blog documents, Corpus A (described in Section 3.1.2) is used for evaluation.

The approach proposed by Klein et al. classifies the sentiment orientation of investor sentiment referring to the "expected future price change" (Klein et al., 2011, p.2) of a stock on the sentence level in positive/negative (Klein et al., 2011). To start with, a dictionary of words was used to determine the negative/positive sentiment orientation of single words (Klein et al., 2011, pp.3–4). An information extraction ontology was used to extract stocks and financial indicators, which allow to indirectly infer a sentiment (Klein et al., 2011, pp.3–5). Using these indicators, a sentiment about a stock's expected future price change was inferred by an indicators correlation (positive or negative) to the stock's expected future price change (Klein et al., 2011, pp.3–5). The classification approach uses a set of manually-crafted rules that relate the sentiment orientation of a word to a stock directly or indirectly by means of the indicators and takes negations into account (Klein et al., 2011, p.5).

For evaluation of the approach by Klein et al. (2011), the General Inquirer Harvard-IV-4 dictionary of positive and negative words was used like in the dictionary-based approach that was evaluated on the same corpus in Section 2.4.1. The sentence level sentiments, referring to the same stock, were aggregated to the document level in the same way like in the application of a baseline dictionary-based approach to investor sentiment classification (see Section 2.4.1). Using 10-fold cross-validation (described in Section 3.3.1) on Corpus A (described in Section 3.1.2), the accuracy was 66.2%.

The accuracy is a bit higher than the accuracy of the baseline dictionary-based approach evaluated on the same corpus. Clearly, the higher accuracies of the reviewed knowledge-based approaches, evaluated on other corpora, were not achieved. Possible reasons might be the use of a different corpus and a higher sophistication of the reviewed approaches and their rules using deeper parsing, which determines subject-verb-object triplets and dependencies among them (e.g., in (Shaikh et al., 2008)). Klein et al. (2011) only use shallow parsing, which does not determine triplets and a dependency tree. Such a deep parse would be computationally expensive and is time-consuming. For instance, parsing the example blog document presented in Figure 7 with the Stanford Parser available online[2] took 10.5s with the text containing 494 tokens. Because in the financial domain a large amount of blog and other web documents become available constantly and because timely investment decisions matter, this seems not desirable.

[2] http://nlp.stanford.edu:8080/parser/index.jsp, applied on 2014-03-05; the online deployment's last update was from 2012-07-10

Assessment with Respect to this Work

Knowledge-based approaches allow for high specificity regarding a language, topic, feature of a topic, or a domain. Using formalized common sense knowledge the covered topics can be broad. The analysis of sentiments is typically fine-grained on the sub-sentence level. However, the sentiments can be aggregated easily to the document level. The classification models can be considered "glass box" because the knowledge for classification is modeled by rules or an ontology, which are directly accessible for human understanding. Regarding the accuracy of knowledge-based approaches, accuracies reported in non-financial-domains are high.

A major problem for adapting a knowledge-based approach for investor sentiment classification is the fact that often important pieces for replication are not fully published or are not available publicly (e.g., full set of rules, dictionary, and the corpus). A reason might be that some of the authors work or have worked in research laboratories of commercial organizations. Thus, at least some of the pieces for a knowledge-based approach to investor sentiment classification in blogs would have to be created from scratch. Sophisticated approaches come with high manual effort for creating rules, suitable dictionaries, and knowledge models. Furthermore, sophisticated knowledge-based approaches require structural, grammatical, and dependency information obtained at high computational cost by deep parsing.

The baseline accuracy of the (rather shallow) approach by Klein et al. with respect to investor sentiment classification was higher than for the baseline dictionary-based approach evaluated on the corpus of investment blog documents used in this work. However, the accuracy was still rather low. Thus, it seems more fruitful in terms of accuracy (in relation to manual effort required) to pursue a supervised machine learning approach. The next section discusses respective approaches. In terms of manual effort they still require a labeled corpus for training. However, a corpus would be also required for evaluating any other type of approach. Furthermore, supervised machine learning approaches are computationally efficient (e.g., (Joachims, 2006)).

2.4.3 Supervised Machine Learning Approaches

This section first discusses the vector space model, which is used to represent blog documents for machine learning text classification. Furthermore, the concept of supervised machine learning and two concrete machine learning approaches are discussed. Finally, the approaches are assessed with respect to this work's design of a classifier of the sentiment orientation of investor sentiment in blog documents.

2.4.3.1 Vector Space Model

For classifying texts with a machine learning approach a numerical vector is usually used as representation (Sebastiani, 2002, p.10). The underlying **vector space model** was invented by Gerard Salton and developed and published in several articles and books, e.g., (Salton, 1979,

1989; Salton et al., 1975). The vector space model builds on the core assumption that a document can be identified by a set of terms (Salton, 1989, p.313; Salton et al., 1975, p.613). Terms are made up by words usually (Sebastiani, 2002, p.10) or other basic text parts (e.g., (Joachims, 2002, pp.12–16)). In machine learning document text classification they are interchangeably called "features" (e.g., (Sebastiani, 2002, p.10)) – which is adopted in the following. The term "feature" is also used when citing literature that uses the term "term". The exact definition of "features" for this thesis is provided in Section 3.2.1.1.

A **set of features** *FT* is defined in this work to comprise all *distinct* features *ft* that are contained in a set *BD* of textual blog documents *bd*. The textual part of a blog document can now be represented by a document vector. Each coefficient of the vector (vectors are denoted in bold) represents a numerical weight that indicates the presence or importance of a feature (or term) that is used to represent the document (Salton, 1989, p.314; Salton et al., 1975, p.613), see the following definition.

Definition: Document Vector (adapted from (Salton et al., 1975, p.613)):

$$\boldsymbol{d} = (w_1, w_2, \dots, w_n) \tag{2.12}$$

where

$\boldsymbol{d} \in \mathbb{R}^n$: Document vector.

$w_i \in \mathbb{R}$: Weight of feature ft_i in the blog document.

$n \in \{0,1,2,3,\dots\}$: Number of dimensions of the vector.

Regarding the **vector space**, to which the vector space model refers to, all documents of a set of documents and all features (or terms) can be represented as vectors in the same vector space (Raghavan & Wong, 1986, p.280). This thesis defines the **feature vectors** to represent the **basis vectors** of the vector space. Thus, all feature vectors are orthogonal per definition, not requiring any assumptions (Dubin, 2004, p.757). The number of dimensions of the vector space is given by the size of the set of features (following a common assumption, see (Salton & McGill, 1983), cited in (Raghavan & Wong, 1986, p.280)). Based on the stated definitions and assumptions, each document vector can be represented by a linear combination of the feature vectors (Salton, 1989, p.314). The coefficients in this linear combination are assumed to be the components of the document vector along the respective feature vectors (Salton, 1989, p.314). That is, the coefficients are the weights associated with respective features.

Per definition, orthogonal feature vectors are linearly independent. Thus, orthogonality means to ignore dependencies (Dubin, 2004, p.757). For instance, the relative order of features (i.e., terms) is not represented (Manning et al., 2009, p.121). However, orthogonal feature vectors are regarded as an acceptable approximation (Raghavan & Wong, 1986, p.280). Also, the simplicity of the vector space model is regarded as one of its advantages (Salton, 1989, p.319). Therefore, the vector space model is popular among researchers and practitioners (Baeza-Yates & Ribeiro-Neto, 1999, p.34).

For using the vector space model, at least two questions have to be answered: (1) How to define a feature?, and (2) How to determine the weights? (cf. (Sebastiani, 2002, p.10)). These questions are answered in Section 3.2.1.

2.4.3.2 Supervised Machine Learning

Machine learning refers to computer programs that learn to improve their performance measured in some metric, regarding some task, provided with experience (Mitchell, 1997, p.2). In this thesis, the task is to classify the sentiment orientation of investor sentiment in the text of a (blog) document referring to a stock. "Experience" is provided by examples of pairs of (blog) documents and the respective correct sentiment orientation in supervised machine learning. The performance is measured in terms of accuracy, i.e., the percentage of correct classifications (e.g., (Yang, 1999, p.75)).

For supervised machine learning from examples, the following components are required (adapted from (Vapnik, 2000, p.17)):

(1) A "generator" of (input) vectors (Vapnik, 2000, p.17). Regarding textual investor sentiment classification (of blog documents), each investment text (blog) document $bd \in BD$ is transformed to a document vector $\boldsymbol{d} \in D$ (see Section 2.4.3.1) using a **document vector transformation** function $df: BD \rightarrow D$. The transformation encompasses several steps (see Section 3.2.1).

(2) A "supervisor" who provides an output value for every input vector (Vapnik, 2000, p.17). Regarding the classification of the sentiment orientation of investor sentiment in textual (blog) documents, the investor sentiment document score $s \in \{-1,1\}$ was defined (see Definition (2.9)) as output value. To this respect, the supervisor is an (typically human) **annotator** of the (blog) documents that has to read and interpret the text of the document and provide either –1 (i.e., negative) or 1 (i.e., positive) as annotations with respect to the sentiment orientation of an investor sentiment regarding to a stock or other financial instrument identified in the document.

(3) A "**learning machine**" that implements a set of functions $f(\boldsymbol{d},pa)$, where $pa \in PA$ with PA being a set of (yet abstract) parameters, which could be scalars or vectors (Vapnik, 2000, p.17). The parameters pa of the learning machine are specific to the learning machine algorithm. Regarding textual investor sentiment classification, the function f is termed **classifier** (e.g., (Sebastiani, 2002, pp.2–3)) and outputs an investor sentiment document score $\hat{s} \in \{-1,1\}$.

The learning problem for classifying the sentiment orientation is to choose a function f that best approximates the output of the supervisor (Vapnik, 2000, p.17). A **training set** of l pairs $(\boldsymbol{d}_i,s_i)$ is used for choosing f: $(\boldsymbol{d}_1,s_1),\ldots,(\boldsymbol{d}_l,s_l)$ (Vapnik, 2000, p.18). This training set has to be provided by the annotator and provides an investor sentiment document score output s for each vector representation $\boldsymbol{d}$ of a (blog) document text. In the learning process, the learning machine is provided with all input-output pairs $(\boldsymbol{d}_i,s_i)$ of the training set to choose f,

such that each $\hat{s}=f(\boldsymbol{d},pa)$ comes closest to the respective desired s (adapted from (Vapnik, 2000, p.18), see also Figure 9).

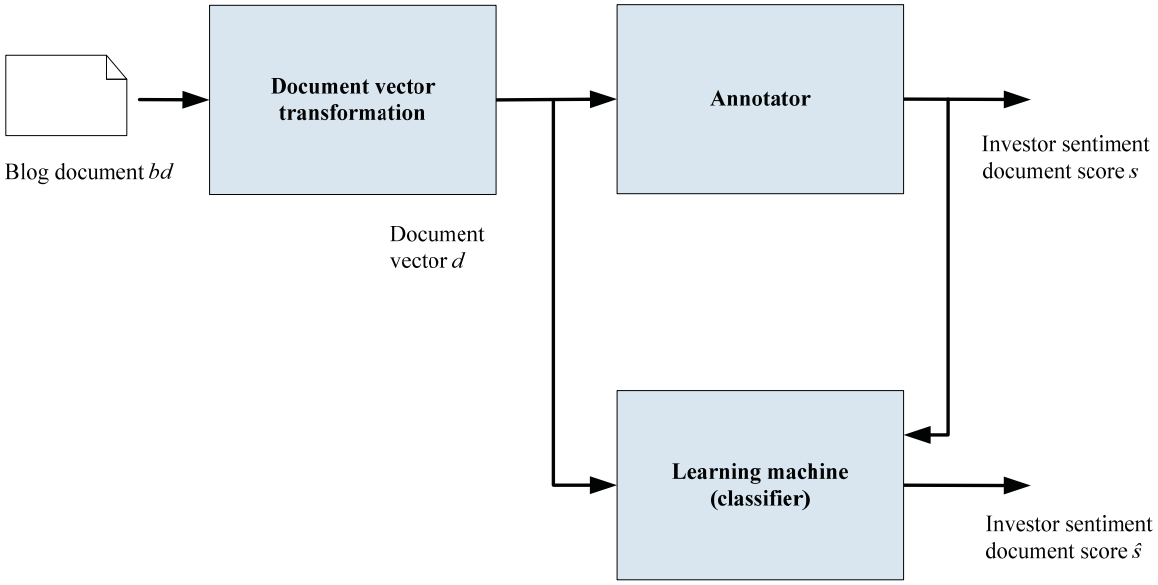


Figure 9: Supervised machine learning of an investor sentiment classifier given training examples provided by an annotator (adapted from (Vapnik, 2000, p.18)).

To guide the learning process, the discrepancy of the classifier's output $\hat{s}$ regarding the annotator's output s can be measured over all training examples by the "empirical risk functional" (adapted from (Vapnik, 2000, pp.18–21)):

Definition: Empirical Risk Functional (adapted from (Vapnik, 2000, pp.18–21)):

$$R_{emp}(pa) = \frac{1}{l}\sum_{i=1}^{l} Lo\big(s_i, f(\boldsymbol{d_i}, pa)\big) \tag{2.13}$$

where

the loss function $Lo\big(s_i, f(\boldsymbol{d}_i, pa)\big) = \begin{cases} 0 & if\ s_i = f(\boldsymbol{d}_i, pa) = \hat{s}_i \\ 1 & if\ s_i \neq f(\boldsymbol{d}_i, pa) \end{cases}$

provides 1 for each classification error (adapted from (Vapnik, 2000, p.19)).

Variables in Definition (2.13) are defined as follows:

$s_i \in \{-1,1\}$: Investor sentiment document score (see Definition (2.9)) of the annotator for a given $\boldsymbol{d}_i$.

$\hat{s}_i \in \{-1,1\}$: Investor sentiment document score output of the classifier function f.

f: Classifier function that is chosen by the learning machine (adapted from (Vapnik, 2000, pp.17–18)).

$\boldsymbol{d}_i$: Document vector, see Definition (2.12).

pa: Parameter (a scalar or vector) of the learning machine (Vapnik, 2000, p.17).

l: Number of pairs $(\boldsymbol{d}_i, s_i)$ that constitute the training examples (adapted from (Vapnik, 2000, p.18)).

The objective of the learning machine is to minimize the empirical risk functional R_{emp} given the parameter (vector) pa of the learning machine and the empirical data of the training examples (Vapnik, 2000, pp.18–21). The document vector transformation also influences the effectiveness of the learning machine in minimizing the empirical risk functional as each (blog) document in the training set has to be transformed to a document vector first.

After learning, the classifier f is supposed to output $\hat{s}$ such that it comes close to the desired s for document vectors (adapted from (Vapnik, 2000, p.18)). To measure the effectiveness of the classifier after learning, the metric accuracy can be used. Formally:

Definition: Accuracy (adapted from (Sebastiani, 2002, pp.33–34; Vapnik, 2000, pp.18–21)):

$$a = \left(1 - R_{emp}(pa)\right) \cdot 100\% \tag{2.14}$$

where

$a \in [0,100]$: Accuracy of a classifier in percent.

R_{emp}: The empirical risk functional as defined above, being equivalent to "error" as defined in (Sebastiani, 2002, pp.33–34).

pa: Parameter (a scalar or vector) of the learning machine (Vapnik, 2000, p.17).

If the number of training examples would tend to infinity, the empirical risk functional for a classifier would converge to the true risk (Schölkopf & Smola, 2002, pp.131–132). However, for real world learning problems, training examples are often scarce. In this case, the empirical risk cannot be expected to converge. Practically, this means that a small error obtained on the training set does not necessarily imply a small error on an independent set (e.g., (Schölkopf & Smola, 2002, p.8)). This set is usually termed **test set** (e.g., (Hastie et al., 2009, p.222)). The value of the accuracy (or error) on the test set using a classifier trained by a learning machine is termed **generalization performance** (e.g., (Hastie et al., 2009, p.219)). A good generalization performance is what machine learning aims at. The generalization performance (precisely: the upper bound on the true risk) depends on the complexity of the (class of) classifier functions (Vapnik, 2000, pp.93–96). The (informal) "complexity" of a

(class of) classifier functions can be measured in terms of the so-called **VC-dimension** (e.g., (Hastie et al., 2009, p.238; Schölkopf & Smola, 2002, p.128)), which is a concept of Vapnik's and Chervonenkis' statistical learning theory (Vapnik & Chervonenkis, 1974; Vapnik, 2000).

Before going into the generalization ability of trained classifier functions more formally, the following intuition is provided. For instance, high-degree polynomial functions should have a higher VC-dimension than linear functions. A highly complex function might well learn a classification problem (or a general function approximation problem) and observe low error (high accuracy) on the training set – but it will generalize badly and observe high error (low accuracy) on the test set (e.g., (Joachims, 2002, pp.35–36)). That is, the classifier has been subject to "overfitting" (e.g., (Joachims, 2002, p.36)).

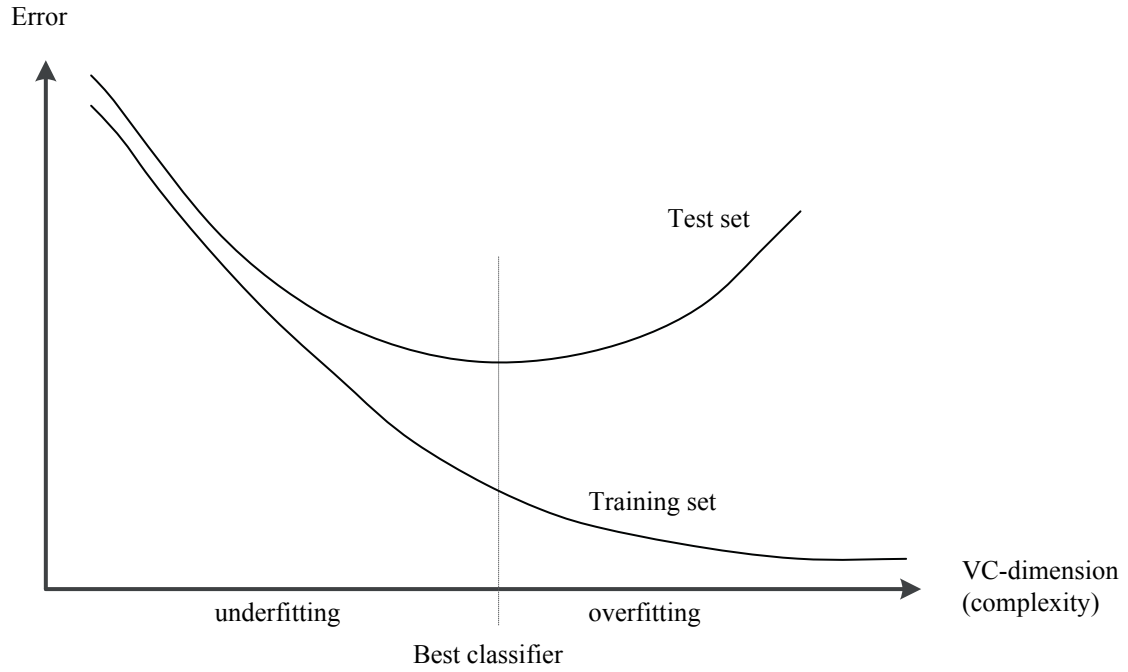


Figure 10: Training and test error as a function of classifier complexity. The best classifier is obtained by restricting the complexity to prevent overfitting to the training set (adapted from (Hastie et al., 2009, p.38; Schölkopf & Smola, 2002, p.139; Vapnik, 2000, p.96).

Figure 10 shows a typical function of the error on the training set and on the test set as a function of the complexity of the classifier. The figure shows that for low complexity of a classifier, the error is high on both, the training set and the test set. To be able to learn the classification function, a certain complexity is required. Thus, on increasing the complexity, the error decreases on the training set and the test set. However, at a certain point the classifier starts to overfit and adapts too strongly to the data presented by means of the examples in the training set. At this point, the error on the test set (i.e., the true error) will increase, as the complexity is increased further. Related to this intuition of the generalization ability of a classifier depending on its complexity, the following upper bound for the true error has been found by Vapnik for "small" sized training sets (e.g., (Vapnik, 1999, pp.993,998, 2000, pp.93,94,123,124)):

Definition: Upper Bound of the True Error (adapted from (Vapnik, 1999, pp.993,998, 2000, pp.93,94,123,124)):

$$R(pa) \leq R_{emp}(pa) + \Omega(h, l, \eta) \tag{2.15}$$

with probability at least 1-η

and where

R: Expected risk functional (Vapnik, 1999, p.989). Also termed true/actual error or actual risk (e.g., (Schölkopf & Smola, 2002, p.132; Vapnik, 2000, pp.94,95)). *R* can be estimated by the error on the test set.

R_{emp}: Empirical risk functional as defined above in (2.13), being equivalent to "error" as defined in (Sebastiani, 2002, pp.33–34) and referring (here) to the (average) error on the training set (Schölkopf & Smola, 2002, p.8).

pa: Parameter (a scalar or vector) of the learning machine (Vapnik, 2000, p.17).

Ω: Confidence interval for the estimate of R_{emp} (Vapnik, 1999, p.998). The confidence interval is a function of h,l,η, see, e.g., (Vapnik, 1999, p.993) for the specification.

h: VC-dimension (Vapnik, 2000, p.93) – informally, it measures "[…] the complexity of a class of functions […]" (Hastie et al., 2009, p.238)).

l: Number of training examples.

The upper bound definition identifies the two factors that control the **generalization ability** of a classifier created by a learning machine: (1) empirical risk (i.e., the error on the training set), and (2) the confidence for the estimate of the empirical risk (e.g., (Vapnik, 1999, p.998)). Essentially, this boils down to the fact that the generalization ability depends on the complexity of a classifier: The first term basically *decreases* with increasing the complexity (Vapnik, 2000, p.95), as Figure 10 also indicates. The second term basically *increases* for a given amount *l* of training examples and a given η, when the VC-dimension (measuring the complexity) is increased (Vapnik, 2000, pp.93,94,123,124).

For exploiting the upper bound of the true error for creating a classifier that generalizes well (i.e., has a low true error), the "**structural risk minimization**" (SRM) inductive principle (Vapnik, 2000, p.94) was introduced ((Vapnik & Chervonenkis, 1974) cited in (Vapnik, 2000, pp.8,94)). It is a key part of the theory for controlling the generalization ability (Vapnik, 2000, pp.93–122). In contrast to empirical risk minimization, which minimizes only R_{emp} on large training sets (Vapnik, 2000, pp.20–21,94), the complexity of the classifier function is taken into account by SRM and is subject to optimization (Vapnik, 2000, pp.93–95). SRM proposes to minimize *both* terms of the upper bound of the true error (e.g., (Vapnik, 1999, p.994)). Because the first term decreases with increasing complexity, and the second term increases with increasing complexity, SRM defines a trade-off between the quality of a classifier function and the complexity of the classifier function (Vapnik, 1999, p.994, 2000, p.95). For the best generalization ability of a classifier, the complexity has to be optimized

(by choosing an appropriate function), such that the *sum* of the two terms of the upper bound of the true error is minimized (Vapnik, 2000, pp.123–124).

Support Vector Machines (SVM) are a text classification approach that implements the structural risk minimization inductive principle (e.g., (Joachims, 2002, p.35)). Support Vector Machines (SVM) and another approach, called Naïve Bayes, are the machine learning approaches best suited for sentiment classification on the document level within one domain according to a literature review (Tang et al., 2009, p.10766). Each of the approaches is discussed below, starting with the simpler one.

2.4.3.3 Naive Bayes

Naive Bayes (NB) is a probabilistic learning approach (Manning et al., 2009, p.258). The basic idea is to find the most probable (i.e., the maximum a posteriori) hypothesis (of the following hypotheses: H1: "the blog document belongs to the class of positive sentiment orientation" and H2: "the blog document belongs to the class of negative sentiment orientation") using Bayes' rule, given knowledge about the prior probabilities of the hypotheses and training examples (Mitchell, 1997, p.156). The terms in the documents of a certain class in the training examples serve as supporting evidence for the hypothesis that the document belongs to the class (e.g., (Manning et al., 2009, pp.258–259)). Typically, a term (or a feature – which is used as a synonym in this thesis) are identified by a single word (Sebastiani, 2002, p.10). This definition of a term or feature is also used in this section.

The NB approach for text classification is based on two different models of a textual document: (1) a multi-variate Bernoulli model, and (2) a multinomial model (e.g., (McCallum & Nigam, 1998)). The NB-approach based on the multinomial model is discussed subsequently as it usually performs better at a lower error rate (McCallum & Nigam, 1998).

In the multinomial model, a document is assumed to be an ordered sequence of words that are drawn independently but from the same multinomial distribution of words, i.e., the vocabulary (e.g., (McCallum & Nigam, 1998)). This implies the naive Bayes assumption of each word occurrence to be independent of other word occurrences given a class' context (e.g., (McCallum & Nigam, 1998)). Furthermore, the conditional probability of a word (or term), given a class, is assumed to be the same for each position in the document (e.g., (Manning et al., 2009, p.267)). Finally, the length of a document, i.e., the number of draws, is assumed to be independent of a class (e.g., (McCallum & Nigam, 1998)).

The NB classifier is based on the probability of a (investor sentiment) class with respect to a textual (blog) document that can be derived by Bayes' rule (e.g., (McCallum & Nigam, 1998; Mitchell, 1997, pp.156–157)):

Definition: Naive Bayes Classifier (adapted from (Joachims, 2002, pp.22–24; Manning et al., 2009, pp.258,265; Mitchell, 1997, p.157)):

$$\hat{s} = \underset{\hat{s}\in\{-1,1\}}{\arg\max}\ p(\hat{s}|bd) = \underset{\hat{s}\in\{-1,1\}}{\arg\max}\ \frac{p(\hat{s})\ p(bd|\hat{s})}{p(bd)} \tag{2.16}$$

where

$\hat{s}\in\{-1,1\}$: Investor sentiment document score (see Definition (2.9)) of the Naïve Bayes classifier. Each possible value of $\hat{s}$ defines a class.

bd: A textual blog document.

p(ŝ): Prior probability of the hypothesis that *bd* belongs to class $\hat{s}$ (e.g., (Manning et al., 2009, p.258; Mitchell, 1997, p.156)).

p(bd): Probability of observing a blog document *bd* (adapted from (Mitchell, 1997, p.156)).

p(bd|ŝ): Conditional probability of observing a blog document *bd* in class $\hat{s}$ (e.g., (Joachims, 2002, p.23)).

p(ŝ|bd): The posterior probability of the hypothesis that *bd* belongs to class $\hat{s}$ (adapted from (Mitchell, 1997, p.156)).

The above classifier function can be rewritten to (adapted from (Manning et al., 2009, pp.258–259)):

$$\hat{s} = \underset{\hat{s}\in\{-1,1\}}{\arg\max}\ log\big(p(\hat{s})\big) + \sum_{i=1}^{|V|} log\big(p(wo_i|\hat{s})\big) \tag{2.17}$$

That is, to classify a document *bd* in a class $\hat{s}\in\{-1,1\}$, the parameters *p(ŝ)* and the conditional probability *p(wo|ŝ)* need to be estimated for all words *wo* from the blog document that are contained in the vocabulary *V* (Manning et al., 2009, pp.258–259). The set of words *V* can be extracted from the blog documents that comprise a training set (Manning et al., 2009, pp.256,260). The parameters of the classifier can be also estimated by the training examples (*bd*,*s*), where *s* is used to distinguish class information provided from annotators from the classifier output $\hat{s}$. First, the prior probability of class *s* occurring is estimated as the relative frequency of the number of documents in the training examples that belong to class *s* (adapted from (Manning et al., 2009, p.259)):

$$p(s) = \frac{l_s}{l} \tag{2.18}$$

Variables in Definition (2.18) are defined as follows:

$s \in \{-1,1\}$: Investor sentiment document score (see Definition (2.9)) of the annotator.

l: Number of all training examples.

l_s: Number of training examples of the class $s \in \{-1,1\}$.

Second, the conditional probability of a word *wo* (representing a term or feature), occurring in a blog document of a given class *s*, can be estimated as the relative word (or term or feature) frequency over all training example blog documents (adapted from (Joachims, 2002, pp.23–24; Manning et al., 2009, pp.259–260; McCallum & Nigam, 1998)):

$$p(wo|s) = \frac{\left(\sum_{bd_s \epsilon BD_{train}} ff(wo, bd_s)\right) + 1}{\left(\sum_{wo' \in V} \sum_{bd'_s \epsilon BD_{train}} ff(wo', bd'_s)\right) + |V|} \tag{2.19}$$

where

$s \in \{-1,1\}$: Investor sentiment document score (see Definition (2.9)) of the annotator.

wo: A word (representing a term or feature).

wo'$\in V$: All words *wo'* of the vocabulary set *V* that are used for classification (e.g., (Manning et al., 2009, pp.258–260)).

$ff(wo,bd_s)$: Word (or term or feature) frequency of the word (or term or feature) *wo* in the textual blog document bd_s classified as class *s* (e.g., (Joachims, 2002, pp.23–24)). Word frequency corresponds to the bag of words model (see Definition (3.2)).

$bd_s \in BD_{train}$: All blog documents of the training set BD_{train} that have been annotated with the class *s* (adapted from (McCallum & Nigam, 1998)).

The formula adds one for each word (or term or feature) per class as a "uniform prior" to prevent zeros (Manning et al., 2009, p.260).

The abstract parameter *pa*, which was introduced in the empirical risk functional (2.13) regarding the NB approach as a "learning machine", is void. Thus, there is no optimization parameter for this approach.

2.4.3.4 Support Vector Machines

The basic idea of the Support Vector Machine (SVM) is to find a hyperplane in a vector space that separates vectors of two classes in the training examples (Boser et al., 1992; Cortes & Vapnik, 1995). Whereas Boser et al. and Cortes & Vapnik are one of the initial developers of the contemporary SVM approach, the approach is based on the "generalized portrait" approach, which dates back to ((Vapnik & Lerner, 1963) cited in (Smola & Schölkopf, 2004, p.199)) and is well grounded in statistical learning theory ((Vapnik, 1982; Vapnik & Chervonenkis, 1974) cited in (Smola & Schölkopf, 2004, p.199); see also (Vapnik, 2000)).

Central to the SVM approach is an optimal hyperplane, which can be defined as a hyperplane "[…] with maximal margin between vectors of the two classes […]" (Cortes &

Vapnik, 1995, p.275). The margin is determined by support vectors, i.e., some document vectors that represent training examples that provide the "[…] largest separation between the two classes." (Cortes & Vapnik, 1995, p.275). In this thesis, the hyperplane separates positive from negative document vectors, see the subsequent Figure 11.

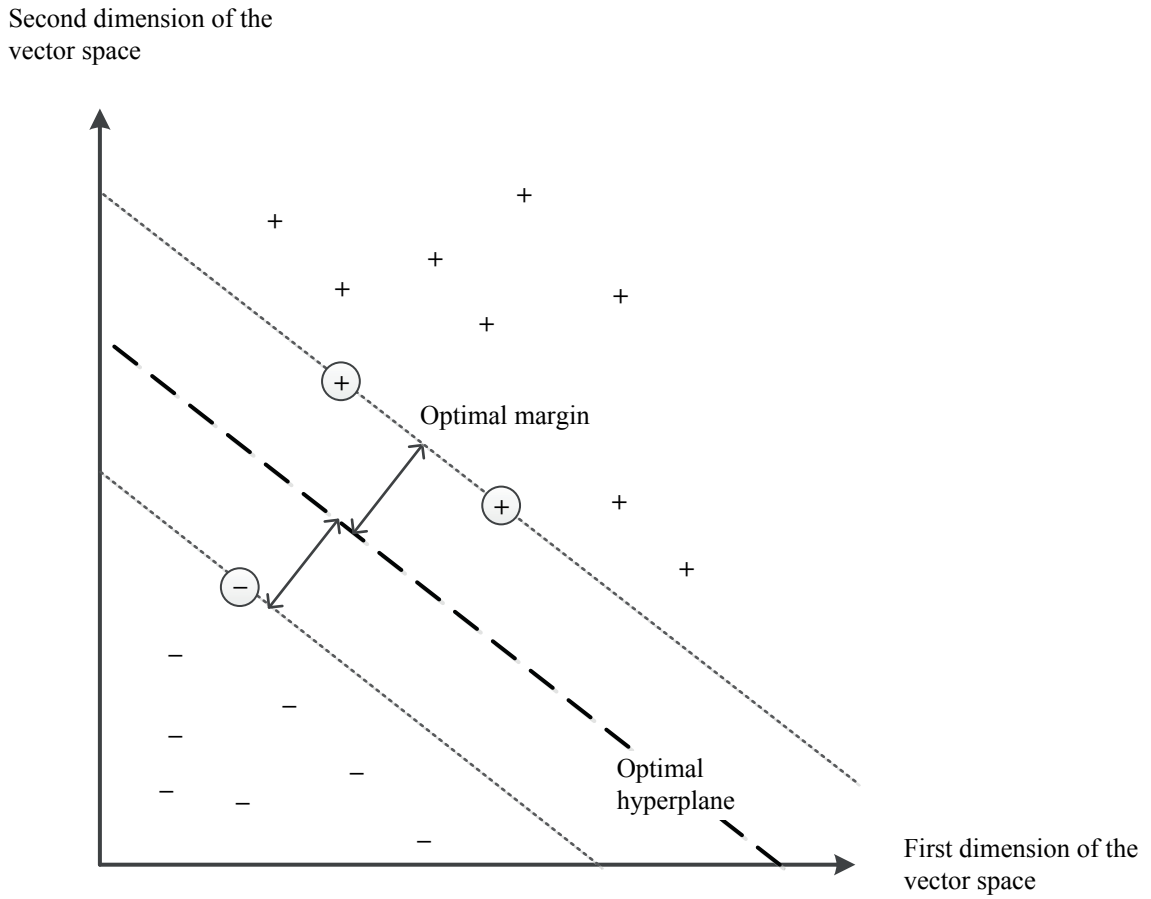


Figure 11: An example of an optimal hyperplane, which separates positive (plus) document vectors from negative (minus) document vectors in a two-dimensional vector space by the optimal margin. The optimal margin is defined by the support vectors (circles) such that the margin is maximal (adapted from (Cortes & Vapnik, 1995, p.275)).

To describe the idea of SVM formally, a hyperplane (i.e., a line in $\mathbb{R}^2$ and a plane in $\mathbb{R}^3$) is defined in vector notation (vectors are in bold) as follows.

Definition: Hyperplane (adapted from (Cortes & Vapnik, 1995, p.278)):

$$\boldsymbol{w} \cdot \boldsymbol{d} + b = 0 \tag{2.20}$$

where

$\boldsymbol{w} \in \mathbb{R}^n$: Weight vector (Boser et al., 1992).

$\boldsymbol{d} \in \mathbb{R}^n$: Arbitrary document vector(s), see Definition (2.12).

$b \in \mathbb{R}$: Bias (Boser et al., 1992).

The hyperplane is spanned by all points (i.e., document vectors $\boldsymbol{d}_i$) that satisfy equation (2.20). The weight vector is orthogonal to the hyperplane (e.g., (Schölkopf & Smola, 2002, p.189)). The optional bias shifts the hyperplane away from the origin (e.g., (Joachims, 2002, p.39)). Central to the hyperplane definition is the scalar product (or dot product, denoted by "$\cdot$").

A (n-1)-dimensional hyperplane divides *all* training examples ($\boldsymbol{d}_i$,s_i) with i=1,2,…,l in the n-dimensional vector space of the document vectors $\boldsymbol{d}_i$ linearly in two parts, if for all i the following inequalities are satisfied (adapted from (Cortes & Vapnik, 1995, pp.277,278)):

$$\begin{aligned} \boldsymbol{w} \cdot \boldsymbol{d}_i + b \geq 1 \quad if\ s_i = 1 \\ \boldsymbol{w} \cdot \boldsymbol{d}_i + b \leq -1 \quad if\ s_i = -1 \end{aligned} \tag{2.21}$$

where

$s_i \in \{-1,1\}$: Investor sentiment document score (see Definition (2.9)) of the annotator for a given $\boldsymbol{d}$.

$\boldsymbol{w},\boldsymbol{d},b$: See Definition (2.20) above.

This translates to the following **necessary requirement for linear separability** of all training examples in a vector space (adapted from (Cortes & Vapnik, 1995, p.278)):

$$\forall_{i=1}^{l}: s_i(\boldsymbol{w} \cdot \boldsymbol{d_i} + b) \geq 1 \tag{2.22}$$

There can be many possible hyperplanes that can satisfy the above requirement. For finding the *optimal* hyperplane, one has to choose $\boldsymbol{w}$ and b, such that the margin, i.e., the distance of the document vectors closest to the hyperplane on either side, is maximized (Boser et al., 1992). The **maximum (i.e., optimal) margin** is $\frac{1}{|\boldsymbol{w}|}$ (Boser et al., 1992). To maximize the margin, one has to minimize $|\boldsymbol{w}|$ subject to the constraint defined in (2.22) (Boser et al., 1992). This term can be translated to minimizing $\boldsymbol{w}\cdot\boldsymbol{w}$ (Cortes & Vapnik, 1995, p.279), which is the *quadratic* norm (i.e., quadratic magnitude) of the vector $\boldsymbol{w}$. Thus, the optimization problem to find the optimal hyperplane with optimal margin can be defined as a *quadratic* programming problem (Boser et al., 1992; Cortes & Vapnik, 1995) with adding the factor ½ for "cosmetic" reasons (Boser et al., 1992).

Definition: SVM Optimization Problem with Optimal Margin (adapted from (Boser et al., 1992; Cortes & Vapnik, 1995)):

$$\min_{\boldsymbol{w},b} \frac{1}{2} \boldsymbol{w} \cdot \boldsymbol{w} \tag{2.23}$$

subject to the constraint defined in (2.22)

A training algorithm for solving this problem to construct the optimal hyperplane, which separates the training examples without errors, was proposed by Boser et al. (1992), building on an approach proposed earlier ((Vapnik, 1982) cited in (Boser et al., 1992)).

The SVM optimization problem defined above follows the structural risk minimization (SRM) inductive principle (see Section 2.4.3.2) (e.g., (Joachims, 2002, p.35)). Thus, it automatically maximizes the generalization ability of the trained classifier by minimizing the two terms that define the upper bound of the true error (see Definition (2.15)). The first term – empirical risk (i.e., the error on the training set) – is minimized to zero by definition of constraint (2.22) for the optimization problem that allows no errors (e.g., (Cortes & Vapnik, 1995, p.286)). The minimization of the second term, which is driven by the VC-dimension (measuring the complexity of a class of classifier functions), is inherent to the optimization problem defined in (2.23) (e.g., (Cortes & Vapnik, 1995, pp.285–286)): The VC-dimension of the maximal margin hyperplane can be estimated on the basis of $|\boldsymbol{w}|^2$ (Vapnik, 2000, p.144) – which is subject to minimization in the optimization problem defined in (2.23) because $|\boldsymbol{w}|^2 = \boldsymbol{w} \cdot \boldsymbol{w}$. More intuitively, the complexity of the SVM classifier function depends on the number of support vectors (Vapnik, 2000, p.141). Because the support vectors are the only vectors required to describe the hyperplane, the corresponding weights are the only non-zero ones (e.g., (Cortes & Vapnik, 1995, pp.279,291-293)) – which are to be minimized as the magnitude of the weight vector is minimized.

Once the SVM optimization problem has been solved using a set of training examples and $\boldsymbol{w}$ and b have been chosen, the following linear classifier (decision) function can be used to classify an arbitrary document vector $\boldsymbol{d}$ not contained in the training set.

Definition: SVM Classifier Function (adapted from (Boser et al., 1992)):

$$\hat{s} = \text{sign}(\boldsymbol{w} \cdot \boldsymbol{d} + b) \tag{2.24}$$

where the sign function $sign(x) = \begin{cases} 1 & if\ x > 0 \\ -1 & else \end{cases}$ where $x \epsilon \mathbb{R}$.

Variables in Definition (2.24) are defined as follows:

$\hat{s} \in \{-1,1\}$: Investor sentiment document score (see Definition (2.9)) of the classifier.

$\boldsymbol{d}$: Document vector not contained in the training set, see Definition (2.12).

$\boldsymbol{w}$: Weight vector, determined by solving (2.23).

b: Bias, determined by solving (2.23).

Note that the input document vector can be transformed by a possibly nonlinear function $\boldsymbol{\varphi(d)}$ (to a high dimensional space), such that the hyperplane separates the training examples in $\boldsymbol{\varphi}$-space (Boser et al., 1992; Cortes & Vapnik, 1995, p.274). Thus, a "nonlinear SVM" (e.g., (Hsieh et al., 2008, p.408)) can be created by using, e.g., a polynomial function (e.g., (Cortes & Vapnik, 1995, p.275)). The higher dimensional space increases the chance to obtain linear separability by means of a hyperplane for a set of training examples ((Schölkopf

& Smola, 2002, pp.200–201), also citing (Cover, 1965)). However, for document classification, the dimensionality is often already high and usually a linear function is used (e.g., (Hsieh et al., 2008, p.408)) by letting $\boldsymbol{\varphi(d)=d}$ like in the classifier function above (Boser et al., 1992). Such an approach is referred to as "linear SVM" (e.g., (Hsieh et al., 2008, p.408)).

The problem with the SVM-approach, seeking an optimal hyperplane by an optimal margin, are sets of training examples that cannot be linearly separated as defined above (Cortes & Vapnik, 1995, p.276). Thus, the above approach was extended for cases in which the separation involves errors (Cortes & Vapnik, 1995). To this respect, a training example document vector residing on the wrong side of the hyperplane would be an error (e.g., (Joachims, 2002, p.40)). Thus, the "soft margin hyperplane"-approach aims at finding a hyperplane with a minimal number of errors and a maximum margin for the correctly classified vectors (Cortes & Vapnik, 1995, pp.280–282). To allow for errors but at the same time minimize the number of errors on finding a separating hyperplane, the following optimization problem has to be solved:

Definition: SVM Optimization Problem with Soft Margin (adapted from (Chang et al., 2008, p.1369; Cortes & Vapnik, 1995, pp.280-281) with $F(u)=u$):

$$\min_{\boldsymbol{w},b} \frac{1}{2}\boldsymbol{w}\cdot\boldsymbol{w} + C\sum_{i=1}^{l}\xi_i \tag{2.25}$$

subject to $\xi_i \geq 0$ and the loss function defined in (2.26)

where

$\boldsymbol{w}$: Weight vector.

b: Bias – part of ξ function, see (2.26).

C: Penalty parameter (Chang et al., 2008, p.1369).

l: Number of training examples.

ξ: Loss function (Chang et al., 2008, p.1369), see (2.26).

The first term in the definition above is equivalent to the optimal margin optimization problem defined in (2.23). Thus, the first term seeks the maximum margin (see the definition above). The second term aggregates the measure of the training errors (provided by the loss function) over all training examples and is subject to minimization to a degree that depends on the penalty parameter. The loss function can be defined as:

Definition: Loss Function (adapted from (Chang et al., 2008, p.1369; Cortes & Vapnik, 1995, pp.280-281)):

$$max\big(0,1 - s(\boldsymbol{w} \cdot \boldsymbol{d} + b)\big)^L \tag{2.26}$$

where

the maximum function $max(x,y) = \begin{cases} x & if\ x > y \\ y & else \end{cases}$ where $x, y \in \mathbb{R}$

$s \in \{-1,1\}$: Investor sentiment document score (see Definition (2.9)) of the annotator for a given **d**.

$\boldsymbol{w},\boldsymbol{d},b$: See Definition (2.20) above.

$L \in \{1,2\}$: Commonly, L=1 (denoting L1-SVM) or L=2 (denoting L2-SVM) (Chang et al., 2008, p.1369).

L1-SVM sums up the losses and L2-SVM sums up the squared losses (Chang et al., 2008, p.1369). Thus, the penalty for losses using L2-SVM is higher. The "loss" is generated by the inner term $s(\boldsymbol{w}\cdot\boldsymbol{d}+b)$, which should evaluate to ≥1 according Definition (2.22) for correctly classified document vectors – leading to a zero "loss". For incorrect classifications, the inner term would be negative, leading to a positive "loss".

An algorithm to solve the soft margin optimization problem has been proposed by Cortes & Vapnik (1995).

Regarding controlling (and maximizing) the generalization ability of the trained classifier, the soft margin classifier also addresses the SRM principle (see Section 2.4.3.2). In contrast to the optimal margin classifier, the soft margin classifier can allow for training errors, which also affect the generalization ability (see Definition (2.15)). The ***C*-parameter** controls the trade-off in the SRM principle between the number of training errors and classifier function complexity (e.g., (Cortes & Vapnik, 1995, p.286)).

Different values of C can be used, e.g., $C=2^i$, where $i \in \{-5, -3, \ldots, 15\}$ (Hsu et al., 2010, p.5). With a small C, there is only a small penalty on the training errors, thus the number of training errors will increase (e.g., (Joachims, 2002, p.40)). An increase of training errors allows the complexity of the classifier to decrease (as essentially the sum of training errors and complexity defines the upper bound for the generalization performance in Definition (2.15) and a less complex classifier function is required to have more training errors). In effect, the generalization ability potentially increases (e.g., (Cortes & Vapnik, 1995, p.286)). Otherwise, with a large C, there will be less training errors. Thus, the solution is closer to the optimal hyperplane (e.g., (Joachims, 2002, p.40)). However, the complexity might be higher and the generalization ability lower.

The C-parameter takes the concrete role of the abstract parameter pa, which was introduced in the empirical risk functional (2.13) regarding the SVM as a "learning machine".

The generalization performance in terms of accuracy (see Definition (2.14)) of the SVM classifier (measured on the test set) can be optimized by this parameter.

2.4.3.5 Assessment with Respect to this Work

Two supervised machine learning approaches for text classification in general and sentiment classification in texts were discussed. This section assesses the performance of these approaches in terms of accuracy and selects an approach for the design of the classifier of the sentiment orientation of investor sentiment in Section 3 on this basis.

Supervised machine learning approaches only require a labeled corpus of training examples for learning. Other kind of manual effort is not required in creating a classifier. Another benefit of the approaches is that classifier model training is automatic and fast (e.g., (Joachims, 2006)), and the trained classifier model is inherently very specific to the corpus, i.e., to the language, the topic, and the domain. Finally, the text classification accuracies are typically high (e.g., (Dumais et al., 1998; Joachims, 1998)). The next paragraphs shed light on respective findings regarding the NB and SVM approaches.

Considering the concrete supervised machine learning approach NB, the assumptions are simplifying and are indeed naïve because terms in a document (given a class) often occur not independently (e.g., (Manning et al., 2009, pp.268–269)). Also, the used text model does not convey the order of the terms (e.g., (Manning et al., 2009, p.269)). However, Naïve Bayes approaches perform well in text classification tasks in terms of accuracy (e.g., (Manning et al., 2009, pp.269–270; McCallum & Nigam, 1998)). For instance, McCallum & Nigam achieved 86% accuracy for classifying university webpages into seven categories (McCallum & Nigam, 1998). They also achieved high accuracies for other corpora (McCallum & Nigam, 1998). Furthermore, Naïve Bayes approaches are computationally efficient (Manning et al., 2009, p.270). Thus, they often serve as a baseline approach for text classification (Manning et al., 2009, p.270).

On comparing SVM's general text classification performance (i.e., not specific to web texts and bipolar sentiment classification) to the one of NB approaches, usually SVM outperforms as evidenced in the following experiments: Using SVM with a non-linear kernel achieved the highest accuracy of 86.5% compared to a NB classifier (72% accuracy) and three other classifier approaches (Joachims, 1998, pp.140–141). This result refers to the topic classification of the newswire documents of a version of the Reuters-21578 corpus (e.g., (Joachims, 1998, p.140; Manning et al., 2009, pp.279–280)). This corpus "[…] was the main benchmark for text classification evaluation." (Manning et al., 2009, p.279). Regarding the same corpus, but using a linear SVM and different (weightings of) features (on feature definition and weighting possibilities, see Section 3.2.1), the accuracy of SVM was again best at 87% compared to NB (which had an accuracy of 75.2%) and three other classifier approaches (Dumais et al., 1998). The basic result that SVM is better than NB and also one of the best performing classifiers in terms of the smallest error was also supported by statistical significance tests (Yang & Liu, 1999). Manning et al. conclude that on average,

NB is less accurate than SVM "[…] when trained and tested on independent and identically distributed (i.i.d.) data […]" (Manning et al., 2009, p.283).

Regarding sentiment classification, it is not that clear, which approach is more accurate. On reviewing several results in the literature, Tang et al. conclude that SVM and NB are both well suited for "single domain document sentiment classification" (Tang et al., 2009, p.10766). To this respect, the classification of the sentiment of movie reviews is widespread in the literature. For instance, Pang et al. found SVM to have the highest accuracy (up to 82.9%) in experiments on the bipolar classification of the sentiment of movie reviews with three different supervised machine learning approaches including NB (which had up to 81.5% accuracy) (Pang et al., 2002). Regarding the bipolar investor sentiment classification in financial blog documents, SVM was shown to have a slightly smaller accuracy (74.4%) vs. the accuracy of NB (75.1%) (O'Hare et al., 2009). The higher accuracy of NB is in line with a comment by Manning et al. who note that many practitioners could not create a classifier performing better than NB – at least not regarding a specific experimental setup, document collection, and class (Manning et al., 2009, p.283). Wang and Manning found some more insights because in their sentiment classification experiments SVM had higher accuracy than NB – but only on two movie review corpora of longer documents with on average 231 and 787 words (Wang & Manning, 2012). Because the average document length of the blog documents in this thesis' corpus for training an investor sentiment classifier is of comparable length (i.e., 664 words on average (see Section 3.1.2 and Table 4)), SVM seems to be the better option regarding this corpus.

Based on the above assessment of the general advantages of supervised machine learning approaches and the discussion of the accuracies of the reviewed concrete approaches, SVM was used for the design of a classifier of the sentiment orientation of investor sentiments in blog documents in Section 3. Whereas SVMs have been shown to achieve high accuracies in sentiment classification tasks (as discussed above), the design of the new classifier required some experiments for determining the optimal parameter settings of the *C*-parameter and the document to vector transformation with respect to a specific corpus of investment blog documents for achieving high accuracy. These issues are discussed in Section 3.

2.5 Effects of Textual Investor Sentiment on Returns

This section reviews findings of studies on effects of investor sentiment classified from textual documents on (abnormal) returns of stocks. The reviewed studies use the various approaches discussed in Section 2.4 to classify investor sentiment in documents from various publication sources. The following publication sources are differentiated: (1) traditional news (e.g., financial newspapers), (2) stock message boards, (3) Twitter posts, and (4) blogs. Whereas this thesis' focus is on blogs, prior studies on news and stock message boards have paved the way for research with respect to blogs. The section on findings related to Twitter

posts discusses results of a recent research stream to identify differences to blogs. Each of the following four subsections is organized with the title of a reviewed paper as a heading.

2.5.1 Effects of Investor Sentiment from News

Evaluating Sentiment in Financial News Articles

Schumaker et al. (2012) studied the effect of subjectivity and (investor) sentiment from financial news articles on the predictability of stock prices. Albeit Schumaker et al. (2012) term it simply "sentiment", it refers to financial news and can thus be categorized as *investor* sentiment. Schumaker et al. (2012) predicted absolute stock prices on a 20 minute time horizon of S&P 500 companies based on 2802 financial news articles (e.g., Financial Times) (Schumaker et al., 2012). The sequential minimal optimization variant of the support vector regression (SVR) machine learning approach with a linear kernel was used for creating the prediction model (Schumaker et al., 2012). Different text representations were tested, i.e., (1) only proper nouns of the text referring to a specific stock and the stock price at the publication time of the text as baseline, (2) baseline enriched with a document level classification of subjectivity (the authors term it "tone"), and (3) baseline enriched with a document level classification of the sentiment orientation (Schumaker et al., 2012).

Schumaker et al. (2012) found that the addition of classifications of subjectivity and the sentiment orientation leads to worse predictive price directional accuracy in comparison to the baseline text representation. Using only articles classified as subjective with the second text representation yielded best overall reported results in (Schumaker et al., 2012): the directional accuracy increases from 50.4% (baseline text representation) to 59% (Schumaker et al., 2012). However, only 61 articles constitute the set of articles classified as subjective (Schumaker et al., 2012). Thus, this result has to be interpreted carefully in terms of drawing general conclusions.

Schumaker et al. (2012) found that using negative articles, price upswings can be better predicted than downswings and using positive articles, price downswings can be better predicted than upswings. They conclude that an inverse relationship between sentiment orientation and future price direction might exist (Schumaker et al., 2012).

Concerning replicability of the results of Schumaker et al. (2012), it is unclear what parameters were used for the training of the prediction model. Furthermore, the document corpus and the implementation of the software system used in Schumaker et al. (2012) for text processing and price prediction seem to be not publicly available (e.g., on a website).

Concerning general validity of the results of Schumaker et al. (2012), the small size of the corpus (2802 documents) might hamper drawing general conclusions. Also, only a five week period – one that "did not have unusual market conditions" (Schumaker et al. 2012, p.14) was used. It remains to be researched whether news articles of time periods with an

economic crisis or debt crisis (accompanied by increased volatility and volume) would yield other price predictability and prediction accuracy results.

From Schumaker et al. (2012) the following conclusions are drawn with respect to this thesis: Because Schumaker et al. (2012) found subjective articles to allow better price predictions (higher directional accuracy) than objective articles, this gives an indication that the content of blog documents might relate to future prices as this thesis focuses on opinionated, and thus inherently subjective blog documents. To tackle this thesis' research question regarding a potential effect of investor sentiment from blogs on abnormal returns, abnormal returns should be rather studied, different market conditions should be studied, and longer time periods could be studied.

More than Words: Quantifying Language to Measure Firms' Fundamentals

Tetlock et al. (2008) studied the effect of a measure of negative words in financial news on unexpected abnormal returns of stocks of S&P 500 companies. The news are from (1) Dow Jones News Service (DJNS), and (2) Wall Street Journal (WSJ) in the time period 1980 – 2004 (Tetlock et al., 2008). The authors found negative words in news texts to (1) contain novel information in relationship to other information, such as stock analysts' forecasts and quantitative historical accounting data, and to (2) predict future returns (Tetlock et al., 2008, p.1439). They used the General Inquirer Harvard IV-4 dictionary (see Section 2.4.1) to obtain negative word classifications (Tetlock et al., 2008, p.1438). Based on this dictionary, Tetlock et al. (2008, p.1443) propose a stationary "measure of negative words" as the standardized fraction of negative words for representing a set of news texts that refer to the same company from a defined time period. For an overall negative (positive) classification of a news story, the measure of negative words is required to be in the highest (bottom) quartile of the distribution from the last year (Tetlock et al., 2008, p.1455).

Using regressions, Tetlock et al. (2008, p.1453) found "[…] that negative words in firm-specific news stories robustly predict slightly lower returns on the following trading day". Statistical significance for this finding is given only for DJNS news and not for WSJ news (Tetlock et al., 2008, p.1453). Also, using an event study, Tetlock et al. (2008, pp.1455–1456) found cumulative unexpected abnormal returns for DJNS news to be much higher than for WSJ news. They found unexpected abnormal returns for DJNS to be 6.6 bps (–4.0 bps) for positive (negative) news on day 1 after the event (Tetlock et al., 2008, pp.1455–1456, Figure 3). For positive news from DJNS, the cumulative unexpected abnormal return increases steadily in the period from day 0 to day 10 of the event study (Tetlock et al., 2008, p.1455). That is, the effect of positive news (from DJNS) on returns lasts at least for 10 days after publication.

Tetlock et al. (2008) performed a long (-short) portfolio simulation with all stocks for which positive (negative) news had been released. Each portfolio was held for one full trading day (Tetlock et al., 2008, p.1456). The returns realized by the portfolio simulation were adjusted by Tetlock et al. (2008, p.1457) for the risk factors of (1) the Fama-French model

(see Section 2.1.2.2), and (2) the Carhart model (see Section 2.1.2.3). They found average *daily* abnormal returns in terms of alpha in the 1995 to 2004 period to be 11.8 bps (11.3 bps) according the Fama-French (Carhart) model (Tetlock et al., 2008, p.1457). They found these results to be statistically significant (Tetlock et al., 2008, p.1457). Tetlock et al. report abnormal (adjusted for Fama-French factors) *annualized* cumulative returns to be 23.17% not taking into account transaction costs and –2.71% with 10 bps round trip transaction costs (Tetlock et al., 2008, p.1459).

Some shortcomings in the approach of Tetlock et al. (2008) with respect to this thesis are as follows. Regarding interpretability, no formal return generation models are provided for realized returns, various forms of expected returns, and abnormal returns. Furthermore, the portfolio simulation has a one day holding period, thus creating lots of transactions and rendering it unprofitable when considering transaction costs. Because they also found cumulative abnormal returns to steadily increase over several days at least for positive news from DJNS (Tetlock et al., 2008, p.1455), longer holding periods seem advisable.

With respect to this thesis it is important to note that Tetlock et al. (2008) do not claim to investigate the effect of investor sentiment on (abnormal) returns. Rather, they study the effect of negative words in news. Also, these news are presumably not much opinionated compared to blogs. Still, their measure of negative words could be also applied to this kind of content and be interpreted as a measure of the level of (negative) investor sentiment represented in a blog document. However, Tetlock et al. do not evaluate the classification performance in terms of classification accuracy. Thus, they omit the problem of obtaining a reference classification. The accuracy might be rather low as discussed in Section 2.4.1. Thus, a more holistic document level classification approach should increase the classification accuracy and also chances for evidencing a potential effect on abnormal returns.

Major differences with respect to this thesis are a focus on longer term (i.e., monthly) effects of investor sentiment on abnormal returns of a portfolio and using document level investor sentiment from blogs with evaluated accuracy instead of measuring negative words in news.

Do U.S. Stock Markets Typically Overreact to Corporate News Stories?

Antweiler & Frank (2006) studied longer term effects of events of publication of news from the WSJ on unexpected abnormal stock returns (using the terminology of this thesis, see Section 2.1.3) using event studies on a sample of more than 200,000 news ranging from 1973 until 2001 with different lengths of post event windows (5, 10, 20, and 40 trading days). The market model was used to compute normal returns (Antweiler & Frank, 2006, pp.7,9).

Antweiler & Frank found indications for overreaction of investors by finding on average statistically significant positive unexpected abnormal returns in (post-)event windows that start before the event, followed by a reversal to negative unexpected abnormal returns after the event (Antweiler & Frank, 2006, pp.9–13). Furthermore, Antweiler & Frank found

indications for stock price drift because the magnitude of the unexpected abnormal return generally increases when increasing the post-event window length to up to 40 trading days (Antweiler & Frank, 2006, p.9). They attribute this price momentum to possible gradual adjustments of large portfolios (Antweiler & Frank, 2006, p.9) and investor psychology effects studied by behavioral finance (Antweiler & Frank, 2006, p.20).

A limitation is that Antweiler & Frank (2006) only report on *average* standardized cumulative unexpected abnormal returns that they found to increase while extending the post-event window. With respect to exploiting this effect, the day-by-day time series of cumulative unexpected abnormal returns would provide for more detailed information. Furthermore, Antweiler & Frank (2006) did not conduct a portfolio simulation. Thus, the realization of portfolio level abnormal returns (in terms of alpha) remains unclear.

With respect to this work, Antweiler & Frank (2006) present interesting evidence for long-term effects of post-event price momentum manifested by increasing unexpected abnormal returns after the publication of news. This corroborates predictions of behavioral finance theory and provides an indication for a potential effect of investor sentiment from blogs on abnormal returns on long time horizons (e.g., monthly). However, Antweiler & Frank (2006) did not study investor sentiment in the news publications or blogs, which might be more related to investor sentiment and stronger effects due to behavioral biases.

Relating News Analytics to Stock Returns

Leinweber & Sisk (2011) studied the effect of (investor) sentiment in news articles on unexpected abnormal return (using the terminology of this thesis) of individual stocks. Leinweber & Sisk (2011) quantified the (investor) sentiment of a news article by the Thomson Reuters NewsScope Sentiment Engine (TRNSE), which they report to estimate a probability score for each of the classes positive, neutral, and negative. Furthermore, the TRNSE is supposed to measure the relevance and novelty of news articles (Leinweber & Sisk, 2011). Leinweber & Sisk (2011) define high sentiment events as days, on which at least one or two news articles were required to be novel, highly relevant, and the sentiment probability score of positive or negative needed to be in the top 5% or top 10% of the daily distribution.

In an event study, the authors study high sentiment events referring to S&P 1500 stocks in the 2004–2008 period (Leinweber & Sisk, 2011, pp.153–158). They use market-adjusted returns as a simple form of unexpected abnormal returns, i.e., a stock's return in excess of the return of a benchmark (Leinweber & Sisk, 2011). Leinweber & Sisk (2011) found that such excess returns after 20 days of a positive (negative) event are 39 (–2) basis points (bps). Whereas the cumulation was not made explicit, the daily time series of (presumably cumulative) excess returns after positive events increases steadily over the 60 day time window (Leinweber & Sisk, 2011). After 40 days, (presumably cumulative) excess returns exceed 60 bps (Leinweber & Sisk, 2011). For the negative events, the excess return time

series starts with negative (presumably cumulative) excess returns and after less than 5 days starts to increase steadily (Leinweber & Sisk, 2011).

From their event studies, Leinweber & Sisk (2011) conclude "[…] that it takes market participants a long time (days) to process a large amount of novel, strongly polar news […] News and event ambiguity, fact validation, cognitive dissonance are all good reasons to hypothesize that investors can take a longer time to process new information that has extreme sentiment." (Leinweber & Sisk, 2011, pp.163–164). They attribute this finding to behavioral effects such as overreaction and herding (Leinweber & Sisk, 2011).

In contrast to Tetlock et al. (2008), Leinweber & Sisk (2011) claim to use only novel stories and they also use stocks of the S&P 1500 instead of the S&P 500, thus including also smaller capitalization stocks. Possibly due to both reasons, Leinweber's and Sisk's event studies show higher and more persistent increases (lasting several days) in (presumably cumulative) excess returns after the event date.

A portfolio simulation from 2006 until 2009-11-01 was used to simulate trades on news events that required at least 4 novel news with sentiments in the top 5% of the daily distribution (Leinweber & Sisk, 2011, pp.164–170). Furthermore, Leinweber & Sisk (2011) restricted the stocks to specific sectors they found to relate to most predictable return responses to events in a period prior to the simulation. Leinweber & Sisk (2011) applied a maximum 20 day holding period, a stop loss rule at 5%, a profit take rule at 20%, and transaction costs of 25 bps per roundtrip. The reported annual Sharpe ratio after transaction costs is 0.76 (Leinweber & Sisk, 2011, p.165). Concerning interpretability of the portfolio simulation, the risk free rate of return, which is required for determining the Sharpe ratio, is not provided. Leinweber & Sisk (2011) report the mean monthly return to be 1.74% and the S&P 500 to be clearly outperformed. Concerning the benchmark, the S&P 1500 (instead of the S&P 500) would be more appropriate because all its constituents are used in the portfolio simulation. The outperformance is realized mainly in the years 2008 and 2009 (Leinweber & Sisk, 2011), which exhibit great losses in the S&P 500. These years also exhibit strong up and down swings in the cumulative return time series of the portfolio (Leinweber & Sisk, 2011, p.164), which is usually not desired by investors.

Shortcomings of Leinweber & Sisk (2011) are as follows. Leinweber & Sisk (2011) is a book section and does not follow scientific rigor to the extent a journal publication would require. A lack of rigor manifests in the following issues: The exact nature of the news articles compiled by the TRNSE as well as their sources are not revealed. Leinweber & Sisk (2011) do not comprehensively specify the portfolio construction rules. Furthermore, the calculation and aggregation of unexpected abnormal returns through time and across stocks is not explicitly defined in Leinweber & Sisk (2011). The approach for determining the sentiment using TRNSE is a black box. Thus, the sentiment quantification approach is not replicable. However, TRNSE outputs are available for everyone at the expense of commercial license fees. Thus, the event studies would be replicable in principle. Another potential shortcoming

is the use of the simple market-adjusted return model for estimating unexpected abnormal returns. That is, the found unexpected abnormal returns depend on the validity of this model. In contrast, for instance, Tetlock et al. (2008) use the more sophisticated Fama-French three risk factor model (see Section 2.1.2.2). Also, no test of the statistical significance of results was conducted.

With respect to this thesis, the work of Leinweber & Sisk (2011) indicates highly positive or negative investor sentiment to have an effect on future unexpected abnormal returns of individual stocks on time horizons of up to 3 months. In contrast to Leinweber & Sisk, an investor sentiment index (see Definition (2.10)) is used in this thesis, not a probability score. Furthermore, portfolio level effects of investor sentiment on abnormal returns are studied. For selecting stocks into a portfolio, using high and low level investor sentiment stocks seems advisable based on Leinweber's & Sisk's findings. Finally, statistical significance of potential effects was tested on a monthly basis in this work.

2.5.2 Effects of Investor Sentiment from Stock Message Boards

Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards

The study of Antweiler & Frank (2004) is concerned with contemporaneous and predictive effects of 1.5 million messages in stock message boards from Yahoo! Finance and Raging Bull with respect to returns, trading volume, and volatility of the 45 stocks in the Dow Jones Industrial Average (DJIA) and the Dow Jones Internet Commerce Index (DJICI) in the year 2000. Antweiler & Frank (2004, p.1292) found support for the hypothesis that these messages contain financially relevant information.

The text messages were classified using both, a NB classifier (see Section 2.4.3.3) and SVM (see Section 2.4.3.4), in classes bullish, bearish, or hold (Antweiler & Frank, 2004, p.1264). For training the classifier, a manually classified corpus of 1000 messages was used (Antweiler & Frank 2004, p.1265). The classifiers of either approach are reported to have comparable classification performance (Antweiler & Frank, 2004, p.1264). For the NB classifier, the overall accuracy (calculated based on numbers in Table I in (Antweiler & Frank 2004, p.1266)) is 88.1%.

Antweiler & Frank define an aggregate measure of bullishness B taking into account the number of messages classified as bullish (M^{BUY}) and bearish (M^{SELL}) of a given time period: $B = \ln((1+M^{\text{BUY}}) / (1+M^{\text{SELL}}))$ (Antweiler & Frank, 2004, p.1267). This measure weighs messages equally and discounts large message numbers (Antweiler & Frank, 2004, p.1267). Although it is supposed to measure "bullishness", Antweiler & Frank also note that it "[…] takes into account the number of traders expressing a particular sentiment." (Antweiler & Frank, 2004, p.1267). The aggregate measure of bullishness B could also be interpreted as a real-valued and unbounded investor sentiment index similarly to Definition (2.10). In case

there are no messages or the amount of positive and negative messages is equal, *B* is zero. In case there are more bullish (bearish) messages, *B* is positive (negative).

Antweiler & Frank tested for relationships between the above aggregate measure of bullishness, number of messages posted in a time interval, and other variables derived from Yahoo! Finance stock message boards with respect to stock returns (Antweiler & Frank, 2004, pp.1276–1281). They found significant but small negative correlations respectively for the number of messages (–0.009) and the bullishness (–0.005) – both aggregated over the last 24 hours – with respect to the current 15 minute interval of log returns (Antweiler & Frank, 2004, p.1277).

Panel regression analyses of contemporaneous 15 minute interval log returns as dependent variable and log of number of messages and bullishness as independent variables did not yield coefficients that are significant at the 5% level (Antweiler & Frank, 2004, p.1278). In this thesis, the significance level denotes the maximum permissible probability of erroneously rejecting the null hypothesis (like in, e.g., (Gujarati & Porter, 2009, pp.833–834)) and accepting the alternative hypothesis of a statistical hypothesis test. This way of denoting the significance level is used in this thesis irrespective of the way used in the papers reviewed.

In a predictive panel regression with respect to the current day daily log return, Antweiler & Frank (2004, p.1279) found statistically significant coefficients at the 1% level for the logarithm of the number of messages aggregated on the prior day (–0.002) and the second-last day (0.002) as independent variables. This means that a 100% increase of the number of messages on the prior day leads to a –0.2% lower stock price today (Antweiler & Frank, 2004, p.1281). For bullishness, the magnitude of the coefficients is almost the same (–0.002 and –0.003) but not statistically significant at the 1% or better significance level (Antweiler & Frank 2004, p.1279).

The study of Antweiler & Frank (2004) has the following limitations regarding their classification approach: Antweiler & Frank (2004) do not report on out of sample classification performance. They only report on classification performance on the training set (i.e., “in sample” (Antweiler & Frank, 2004, p.1266)). This performance clearly does not constitute a proper evaluation of the classifiers. With respect to the small set of manually classified documents, a cross-validation seems advisable. Antweiler & Frank (2004) also do not provide details of the manual annotation process. It is unclear by which criteria they selected the 1000 messages they annotated. It is also unclear how many messages were annotated by whom of the two authors and to which extend there is inter-annotator agreement, which can be measured by (some variant of) the Kappa metric (e.g., (Cohen, 1960, 1968)).

Although the study of Antweiler & Frank (2004) has some limitations, with respect to this thesis it is interesting that Antweiler & Frank (2004) found predictive information in internet stock message boards with respect to total daily stock returns. This finding indicates

investment-related web content to contain novel (i.e., not yet priced) information regarding stock returns. However, there are differences between this type of content and investment-specific blogs. The messages studied by Antweiler & Frank often contain predicted price changes and they are short because they "most frequently" contain "between 20 and 50" words (Antweiler & Frank, 2004, p.1263). In contrast to these characteristics, investment blog documents are typically much longer and often resemble full length edited articles (see Section 2.3 and Section 3.1.2), possibly containing pro and contra arguments.

As mentioned above and referring to Antweiler & Frank (2004, p.1267), one can interpret the measure of bullishness as an investor sentiment index. Thus, using the investor sentiment index for stock ranking and portfolio construction to study effects on abnormal returns on this level would be possible but was not conducted by Antweiler and Frank (2004). Unlike this work, they did not consider longer term effects at monthly frequency.

Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web

Das & Chen (2007) evaluated five tripolar (buy/positive, sell/negative, hold/neutral) investor sentiment classifier approaches for messages from stock message boards on the web and studied predictive effects of aggregated investor sentiments to the 24 stock Morgan Stanley High-Tech Index (MSH). Generally, Das & Chen (2007, p.1382) found classification of messages in stock message boards to be a hard problem due to the messages being highly ambiguous, often not adhering to grammar, and containing non-standard vocabulary.

Das & Chen (2007) propose the following types of (rather simple) classifier approaches: (1) net scores of all positive (score +1) and negative (score –1) words occurring in a document using a hand-selected lexicon of classified words, (2) representing a manually labeled and also an unclassified document as vector with term frequencies of all terms in a document and use the angle between the vectors as measure of closeness, (3) adding weights to the terms of the previous approach, (4) net scores of positive/negative words that occur in text parts around an adjective or adverb that occur in a noun phrase, and (5) a Bayesian classifier using only the words that appear in the hand-crafted lexicon (Das & Chen, 2007, pp.1378–1380). Furthermore, a majority vote classifier of classifiers (1) to (5) has been proposed (Das & Chen, 2007, p.1380).

The accuracy of the classifiers was found to be (1) relatively low (best results ranging between 60 and 70%) due to ambiguity in messages, (2) to increase when using more training samples, (3) to increase when sorting out ambiguous documents, reducing also the rate of false positives (Das & Chen, 2007, pp.1380–1383).

The majority vote classifier has a low false positive rate (~4%) and was used for experiments that correlate investor sentiment and stock market variables based on 145,110 messages from 2 months in 2010 referring to 24 MSH stocks (Das & Chen, 2007, p.1383). Investor sentiment was aggregated per day and per stock using all respective messages to a normalized investor sentiment time series, which was further aggregated over all stocks to an

investor sentiment index (Das & Chen, 2007, p.1384). Using a regression analysis, the MSH index was found to be "[…] weakly related to the sentiment index value from the previous day at the 10% significance level." (Das & Chen, 2007, p.1384).

Das & Chen (2007) conducted a tripolar classification but used only two classes of investor sentiment orientation in aggregating investor sentiment to an investor sentiment index. Because O'Hare et al. (2009) have shown that bipolar classification achieves better classification performance than a tripolar one, this seems to be a better option for the design of a classifier in this work (see below for the review of O'Hare et al.'s paper).

With respect to this work, the finding of Das & Chen (2007) that the classification of the sentiment orientation of investor sentiment in noisy and ambiguous stock message board text content from the web is a hard problem should apply also to blog documents. Section 3 of this thesis explores whether the moderate classification performance reported by Das & Chen (2007) can be exceeded with respect to blog documents.

A limitation of the experiments in Das & Chen (2007) relating investor sentiment and price time series is that only 2 months of data were used for testing on statistically significant relationships. Furthermore, in the regressions, no control variables (which could also explain the relationship) except the lagging price were included. In Gilbert & Karahalios (2010), also trading volume and price volatility were used (see below for a review of their paper).

Although there are some limitations, the results of Das & Chen (2007) provide evidence of a (weakly statistically significant) relationship between their aggregated investor sentiment index and the next day stock index price. This thesis extends on these findings by studying (1) investor sentiment from (investment) blogs, which constitute also web content but somewhat different (see Section 2.3 and Section 3.1.2), (2) (longer term, i.e., monthly) effects of investor sentiment on abnormal portfolio stock returns, (3) effects relating to the highest and lowest levels of investor sentiment, and (4) also non-technology stocks.

2.5.3 Effects of Investor Sentiment from Twitter

Messages from Twitter, i.e., tweets, allow for microblogging, i.e., blogging with short text messages. As exemplary research results regarding effects of investor sentiment from tweets on the stock market, the following well-known paper is reviewed. Note that in recent years, research in this direction has gained momentum (e.g., (Bar-Haim et al., 2011; Oh & Sheng, 2011; Smailović et al., 2014)). However, one of the starting points has been the following paper.

Twitter mood predicts the stock market

Bollen et al. (2011) argue that financial decision making is driven in part by public mood based on a study tackling the research question whether public mood extracted from large amounts of tweets and aggregated into a daily time series predicts DJIA stock index prices.

Bollen et al. (2011) analyzed 9.8 million tweets from a 10 month period in 2008 using (1) a positive/negative polarity classifier by means of the publicly available OpinionFinder software and the accompanied lexicon of positive and negative words, and (2) a six-dimensional mood classifier for mood states “calm”, “alert”, “sure”, “vital”, “kind”, and “happy” (Bollen et al., 2011, pp.2–3). Both classifiers work on the word (or phrase) level of tweets and served as basis for creating an aggregated normalized score for each mood dimension on a daily basis using all tweets from the same day respectively and taking into account each occurrence of a classified mood word (Bollen et al., 2011, pp.2–3).

Bollen et al. (2011) similarly to Gilbert & Karahalios (2010) (cited in (Bollen et al., 2011, p.4) and reviewed below) conducted a Granger-causality analysis and found a predictive linear Granger-causal relationship between lagging values of the “calm” mood score time series (with lags of 2 to 6 days) and current daily closing price changes of the DJIA (Bollen et al., 2011, pp.4–5). However, they also found price changes triggered by unexpected news to be not predicted by the mood time series (Bollen et al., 2011, p.5).

Bollen et al. (2011) created a non-linear prediction model for the current day DJIA closing price using a self-organizing fuzzy neural network that uses an online learning algorithm (Bollen et al., 2011, pp.5–7). Bollen et al. (2011) found the model with variables of the past 3 days of DJIA closing prices and the past 3 days of the not normalized “calm” mood time series to have a price change (up or down) bi-directional accuracy of 86.7% (Bollen et al., 2011, pp.5–7). This accuracy substantially exceeds the directional accuracy (73.3%) of the baseline model with only 3 past DJIA closing price variables (Bollen et al., 2011, p.6). In contrast to this, the mood extracted with OpinionFinder was not able to increase the directional accuracy (Bollen et al., 2011, p.6). The predictive success of the “calm“ mood seems to be in contrast to Tetlock et al. (2008). They prefer a measure of negative words to positive ones for representing texts and predicting unexpected abnormal returns.

A shortcoming in the study of Bollen et al. (2011) is that in contrast to Gilbert & Karahalios (2010, pp.61-62), who conducted a similar kind of Granger-causality analysis, Bollen et al. (2011) used a baseline model of only lagging price change time series and omitted price volatility and volume. That is, the model of Bollen et al. (2011) misses further variables that could explain stock index price changes. Therefore, the explanatory power of the mood score variable(s) may appear higher and the null hypothesis (that a mood score time series does not predict DJIA price changes) can be suspected to be more likely to be rejected in the study of Bollen et al. (2011). Also, concerning the Granger-causality analysis, Bollen et al. (2011) have not reported on coefficients for the “calm” mood score in the linear predictive model for DJIA price changes. Thus, it is unclear whether the relationship is positive or negative on average.

Regarding the non-linear DJIA price prediction model, the following shortcomings can be identified: (1) Only a short 15 trading day test period was used to compute the metrics (Bollen et al., 2011, p.6). (2) Regarding the directional accuracy, the majority percentage (of

either up or down price movements) in the sample is not reported. Because on the long run there should be more days with up price movements, the majority percentage should exceed 50%. When considering only parts of the year 2008 of the financial crisis like Bollen et al. (2011) did, there might be more days with down price movements. Still, the majority percentage would probably not be 50%. Thus, the baseline of directional accuracy, on which to judge prediction performance, is unclear. (3) In contrast, Bollen et al. (2011, p.6) assume a 50% chance of success of guessing the price direction. This assumption seems to be flawed for the reason stated. Thus, the calculation in Bollen et al. (2011, p.6) of the odds of achieving the reported price prediction performance by chance is presumably also flawed. The actual probability should be higher than reported.

A limitation with respect to the whole study of Bollen et al. (2011) is that the data relates to only a 10 month time period in 2008, which has been a year with distinct and large price movements due to a general financial crisis. Thus, generalizing the results and conclusions requires some caution.

Despite the outlined limitation, Bollen et al. (2011) provide evidence for stock index price predictability based on tweets and based on using rather simple natural language processing approaches. However, Bollen et al. (2011) did not use investment-specific tweets. Also, their analysis is not specific to certain stocks. Consequently, they predict stock *index* prices. In contrast, this thesis studies effects on abnormal returns of stock portfolios. Furthermore, this thesis studies longer term effects on a monthly basis and uses a much longer overall period for the study.

Bollen et al. (2011) use the term "mood" seemingly interchangeable with the term "sentiment", e.g., in "[…] progress has been made in sentiment tracking techniques that extract indicators of public mood directly from social media content […]" (Bollen et al., 2011, p.2). Thus, the one-dimensional mood can be considered a synonym for sentiment. However, Bollen et al. (2011) study "public mood", which is not *investor* sentiment because the tweets are not necessarily related to investment topics. Also the mood is not specific to certain stocks but rather a general sentiment. In this thesis, investor sentiment from investment-related documents with respect to certain stocks is studied. Whereas Bollen et al. (2011) study tweets that constitute microblogs, this thesis studies investment blogs. They contain much more textual content per "document" and also possibly use different perspectives for their stock analysis (see Section 2.3).

2.5.4 Effects of Investor Sentiment from Blogs

Topic-Dependent Sentiment Analysis of Financial Blogs

O'Hare et al. (2009) is one of the first papers to address (investor) sentiment analysis of *financial* blogs. O'Hare et al. found financial blog documents to often discuss multiple stocks or companies (O'Hare et al., 2009, pp.11-12). Thus, their approach for sentiment analysis first extracted all parts of a document that refer to a specific stock or company by extracting

the stock or company by its name and a specific number of the surrounding (a) words, (b) sentences, and (c) paragraphs (O'Hare et al., 2009, p.13). All stock or company-specific parts of the document were then used for machine learning classification using (1) linear-kernel SVM (see Section 2.4.3.4), and (2) Multinomial NB (see Section 2.4.3.3) (O'Hare et al., 2009, p.13). For training the classifier, O'Hare et al. (2009, pp.11-12) created a manually labeled corpus of 979 unique documents and 1526 unique pairs of (parts of a) document and a stock (or the respective company) from the S&P 500 stock index. Each document-stock pair was labeled with 5 labels expressing polarity (very negative, negative, neutral, positive, very positive) (O'Hare et al., 2009, p.11). O'Hare et al. (2009, p.12) found fewer labels to allow for increased inter-annotator agreement. Thus, they created a bipolar (positive/negative) and a tripolar (positive/neutral/negative) classifier using a word-unigram text representation (see Section 3.2.1.1) of the labeled training documents in their corpus and Boolean model-weighting of the unigrams (see Definition (3.1)) (O'Hare et al., 2009, p.13).

In their experiments, O'Hare et al. found (1) stock-specific classification to be better than document level classification, (2) Multinomial NB to be slightly better than SVM in terms of accuracy, (3) bipolar classification to be better than a tripolar one in terms of accuracy, and (4) extraction of object-specific content based on 30 surrounding *words* to perform best for bipolar classification in terms of accuracy (75.07%) compared to surroundings based on sentences and paragraphs (O'Hare et al., 2009, pp.14-15).

Some shortcomings in the approach of O'Hare et al. (2009) are as follows. O'Hare et al. (2009) did not ground or explicate the meaning of the (sentiment orientation) labels they use for classification. Thus, the meaning of, e.g., the neutral label is unclear. Neutral could mean, for instance, (a) no sentiment, or indicate (b) a sideways price development of a stock. Another shortcoming is that O'Hare et al. (2009) use blog documents that were collected in eight weeks in 2009 only. A more diverse set of documents covering multiple types of market periods over longer and multiple time periods would allow for possibly creating a more robust classifier. Furthermore, from O'Hare et al. (2009, p.12) it is unclear, which kappa measure and calculation formula for inter-annotator agreement the authors use (for instance, Cohen's original kappa (Cohen, 1960) or the weighted kappa (Cohen, 1968) or another modified version). Thus, the interpretation is hampered.

With respect to this work, the corpus of O'Hare et al. (2009) is of special interest because O'Hare et al. (2009) have created one of the first corpora of manually labeled financial blog documents. Such a corpus is a necessity for (1) creating a machine learning classifier, and (2) evaluating any kind of classifier. Thus, the corpus is also of interest for this thesis. However, it is not publicly available.

O'Hare et al. assume the slightly lower accuracy of SVM compared to NB to be possibly due to the linear kernel and default parameters (C=1; see the end of Section 2.4.3.4 for an explanation of the C-parameter) (O'Hare et al., 2009, pp.13–14). However, there is evidence that SVM generally achieves higher accuracy than NB (see the discussion in Section 2.4.3.5).

Thus, this thesis resorts to SVM and reports on experiments on choosing the *C*-parameter (in Section 3.3.2). Furthermore, based on O'Hare et al.'s findings, bipolar classification of the sentiment orientation seems advisable. This finding correlates with the model of investor sentiment in this work (see Section 2.3.4). Finally, O'Hare et al. (2009) used a sub-document level investor sentiment classification that aims to be specific to a certain stock or company. In contrast, in this thesis, document level investor sentiment classification is used for the following reasons: (1) simplicity; if this thesis' study finds evidence for tackling its research question with an even simpler approach, this seems reasonable, (2) the analysis of this thesis' corpus of investment blog documents reveals that many blog documents that discuss multiple stocks still exhibit the same sentiment orientation on a bipolar level for all these stocks (see Section 3.1.2).

The Impact of Blog Recommendations on Security Prices and Trading Volumes

Fotak (2007) studied 500 blog documents from 2006 from the Seekingalpha blog platform (see Section 2.3.2) and the effects of stock recommendations in these documents on unexpected abnormal returns using the event study method ((Campbell et al., 1997) cited in (Fotak, 2007, p.10)). He found the blog documents to provide *some* genuine novel information and to not attempt to manipulate markets (Fotak, 2007, p.26). Fotak found stock recommendations in blog documents on the Seekingalpha platform to concern large firms (Fotak, 2007, p.14). Regarding the event study, Fotak (2007) estimated parameters of the Fama-French model (see Section 2.1.2.2) to calculate unexpected abnormal returns for the event window, i.e., the days before, on, and after the event of publication of a Seekingalpha blog document. Fotak (2007) used the J2 test statistic (Campbell et al., 1997) to test for the hypothesis that mean (cumulative) unexpected abnormal returns (using the terminology of this thesis, see Section 2.1.3) are different from zero during the event window. Generally, the cumulation in the event study method is calculated over all or some consecutive days of the event window (e.g., (Campbell et al., 1997, p.160)). Furthermore, the mean is generally calculated over the unexpected abnormal returns of all stocks on the same day relative to the respective events (e.g., (Campbell et al., 1997, p.161)).

Fotak (2007) applied the event study method on (1) all blog documents in the sample, and (2) a dataset controlled for other news or blog documents, consisting of 148 long recommendations and 60 short recommendations (Fotak, 2007, p.13). Regarding all blog documents, Fotak (2007) found long and short stock recommendations to be preceded by negative mean (cumulative) unexpected abnormal returns on several days. Because the mean cumulative unexpected abnormal returns – cumulated over days 0 until 19 following the stock recommendation in a blog document – are *positive* with 0.68% and statistically significant at the 10% level (*negative*, –3.2%, not statistically significant) for the *long* (*short*) recommendations, Fotak (2007, pp.15,16,33,34) concludes that recommendations are correct and that short recommendations are momentum-based (i.e., assuming a continuation of the price development) and long recommendations are contrarian-based (i.e., assuming a price

development reversal). However, he also found that the correct recommendations are at least partly due to momentum (Fotak, 2007, p.22). That is, the price changes (that could be observed prior to a blog document publication) continue in the same direction (e.g., (Jegadeesh & Titman, 1993, 2011)).

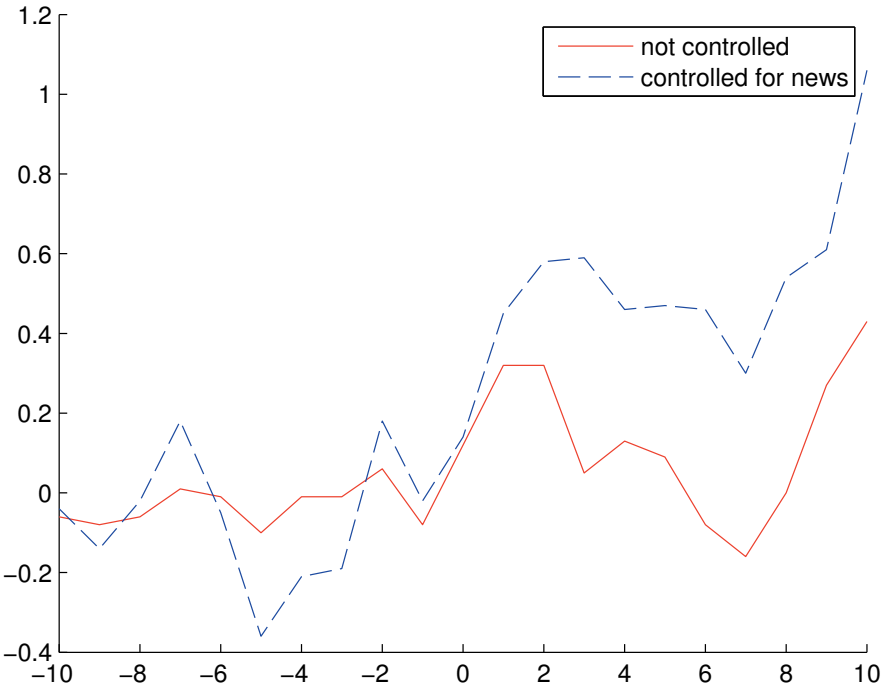


Figure 12: Mean cumulative unexpected abnormal return for events of long recommendations in Seekingalpha blog documents (in %) using data from (Fotak, 2007, pp.33,37). The dashed line is based on events that have been controlled for news and other blog documents (Fotak, 2007, p.13). The horizontal axis displays time in days relative to the event day (day 0).

Regarding the long recommendations, Figure 12 shows positive mean unexpected abnormal returns on day 0 and 1 (see also (Fotak, 2007, pp.33,37)). Thus, there seems to be a positive price reaction on the blog documents with long stock recommendations, and recommendations seem to be correct (see also (Fotak, 2007, p.22)). On day 1, the reaction in mean unexpected abnormal return is 0.2% for the not controlled for news dataset (statistically significantly different from zero at the 5% level) and 0.3% for the controlled for news dataset (statistically significantly different from zero at the 10% level) (Fotak, 2007, pp.33,37). After day 2, the price reaction seems to revert to some degree for the dataset not controlled for news and to a stronger degree for the dataset controlled for news (see also (Fotak, 2007, pp.33,37)). Starting on day 8, some further positive mean unexpected abnormal return is realized. Regarding day 10, this mean unexpected abnormal return is statistically significantly different from zero at the 10% level for both datasets (Fotak, 2007, pp.33,37). It is unclear, whether this mean unexpected abnormal return is due to the stock recommendations in blog documents published on day 0. However, taking this positive mean unexpected abnormal return into account, the mean *cumulative* unexpected abnormal return for days 1 to 10 for the controlled for news dataset amounts to 0.92% and reverts to 0.4% for days 1 to 21 (based on data from (Fotak, 2007, p.37)). The reason for this reversal is unclear.

It might indicate some opposing information, e.g., an overreaction and correction regarding the stock recommendation, or uncontrolled for new opposing information. Compared to the price reaction on day 0 and day 1 (0.47% mean unexpected abnormal return in the controlled for news dataset (Fotak, 2007, p.37)), the results of Fotak (2007) do not indicate a persistent price drift over the 21 day-post-event-time-horizon for long recommendations in blog documents. Thus, it remains unclear whether monthly effects on abnormal returns for long stock recommendations exist on the level of a portfolio selected based on the recommendations throughout a month.

Figure 13: Mean cumulative unexpected abnormal return for events of short recommendations in Seekingalpha blog documents (in%) using data from (Fotak, 2007, pp.34,38). The dashed line is based on events that have been controlled for news and other blog documents (Fotak, 2007, p.13). The horizontal axis displays time in days relative to the event day (day 0).

Regarding the short recommendations, Figure 13 shows only small mean unexpected abnormal returns up to day –2 prior the event of a stock recommendation in a blog document. On day –1, a large negative mean unexpected abnormal return of –2% is realized in the dataset controlled for news (Fotak, 2007, p.38). The same kind of reaction of –1,6% is realized one day later (i.e., on the event day) for the dataset not controlled for news (Fotak, 2007, p.34). Both mean unexpected abnormal returns are statistically significantly different from zero at the 1% level (Fotak, 2007, pp.34,38). This finding indicates that on average, there is a negative price reaction related to short recommendations and that the recommendations are correct (see also (Fotak, 2007, p.22)). Furthermore, the dataset with information also contained in news seems to contain more recent information than the blog only dataset. The blog only information seems to be (partially) known prior to publication when considering the day –1 price reaction. In both cases, the main price reaction happens on days –1 and 0. Therefore, the price reaction could be hardly exploited.

After the event day, between days 1 and 10, the mean cumulative unexpected abnormal returns stay roughly on the same level (Fotak, 2007, pp.34,38). When considering the mean *cumulative* unexpected abnormal return on a longer time horizon starting one day after the event day, the controlled for news dataset yields –2.89% (for days 1 to 21) and the not controlled dataset yields –1.63% (for days 1 to 19), based on reported data (Fotak, 2007, pp.34,38). Although not reported by Fotak (he reports the cumulations from days 2 to 21 and 0 to 19 to be not statistically significant (Fotak, 2007, pp.34,38)), these results are presumably statistically not significantly different from zero. Thus, one cannot judge on monthly effects, in which this work is interested in.

Relevant shortcomings in the approach of Fotak (2007) with respect to this thesis are as follows. Fotak (2007) presumably classified the stock recommendations into long (i.e., expecting a positive future price development) and short (i.e., expecting a negative future price development) *manually*. The annotation process is not described in Fotak (2007), i.e., who annotated the blog documents? How many annotators were involved? What professional background did annotators have? Due to ambiguity in financial web documents (e.g., (Das & Chen, 2007)) annotation requires knowledgeable annotators and possibly multiple annotators to be able to measure inter-annotator-agreement (e.g., (Cohen, 1960)).

Apparently Fotak (2007) did not use an approach for *automatic* investor sentiment classification (with the positive (negative) sentiment orientation referring to a long (short) stock recommendation). Therefore, there would be large manual effort involved in a large-scale study over a long time period (e.g., using a portfolio simulation, which is missing in Fotak (2007)). Furthermore, an event is constituted by one blog document only in Fotak's paper, which might have a negligible effect on unexpected abnormal returns. Also, a single blog document author might be wrong in a stock recommendation. Therefore, studying an aggregated investor sentiment (index) of several blog documents from a certain time period seems reasonable. Also, the 500 blog documents stem from 2006 and from the Seekingalpha platform only. Different results might be observed based on more recent blog documents and other blog platforms.

Regarding controlling for confounding events, Fotak (2007, p.13) only controlled for news and other blog documents in the interval between days –1 and +1 relative to the event date of publication of a stock recommendation in a blog document. It seems more reasonable to control for other events in the whole event window. Otherwise some findings regarding post-event mean (cumulative) unexpected abnormal returns on days 2 until 21 might have been influenced by other events.

Interpretability of the results of Fotak (2007) would be improved by providing the mean unexpected abnormal returns for event days –20 to –11 and 11 to 21. The approach of Fotak (2007) and his findings would be replicable provided the dataset of 500 blog documents and their long/short classification would be available.

Summarizing, Fotak (2007) supports the study of Seekingalpha in this thesis by providing indications for stock recommendations in Seekingalpha blog documents to provide *some* novel (i.e., not yet priced) and predictive information regarding unexpected abnormal returns. However, the exploitability of *single* stock recommendations in a trading strategy can be questioned. Also, it remains an open research question, whether an aggregate investor sentiment from *multiple* blog documents has an effect on abnormal returns on the portfolio level. Thus, in this thesis, (a) more, and (b) more recent data is used, higher levels of aggregation (i.e., monthly investor sentiment) are studied, and portfolio level effects on abnormal stock returns are studied. This thesis proposes a design of a classifier for blog documents regarding the sentiment orientation (in long/positive and short/negative) of investor sentiments, which allows for a large-scale study, vastly exceeding the 500 blog documents used by Fotak (2007). Whereas Fotak (2007) has found many bloggers on the Seekingalpha platform to be finance professionals, this might be not the case for other platforms such as Blogspot (see Section 2.3.2). This thesis also studies effects of investor sentiment from blog documents of the Blogspot platform. Because Fotak (2007) has found Seekingalpha documents to mostly refer to large capitalization stocks, this finding provides an argument for this thesis to focus on DJIA stocks.

Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media

Chen et al. (2014) studied 97,070 opinionated Seekingalpha investment blog documents (see Section 2.3.2) from the time period 2005–2012 (each referring to a single stock) with respect to effects on future abnormal stock returns. The blog documents refer to 7,422 different stocks (Chen et al., 2014, p.13) with larger than average market capitalization (Chen et al., 2014, p.15). Like Tetlock et al. (2008) (reviewed above), Chen et al. (2014) related the fraction of negative words in all blog documents of a day referring to the same stock to future abnormal stock returns using the list of negative words by Loughran and McDonald (see Section 2.4.1). They found Seekingalpha blog documents to be relevant for the pricing of stocks (Chen et al., 2014).

In a regression analysis, they found the daily fraction of negative words with respect to a stock to be statistically significantly negatively related to future abnormal stock returns. They calculated abnormal returns according ((Daniel et al., 1997) cited in (Chen et al., 2014, p.14)) as the difference of total stock returns and returns of a portfolio of stocks of "[…] similar size, book to market ratio, and past returns" (Chen et al., 2014, p.14). They found a 1% increase in the daily fraction of negative words to relate to a 0.379% decrease of abnormal stock returns over 58 trading days, starting two days after publication (Chen et al., 2014, pp.15–16). Results are reported to be robust for controlling for DJNS news articles (Chen et al., 2014, pp.16–17).

Chen et al. (2014) conducted a portfolio simulation that selects stocks with the daily fraction of negative words (in blog documents that refer to the respective stock) being in the top (bottom) quintile or quartile in a long (short) portfolio and holds them for three months,

starting two days after portfolio formation. The difference between the average daily abnormal return of the long and the short portfolio based on the quintile (quartile) is 2.6 bps (2.4 bps) (Chen et al., 2014, p.20). A shortcoming is that it is unclear over which period the distribution of the fraction of negative words is determined. The strongest gain in combined long and short portfolio returns is observed in years 2007 and 2008 (i.e., roughly 40%) (Chen et al., 2014, p.21). From 2009 until 2012 the portfolio returns are rather low (Chen et al., 2014, p.21).

With respect to this thesis, it is important to note that Chen et al. (2014) did not study "investor sentiment" (in blog documents), they rather studied the fraction of negative words (representing the "tone" (e.g., (Chen et al., 2014, p.20))) in blog documents, similar to Tetlock et al. (2008), reviewed above. Like Tetlock et al. (2008), they also did not measure the accuracy of this measure of "tone". Such a dictionary-based approach to text classification typically observes a lower accuracy compared to a machine learning-based approach (see Section 2.4). Thus, this thesis considers a machine learning approach to be better suited for studying the effects of investor sentiment in blog documents on abnormal stock returns.

A limitation regarding the portfolio simulation of Chen et al. (2014) is the daily aggregation of the measure of negative words. This thesis proposes to study higher aggregates over a month, allowing the aggregate to be based on a larger number of blog documents and to represent the accumulation of investor sentiment over a longer time period and more investors. It is an open research question tackled in Section 4, whether such a measure of investor sentiment relates to higher abnormal stock returns. Another limitation of the portfolio simulation of Chen et al. (2014) is that it did not take transaction costs into account. Because transactions would be created at daily frequency in their portfolio simulation, this would create substantial transaction costs, possibly rendering the positive portfolio level abnormal returns not existent.

The reported evidence in Chen et al. (2014) of a measure of negative words in Seekingalpha investment blog documents to have an effect on portfolio level abnormal returns supports the proposed study of (investor sentiment from) Seekingalpha blog documents in this thesis. In contrast to Chen et al. (2014), this thesis explicitly seeks to classify (investor) sentiment orientation at high accuracy, aggregate it over longer time periods (i.e., monthly), and also take transaction costs into account. Furthermore, this thesis also compares these results to the effects of investor sentiment from Blogpost's documents on abnormal returns (in Section 4).

Trading Strategies to Exploit Blog and News Sentiment

Zhang & Skiena studied relationships between sentiment in documents from blogs, Twitter, and news to returns of stocks (Zhang & Skiena, 2010, p.1). The sentiment is not termed *investor* sentiment by Zhang and Skiena (2010) because their textual sources are not explicitly related to investors and investment topics. The blogs were sourced from Spinn3r

(containing world-wide blogs) and LiveJournal (Zhang & Skiena, 2010, p.3). The news comprise 500 nationwide and local newspapers (Zhang & Skiena, 2010, p.2).

Zhang & Skiena constructed a sentence level sentiment measure by analyzing co-occurrences of words with positive or negative orientation and companies within the same sentence (Zhang & Skiena, 2010, p.3). Godbole et al. (2007) (cited in (Zhang & Skiena, 2010, p.3)) are supposed to describe details of the sentiment analysis approach used by Zhang & Skiena (2010). The model of co-occurrence is simply given by a sentiment word and a company to occur in the same sentence (Godbole et al., 2007, p.221; Zhang & Skiena, 2010, p.3). Godbole et al. (2007, p.221) confess that their approach of associating entities and sentiment words is not always accurate. Zhang & Skiena used a dictionary of almost 5000 sentiment words, which they constructed by expanding synonyms and antonyms of seed words with classifications of either positive or negative (Zhang & Skiena, 2010, p.3). Godbole et al. (2007, p.221) reverse polarities in case of negations in a sentence and also take into account intensifiers (e.g., "very" (Godbole et al., 2007, p.221)) by counting respective sentiment words two times. The document level sentiment is a real-valued ratio of the number p of positive and number n of negative sentence level sentiments: (p–n)/(p+n) (Zhang & Skiena, 2010, p.3). Zhang & Skiena (2010, p.1) claim their sentiment computed from a large dataset from multiple sources covering Twitter, blogs, and news to be more significant and reliable than the one of Tetlock et al. (2008) from DJNS and WSJ. However, the authors don't specify what they exactly mean by significant and reliable.

Zhang & Skiena found correlations between a day's sentiment and stock returns to exist only contemporaneously for news and correlations were almost zero for subsequent future days' returns (Zhang & Skiena, 2010, p.5). Removing companies with weak sentiments was found to increase the correlation (Zhang & Skiena, 2010, p.5). For blogs from Spinn3r a positive correlation on future day 1 for day 0 sentiment and a negative correlation on future day 2 was found (Zhang & Skiena, 2010, p.5). The correlations for Twitter on future day 1 and future day 2 were found highly positive (Zhang & Skiena, 2010, p.5). These results indicate the highest potential for exploitation in a trading strategy for "documents" from the Twitter source in relation to the other sources studied. However, the results of the trading simulation of Zhang & Skiena (2010) do not confirm this.

Zhang & Skiena conducted a stock portfolio simulation with 3238 New York Stock Exchange (NYSE) stocks in the five year period 2005–2009 with daily data by ranking stocks according their sentiment ratio and selecting a specified number of top (and bottom) stocks for going long (and short) and holding them for a specified period (Zhang & Skiena, 2010, p.1). Zhang & Skiena (2010) yielded "consistently favorable returns with low volatility" (Zhang & Skiena, 2010, p.1). Zhang & Skiena (2010, pp.6-7) found (1) returns to be greater for large and small capitalization stocks compared to mid-capitalization stocks, (2) returns to decrease as the holding period increased from one day to ten days, (3) returns to decrease in most years with increasing the historic sentiment analysis period, (4) returns to decrease by the number of stocks held, and (5) higher sentiment levels to be associated with higher

returns. Furthermore, some results indicate yearly returns for Spinn3r blogs to be highest relative to the other sources (i.e., Twitter, LiveJournal, and news) for a small number of held stocks (<6) and for a holding period <3 days (Zhang & Skiena, 2010, Figure 6 & 7 on p.7). However, note that interpretability and comparability of these results of Zhang and Skiena (2010) is impaired by partially using data from different time periods per information source. Furthermore, the first finding of larger returns for large capitalization stocks contrasts other authors' findings, e.g., by Leinweber & Sisk (2011) who found small capitalization stocks to yield highest returns.

The relevant shortcomings of the approach of Zhang & Skiena (2010) with respect to sentiment analysis of companies are identified as follows. The sentiment analysis implementation of Godbole et al. (2007) used by Zhang & Skiena (2010) seems to be not publicly available (e.g., on a website). Thus, replication is at least not straightforward. Neither Zhang & Skiena (2010) nor Godbole et al. (2007) evaluated the accuracy of the sentiment classification approach with respect to a corpus of manually classified documents. Thus, the quality of the sentiment classification is not transparent. Furthermore, there is no detailed description of the datasets used in the portfolio simulation. With respect to this thesis especially the set of blogs, relatedness to investment topics, and quality of these blogs covered by Spinn3r would be interesting.

With respect to studying effects of the sentiment of documents on stock returns, the following shortcomings are identified. In contrast to Tetlock et al. (2008) Zhang & Skiena (2010) only determined correlations of stock returns to contemporaneous returns and future stock returns – they did not conduct a regression analysis. Also, Zhang & Skiena (2010) report some correlations to be significant but do not define the term (i.e., statistical significance at a certain significance level).

Concerning shortcomings of the portfolio simulation, Zhang & Skiena (2010) sorted stocks by the sentiment level but did not study the effect of the level of sentiment on unexpected abnormal returns or abnormal returns (as defined in Section 2.1) – rather, it seems they used raw returns. Statistical significance for the returns (in relation to a benchmark or risk-adjustment) yielded in the portfolio simulation is not provided. Also transaction costs were not considered. Thus, results were not rigorously evaluated and real-world relevance is not directly assessable.

With respect to this thesis, Zhang & Skiena (2010) found supportive indications in the sense that sentiment from blog documents (collected by Spinn3r) was found to yield higher returns than sentiment from news, Twitter, and LiveJournal in their portfolio simulation. That is, the proposed study of investor sentiment from blog documents is suggested to be worthwhile in comparison to other textual sources of investor sentiment. However, Spinn3r retrieves and collects content from all kinds of blogs without a focus on specific domains or topics (Spinn3r, 2015). Also LiveJournal is not investment-specific (Livejournal, 2015). This thesis rather proposes to study *investment*-specific blog documents from platforms such as

the investment specific blog platform Seekingalpha (see Section 2.3.2). It seems reasonable to assume that investor sentiment from investment-specific blog documents have a more substantial effect on abnormal stock returns than sentiment from generic blog platforms. This thesis investigates the assumption in a portfolio simulation. Furthermore, this thesis evaluates portfolio simulation results more rigorously, also taking transaction costs into account.

Widespread Worry and the Stock Market

Gilbert & Karahalios (2010) studied predictive effects of an aggregate measure of anxiety and worry in not investment-specific blog documents regarding stock market returns. Gilbert & Karahalios (2010) constructed an "Anxiety Index" using two machine learning classifiers ((1) a decision tree variant, and (2) a NB (see Section 2.4.3.3) variant) for LiveJournal blog documents and related it to daily S&P 500 log returns. The LiveJournal website hosts documents that discuss topics of daily life and often express moods (Gilbert & Karahalios, 2010, pp.58–59). The binary classifiers (distinguishing anxious and not anxious) were trained on a corpus of almost 13,000 LiveJournal documents from 2004 that represented the anxious class (i.e., they were tagged by LiveJournal users as anxious, worried, nervous or fearful on the document level) and other non-anxious documents (Gilbert & Karahalios, 2010, p.59). A shortcoming is that the size of the set of documents referring to the non-anxious class is not stated by Gilbert & Karahalios (2010). Ten-fold cross-validation (see Section 3.3.1) resulted in a correct classification of anxious documents in 28% and 32% of the cases but false positive rates of 3% and 6% respectively for the two classifiers (Gilbert & Karahalios, 2010, p.59). That is, recall is low for class anxious and precision should be high (see, e.g., (Yang, 1999) for a definition of these metrics). However, relevant metrics such as accuracy and also not all inputs to calculate these metrics are provided by Gilbert & Karahalios (2010).

Gilbert & Karahalios (2010) applied the anxious-classifiers to more than 20 million blog documents from the LiveJournal website from 2008. For aggregation, they define an "Anxiety Index" as the "log-return" (see Definition (2.3)) of the standardized proportion of anxious blog documents from this day's market close differenced by the one of the last trading day's close (Gilbert & Karahalios, 2010, p.60). As anxious classification per period, they used the classifier with the higher proportion (Gilbert & Karahalios, 2010, p.60).

Gilbert & Karahalios (2010) claim to be the first work, "[…] documenting a clear (Granger-causal) link between web-based social data and a broad stock market indicator like the S&P." (Gilbert & Karahalios 2010, p.58). For this kind of "causality" analysis ((Granger, 1969) cited in (Gilbert & Karahalios, 2010, p.61)), Gilbert & Karahalios created two predictive linear regression models for a day's first difference (i.e., subtracting this day's value from the last day's value) of S&P 500 log returns using the three previous days' first difference of respectively (1a) volatility, (1b) logarithm of the trading volume, and (1c) S&P 500 log returns as independent variables, and (2) in addition to (1) also the three previous days' "Anxiety Index" as independent variables (Gilbert & Karahalios, 2010, pp.61–62). Gilbert & Karahalios (2010, p.62) found the second predictive model that includes the

"Anxiety Index" to perform significantly better than the first model. Gilbert & Karahalios (2010, p.58) found their "Anxiety Index" to predict falling prices in the S&P 500 stock index over 174 trading days in 2008. Gilbert & Karahalios found that "[…] a one standard deviation increase in the Anxiety Index corresponds to 0.4% lower returns (actual returns, not log-returns)." (Gilbert & Karahalios, 2010, p.64). Presumably, Gilbert & Karahalios (2010) mean next day returns – however, it is not explicitly mentioned.

With their "Anxiety Index", which captures some negative moods of blog posters and predicts S&P 500 stock index returns, Gilbert & Karahalios (2010) support findings of Tetlock et al. (2008) who found their measure of negative words to contain predictive value regarding unexpected abnormal stock returns. Depending on the availability of the LiveJournal training corpus used by Gilbert & Karahalios (2010), the approach would be reproducible.

A major limitation, also mentioned by Gilbert & Karahalios (2010), is the usage of data with respect to parts of the year 2008 only, which is well-known for periods of abnormally high volatility and substantial negative stock (index) returns during the financial crisis. Thus, studying the effects of the "Anxiety Index" on stock (index) returns in other market periods would help in creating a more complete understanding.

Gilbert & Karahalios claim their "Anxiety Index" to capture the mood of people and their emotional state (Gilbert & Karahalios, 2010, p.64). Thus, the "Anxiety Index" of Gilbert & Karahalios (2010) could be loosely interpreted as a measure of negative sentiment, which relates it to this thesis. However, it is not directly comparable to sentiment measures because it captures only some specific aspects of negative emotions.

The study of Gilbert & Karahalios (2010) found predictability of stock index returns based on a form of negative sentiment gathered from blogs. This finding provides support for the proposed study of effects of investor sentiment in blog documents on abnormal returns. However, Gilbert & Karahalios (2010) address only a blog platform hosting documents of generic and broad topics, which are not necessarily related to financial markets. In contrast, this thesis proposes to study investment-specific blog documents and thus also *investor* sentiment with respect to specific individual stocks.

In contrast to Tetlock et al. (2008), Gilbert & Karahalios (2010) did not predict individual stocks' unexpected abnormal returns and did not predict abnormal returns of portfolios but rather (raw) returns of a stock index. To this respect, this thesis studies the portfolio level effects of investor sentiment from blog documents on abnormal returns.

2.6 Summary of the Research Gap

Table 1 provides a descriptive overview of the approaches reviewed in Section 2.5.

Table 1: Overview of approaches to study effects of investor sentiment on stock returns.

Authors	Invest. sent.	Own corpus	Accuracy	Invest. blogs	Abn. ret.	Fre-quency	Time span	Stocks
News								
Schumaker et al. (2012)	Yes	Un-known	Unknown	N/A	No	Minutes	5 weeks	S&P 500 stocks
Tetlock et al. (2008)	Yes, only negative	N/A	Unknown	N/A	Yes	Daily	1980–2004	S&P 500 stocks
Antweiler & Frank (2006)	No	Yes	Unknown	N/A	Yes	Daily	1973–2001	Mostly large cap. stocks
Leinweber & Sisk (2011)	Yes	No	Unknown	N/A	Yes	Daily	2004–2009	S&P 1500 stocks
Stock message boards								
Antweiler & Frank (2004)	Yes, only bullish-ness	Yes	88.1% (on the training set)	N/A	No	Daily	Year 2000	DJIA and DJICI stocks
Das & Chen (2007)	Yes	Yes	<70%	N/A	No	Daily	2 months of 2001	MSH stocks
Twitter								
Bollen et al. (2011)	Yes	N/A	Unknown	N/A	No	Daily	~10 months of 2008	DJIA stock index
Blogs								
O'Hare et al. (2009)	Yes	Yes	75.1%	Financial blogs	N/A	N/A	N/A	S&P 500 stocks
Fotak (2007)	Yes	N/A	N/A	Seeking-alpha	Yes	Daily	2006	Mostly large cap. stocks
Chen et al. (2014)	Yes, only negative	N/A	Unknown	Seeking-alpha	Yes	Daily	2005–2012	7422 stocks
Zhang & Skiena (2010)	No	N/A	Unknown	Generic topic blogs	No	Daily	2005–2009	NYSE stocks
Gilbert & Karahalios (2010)	No, only negative	No	Unknown	Generic topic blogs	No	Daily	~10 months of 2008	S&P 500 index
This thesis	Yes	Yes	79.2%	Seeking-alpha, Blogspot	Yes	Monthly	2007-2011	DJIA stocks

NOTES per column:
- "Invest. sent.": Did the study use a measure of investor sentiment used as defined in Section 2.2.2?
- "Own corpus": Did authors create their own corpus for evaluation and/or training of a classifier? "N/A": authors did not necessarily require a corpus. "No": a corpus by others was used.
- "Accuracy": Did the study evaluate the accuracy of the investor sentiment classifier?
- "Invest. blogs": Did the study investigate investment blogs? From which platform?
- "Abn. ret.": Did authors study effects on (unexpected) abnormal returns?
- "Frequency": What is the frequency of the data used in the study of effects on (various forms of) returns?
- "Time Span": What is the time period of the data used in the study of effects on returns?
- "Stocks": Which specific stocks or stock indexes were used in the study of effects on returns?

From the review and Table 1, the following specific gaps are identified in the literature.

Explicit Investor Sentiment Classification

With respect to blogs, two studies were based on general-topic blogs and extracted general sentiment (Gilbert & Karahalios, 2010; Zhang & Skiena, 2010). That is, the sentiment does not necessarily relate to investment topics. Fotak (2007) classified individual stock recommendations in investment blog documents. However, the study of Fotak (2007) dates back to 2006, used only 500 documents, and did not employ a large-scale investor sentiment classification of blog documents over several years. Furthermore, the study of Chen et al. (2014) does not explicitly relate to and did not classify investor sentiment. Thus, there is a research gap of long term studies that are based on an automatic classifier of the sentiment orientation of investor sentiment from blog documents. Unlike prior studies, this thesis focuses on classifying *investor* sentiment from *investment* blog documents.

Corpus of Investment Blog Documents

For evaluating the accuracy of a classifier, a corpus of manually classified documents is required. Unlike for movie review sentiment or subjectivity classification (e.g., (Pang & Lee, 2004)), there is no standard corpus for investor sentiment (in blog documents). The reason might be that creating a corpus is laborious, e.g., O'Hare et al. (2009) found sentiment annotation in blog documents to be difficult. Thus, many authors choose not to create a corpus and rather use a dictionary-based approach or manual classification of investor sentiment (e.g., (Chen et al., 2014; Fotak, 2007; Tetlock et al., 2008)). Consequently, there is a research gap of evaluated classifiers, on which the studies are based. O'Hare et al. (2009) have created one of the first *financial blog* document corpora with classifications of the sentiment orientation of investor sentiment – however, it is not available publicly. Consequently, *a novel corpus* has been designed in the scope of this thesis. Like for the approaches that study investor sentiment in stock message boards (e.g., (Antweiler & Frank, 2004)), this is quite common in domains studied the first time.

Accurate Classifiers

Accuracy of the investor sentiment classifiers is reported only sparsely by studies (i.e., by (Antweiler & Frank, 2004; Das & Chen, 2007; O'Hare et al., 2009)). However, Antweiler & Frank (2004) determined accuracy erroneously and Das & Chen (2007) and O'Hare et al. (2009) did not study effects of investor sentiment on abnormal returns. In studies that use no corpus, there can be of course also no measuring of the accuracy (e.g. (Chen et al., 2014; Tetlock et al., 2008; Zhang & Skiena, 2010)). Even in studies that use a corpus, sometimes accuracy is not reported – but only some other metric (e.g., in (Gilbert & Karahalios, 2010)). Thus, there is a research gap in all reviewed approaches that study effects on abnormal returns concerning the suitability of the investor sentiment classification for the study. Therefore, this thesis *rigorously evaluates the investor sentiment classification performance* on a corpus using the standard metric accuracy (Sokolova & Lapalme, 2009).

Regarding the design of an *accurate classifier* of the sentiment orientation of investor sentiment in blog documents, O'Hare et al. (2009) proposed a machine learning-based approach, which achieved about 75% accuracy. This thesis also pursues a machine learning-based approach because it has been shown to be usually highly accurate (see Section 2.4.3). O'Hare et al.'s accuracy serves as baseline accuracy because the accuracies of the other studies on blogs are either unknown (as discussed above) or are presumably lower: Chen et al. (2014) and Zhang & Skiena (2010) used a dictionary-based approach for classification, which has been indicated to usually have lower accuracies compared to machine learning-based approaches (see Section 2.4.1). O'Hare et al.'s baseline level of accuracy seems reasonable because investor sentiment classification in web documents is a hard problem due to ambiguity (Das & Chen, 2007). To address the research gap of designing an accurate classifier by using a machine learning-based approach, this thesis uses a SVM approach (see Section 2.4.3.5). To choose, the *C*-parameter, which influences the accuracy, experiments were conducted (see Section 3.2.2). Further parameters that influence the accuracy are embodied in the document-vector-transformation (see Section 2.4.3.2). The settings of these parameters were chosen based on the literature (see Section 3.2.1).

Investment Blog Dataset: Seekingalpha vs. Blogspot

Almost no study covers an *investment blog dataset*. An exception is the recent work of Chen et al. (2014) who use a several year Seekingalpha investment blog document dataset. However, the studies on findings of effects of investor sentiment from blog documents mostly refer to a single blog platform (i.e., Fotak (2007) and Chen et al. (2014) refer to the Seekingalpha blog platform, and Gilbert & Karahalios (2010) refer to the LiveJournal blog platform). Zhang & Skiena (2010) might be an exception because they use datasets from LiveJournal and Spinn3r. However, LiveJournal is not investment specific and Spinn3r consists of blog documents from many sources (Spinn3r, 2015). However, these sources are not transparent. Thus, it is not possible to trace back any effects to a certain blog platform. In contrast, this thesis studies effects of investor sentiment from *investment blog documents* from *two specific blog platforms*: (1) Seekingalpha, and (2) Blogspot. This thesis is one of the first to study effects of investment blog documents from Blogspot. Furthermore, this thesis compares (magnitudes and statistical significance of) effects related to the two blog platforms.

Effects on Abnormal Returns

The findings on effects of investor sentiment from textual content on returns are with respect to various forms of returns (i.e., total returns of stock indexes, total returns of stocks, unexpected abnormal returns of stocks, and abnormal returns of stock portfolios) in prior studies. There is substantial evidence for unexpected abnormal returns of individual stocks on time horizons of up to 60 days (e.g., (Leinweber & Sisk, 2011)). The evidence for unexpected abnormal returns indicates price drift and that investor sentiment takes its time to be incorporated into prices. This evidence corroborates predictions of behavioral finance

theory (see Section 2.2). There is also some evidence for unexpected abnormal returns of stocks related to Seekingalpha investment *blog* documents containing long and short stock recommendations (Fotak, 2007). However, Fotak (2007) did not study portfolio level effects of investor sentiment on abnormal returns in various market phases. Some kind of evidence in this direction is provided by Chen et al. (2014), without explicitly relating to investor sentiment. Two studies have evidenced effects of sentiment from blog documents on stock indexes (Gilbert & Karahalios, 2010; Zhang & Skiena, 2010), thus they support the proposed study of effects of investor sentiment from blog documents. However, in contrast to these studies, this thesis studies effects on *abnormal* returns of stock portfolios. This thesis also considers transaction costs, which is in contrast to most prior studies except for Tetlock et al. (2008) and Leinweber & Sisk (2011).

Monthly Frequency

All reviewed studies of effects on (abnormal) returns focus on effects based on daily (or higher) frequency data. That is, investor sentiment is typically aggregated into a daily score and effects on (abnormal) returns are studied on the following day(s). Thus, there is a research gap with respect to higher aggregates and longer term effects. Unlike previous studies, this thesis aggregates investor sentiment into a monthly investor sentiment index and studies effects on abnormal returns on the portfolio level at *monthly frequency*. Thus, the aggregate accumulates investor sentiments over a longer time period and also a larger number of investors with the benefit of potentially reducing noise. The monthly aggregates of investor sentiment are assumed to have long term effects (i.e., at least one month into the future) based on predictions of behavioral finance theory (see Section 2.2). The monthly frequency is also related to mutual fund performance evaluation (e.g., (Carhart, 1997)). Thus, such effects would be relevant for exploitation in a fund context.

Multiyear Time Span

The time span of the datasets used in the studies with respect to news content is typically quite long and can stretch over several decades. However, when studying web information content (i.e., stock message boards, Twitter posts, and blogs), the datasets are typically much shorter, i.e., less than a year. Datasets spanning less than one year might come with the problem of covering only a distinct market phase such as the financial crisis year of 2008 (e.g., (Bollen et al., 2011; Gilbert & Karahalios, 2010)). The datasets of Zhang and Skiena (2010) cover longer time periods. However, several datasets are used by Zhang and Skiena (2010) for different textual sources, covering a maximum period of three years only for web content (i.e., Twitter posts and blogs). In contrast, the datasets used in the study of this thesis cover a *five year period* including different market phases to allow for a meaningful portfolio simulation to study effects on abnormal returns of a portfolio of stocks.

Specific Stocks

Some (micro) blog-related studies focus on effects of sentiment on the level of a stock index (Bollen et al., 2011; Gilbert & Karahalios, 2010). That is, they do not relate to specific, individual stocks. The study of Fotak (2007) is on the stock level but only with respect to mentions in 500 blog documents. The related work of Chen et al., 2014 (Zhang & Skiena, 2010) did not restrict their datasets on specific stocks as their datasets contain documents referring to 7422 (3238) different stocks. Thus, their daily aggregate of their measure of negative words or investor sentiment is presumably based on a few (blog) documents only for most of the stocks because they are simply not very well known. The low number of (blog) documents may result in a low quality (i.e., noisy) measure. The investor sentiment index proposed in this thesis would be of even lower quality because it aggregates a document level measure of investor sentiment instead of a word level one. Thus, this thesis uses different datasets of blog documents that were specifically retrieved from Seekingalpha and Blogspot to refer to large capitalization DJIA stocks, for which there should the highest number of blog documents on these blog platforms.

3 Design of a Classifier for Investor Sentiment in Blogs

This section describes a novel classifier for determining the sentiment orientation of investor sentiment in blog documents. The design comprises setting the parameters of the document-vector transformation and the learning machine. For training of the classifier, a new corpus specific for investment blog documents is proposed.

3.1 Corpus

Designing a corpus is common practice for new domains but also for not so new domains for which corpora are often not disclosed to the public. Central to corpus design are the sampling and annotation approaches, which are discussed in this section for selecting the approaches to be used in this thesis. Then, the corpus is described and its properties are analyzed. Finally, the quality of the corpus is evaluated.

3.1.1 Sampling and Annotation Approaches

The annotation approach has to make sure that the corpus fits its research purpose (see also (Biber, 1993)). The role in the research process of this thesis' corpus is to enable training and evaluating a machine learning classifier with high accuracy for the sentiment orientation of investor sentiment in the text of blog documents. To meet this objective, financial and investment experts are needed to manually annotate the document level sentiment orientation in a set of blog documents. More precisely, the experts are needed to annotate each element of the tuple of an investor sentiment document (see Definition (2.8)) for each stock they identified in a blog document.

To determine the quality of annotations, it is required to have at least two annotators who independently annotate parts or all of (blog) documents in a corpus. On this basis, the level of agreement among the annotators can be measured. For instance, Das & Chen (2007) had two annotators annotate the whole corpus. In the work of Antweiler & Frank (2004), reporting of annotator agreement is missing. In the scope of the creation of this thesis' corpus only parts of the corpus were annotated by more than one annotator for reasons of time and cost efficiency. Still, the parts annotated by more than one annotator allow for estimating the level of agreement.

The overall design objective for the corpus is to be representative of the following population, to which results from using the corpus should generalize: all texts of English language investment blog documents that refer to stocks (as defined in Section 2.3). The sampling frame for the corpus is defined to be full articles from the Seekingalpha blog

platform stemming from the time period 2006 – 2013. The reason for selecting Seekingalpha is that it hosts a large number of (semi-) professional investment blog documents in English language, contributed by many different authors (see Section 2.3.2). The high variety of different articles by different authors should help increasing the representativeness of the population. Confining the sampling to Seekingalpha made sure that only *investment* blog documents were sampled. Also it simplified the retrieval of the blog documents because only one major source had to be used. That is, good representativeness, efficiency and costs have to be balanced on defining the sampling frame (Biber, 1993, p.244). The multiyear time period made sure that multiple market phases are represented in the sample, potentially benefiting the representativeness of blog documents with positive and negative sentiment orientation. The multiyear time period is in contrast to the eight week time period covered in the related corpus of blog documents of O'Hare et al. (2009).

The most important design factor for achieving good representativeness of the corpus with respect to the population is the sampling approach. Basically, two approaches can be followed: (1) simple random sampling, or (2) stratified sampling (Biber, 1993, p.244). Simple random sampling selects each blog document from the population with equal chance (Biber, 1993, p.244), whereas stratified sampling first identifies subgroups (i.e., the "strata" (Biber, 1993, p.244)) in the population and then conducts random sampling within each subgroup (Biber, 1993, p.244). Stratified sampling has been recommended because of achieving higher representativeness (Biber, 1993, p.244). Stratified sampling was also used in the design of this thesis' corpus. Many potential subgroups can be defined: Biber (1993, p.245) proposes the subgroups of "Primary channel", "Format", "Setting", "Addressee", "Addressor", "Factuality", "Purposes", and "Topics". However, considering the research purpose of the design of the corpus of this thesis and the population definition, only the classes that represent the sentiment orientations make up reasonable subgroups. As discussed in Section 2.3.4, the model of investor sentiment document contains only two classes, i.e., a positive class and a negative class. Furthermore, according to that model, multiple stocks and associated investor sentiment can be present in a single blog document (see Definition (2.8)).

Other design factors are given by the subgroup size and the corpus size (Biber, 1993). The relative subgroup size, i.e., the number of investor sentiment document with positive (or negative) sentiment orientation in relation to all investor sentiment document in the corpus, can be determined by (1) proportional sampling (Biber, 1993, pp.247–248), or (2) non-proportional stratified sampling (e.g., (Sudman, 1976) cited in (Biber, 1993, p.254)). Proportional sampling means that the number of investor sentiment document per sentiment orientation should be proportional to their representation in the defined population (Biber, 1993, p.247) to achieve high representativeness ((Henry, 1990; Williams, 1978) cited in (Biber, 1993, p.247)). Proportional sampling has been recommended for corpus design ((Biber, 1993, pp.247,254), referring to (Henry, 1990; Williams, 1978)). However, when the subgroups of the corpus are of primary interest, each subgroup can have equal size ((Sudman, 1976, pp.110–111) cited in (Biber, 1993, p.254)). This approach is termed "non-proportional

stratified sampling" ((Sudman, 1976, pp.110–111) cited in (Biber, 1993, p.254)) and has also been selected for the design of this thesis' corpus. Thus, the vocabulary that represents the subgroups (i.e., classes) of positive and negative sentiment orientations is represented in equal shares in the corpus. By means of the equal shares, the learning machine should be able to discriminate well among the two classes. In contrast, a proportional corpus with unequal shares of the classes could cause the SVM learning algorithm to always vote for the prevailing class (e.g., (Akbani et al., 2004)).

Related to the design of a corpus with two subgroups of equal size in a scenario with subgroups of unequal shares in the population, two basic sampling approaches for creating subgroups of equal size can be followed: (1) undersampling the prevailing subgroup, or (2) oversampling the minority subgroup (Akbani et al., 2004). Akbani et al. (2004) discuss problems with undersampling and conclude that oversampling should be better. Based on this finding and because the annotators of this thesis' corpus (who also retrieved the blog documents) reported blog documents with a positive (investor) sentiment orientation to prevail, the oversampling approach was adopted for creating the subgroup representing the negative sentiment orientation of investor sentiment in blog documents within the corpus of this thesis.

For determining the required corpus size, various approaches exist, depending on the research purpose of the corpus. For instance, regarding the research purpose of studying a linguistic feature, a reasonable corpus size can be determined by means of the standard error of the mean of the number of occurrences of the linguistic feature per textual document in the corpus (Biber, 1993, p.248). With an increasing corpus size, the standard error is expected to decrease, indicating convergence of the mean in the corpus to the mean in the population (Biber, 1993, p.248). Regarding the research purpose of this thesis, the accuracy of the classifier trained on the corpus is central. Statistical learning theory suggests that the accuracy (on the test set) of a consistent classifier increases and converges to the maximum possible accuracy on increasing the number of training examples up to infinity (Schölkopf & Smola, 2002, p.131; Vapnik, 2000, pp.35–38, Figure 2.1). Thus, on increasing the corpus size, the accuracy of the trained classifier should increase. Experimental evidence corroborates this prediction regarding sentiment classification in (blog) documents for corpus sizes that were increased up to 1800 training examples (Melville et al., 2009, pp.1279–1280). However, the increase in accuracy per unit increase of the number of training examples typically gets smaller with an increasing number of training examples (Melville et al., 2009, p.1280). Furthermore, the creation of training examples by human annotation is laborious. Thus, manually created corpora in the literature are of limited size. For instance, Antweiler & Frank (2004) created a corpus of 1000 messages from stock messages boards and O'Hare et al. (2009) created a corpus of 979 financial blog documents. However, also smaller corpora have been created and used in the financial domain. For instance, a classifier was trained on a corpus of 423 corporate disclosures and evaluated successfully (Groth & Muntermann,

2011). The size of the corpus of this thesis falls in between and comprises 638 blog documents.

Other potential design factors of a corpus that do not relate to the research purpose of the corpus of this thesis were neglected. For instance, design factors related to linguistic variation and linguistic features in textual documents, demographic variation among authors, and length of textual documents (Biber, 1993) were neglected.

3.1.2 Corpus Description and Analysis

The corpus of this thesis comprises two document sets of investment blog documents from the Seekingalpha blog platform: (1) the first set was sampled and annotated by two undergraduate students of economics based on the sampling and annotation approach described above, and (2) the second set was annotated by four investment industry professionals according to a related annotation approach (Klein et al., 2013, p.701). All annotators possessed a good command of English.

Document Set 1

The first document set comprises 527 distinct blog documents from Seekingalpha and was created in the scope of this thesis. Prior to starting with the blog document annotations, both annotators received a training to ensure consistent annotations. The annotation of this document set was conducted according to this thesis' model of investor sentiment (see Section 2.3.4). That is, annotators assigned a positive or a negative label for each stock in a blog document, representing the document level (investor) sentiment orientation. Only the text (from the title and the body) of blog documents was used and all non-textual content (e.g., images), advertisements, and comments were discarded. Furthermore, all disclosure and disclaimer sentences, links, references to charts and figures, the publication date, and author mentions were removed. All these elements should not be relevant for creating a text-based classifier with high generalization ability.

The blog documents to be annotated were selected from different time periods to cover different market phases (see below for a detailed analysis). Most of the blog documents refer to at least one DJIA-stock. The focus on DJIA-stocks correlates with the portfolio simulation using investor sentiment from blog documents in Section 4, which also focuses on DJIA-stocks. However, also other stocks were included (see below for a detailed analysis). In fact, the investor sentiment document of *all* stocks contained in a blog document were annotated in case there was an explicitly and distinctly (by at least some sentences) expressed one. The annotation of all stocks allows analyzing the number of blog documents with multiple (investor) sentiment orientations.

For evaluating the quality of the annotations (see Section 3.1.3), a part of the blog documents was annotated independently by both annotators. The other blog documents were annotated each by one annotator. For blog documents annotated by both annotators, only the

annotation of the first annotator was included respectively in the corpus for training and evaluating this thesis' classifier.

Document Set 2

The second document set comprises 111 distinct blog documents from Seekingalpha sourced from Klein et al. (2013). A subset of blog documents from *Seekingalpha* only was used from the new corpus parts developed and described in Klein et al. (2013, p.701, part (2) and (3)). These blog documents had been annotated according to an investor sentiment annotation schema with five levels of positive and negative (investor) sentiment orientation (Klein et al., 2013, p.701). To obtain positive-or-negative (investor) sentiment orientations (according to this thesis' model of investor sentiment, see Section 2.3.4 and Definition (2.8)), the transformation proposed by Klein et al. (2013) was adopted. For evaluating the quality of the annotations, a part of the blog documents had been annotated each by four persons (Klein et al., 2013, p.701). The rest of the blog documents had been annotated each by one (randomly chosen) person (from the four persons) (Klein et al., 2013, p.701). The investor sentiments in the second document set refer to U.S. or EU stocks (Klein et al., 2013, p.701).

Analysis of the Overall Corpus

The overall set of annotated documents comprises the two sets described previously. Table 2 provides a descriptive overview of all annotated documents. Each document can have multiple document level annotations of investor sentiment document referring to different stocks. From this overall set of annotated documents, two corpora for training and evaluating this thesis' classifier of investor sentiment in blog documents (in Section 3.3) were constructed: (1) Corpus A, consisting of all blog documents, and (2) Corpus B of blog documents with one sentiment orientation annotation only (possibly annotated with respect to multiple stocks in the same blog document). For both corpora, one stock per blog document was selected randomly because this thesis' document level classifier does not differentiate stocks within a blog document. Both corpora were constructed with equal shares of the number of blog documents of positive vs. negative sentiment orientation. The balancing of positive vs. negative required randomly removing three blog documents from Corpus B. Table 2 provides a descriptive overview of the final corpora.

The overall set of all annotated documents comprises 638 distinct blog documents from Seekingalpha, of which 527 (82.6%) originate from document set 1, which was created in the scope of this thesis. Document set 1, in which annotators were supposed to annotate the investor sentiment document for all stocks for which an investor sentiment was explicitly and distinctly expressed by the author of the respective blog document, contains only 40 blog documents with multiple sentiment orientations (7.6% of the blog documents in document set 1). That is, most blog documents were annotated only with a single (investor) sentiment orientation. It is possible, however, that blog documents annotated with a single (investor) sentiment orientation contain multiple stocks for which the same (investor) sentiment orientation was annotated. Occurrences of such blog documents seem plausible because

many blog documents from Seekingalpha are titled like "5 Stocks to Buy" (e.g., "5 Stocks Under $10 To Buy With Solid Fundamentals and Upside" (Clark, 2013)).

Table 2: Overview and descriptive analysis of the Seekingalpha blog document corpus, annotated with document level stock-specific investor sentiment orientations.

Number of blog documents	Document set 1	Document set 2	Total
Annotated document set			
in total	527	111	638
with one sentiment orientation	487	104	591
with multiple (>1) sentiment orientations	40	7	47
Corpus A: all blog documents			
in total, with one stock selected	527	111	638
with positive sentiment orientation	248	71	319
with negative sentiment orientation	279	40	319
Corpus B: blog documents with one sentiment orientation			
in total, with one stock selected	485	103	588
with positive sentiment orientation	228	66	294
with negative sentiment orientation	257	37	294

NOTES: The set of all annotated documents comprises two parts: (1) a novel contribution from this thesis, and (2) sourced from Klein et al. (2013). Using subsets of these annotated documents, Corpus A and Corpus B were constructed for training and evaluating a classifier.

The 92.4% blog documents in the document set 1, annotated with a single (investor) sentiment orientation, indicate creating and using a document level classifier of the (investor) sentiment orientation in Seekingalpha blog documents should have a low error rate for the stock-specific classification task related to the possibility of multiple stocks being referred to in such blog documents. This prediction assumes the document set 1 to be a representative sample of Seekingalpha blog documents with respect to the aspect of the number of stocks referred to in the blog documents. A document level classifier does not differentiate multiple (investor) sentiment orientations regarding possibly multiple stocks contained in a blog document. The document level classification task simplifies the creation of the classifier. Experiment results in Section 3.3.2 demonstrate that the document level investor sentiment classifier for blog documents has a reasonable accuracy regarding the overall Corpus A, which includes the blog documents annotated with multiple (investor) sentiment orientations. The blog documents of Corpus A are listed in Table 57 and Table 58 in the Appendix A.2. For experiments on classification accuracy, a second corpus, i.e., Corpus B, was created, which includes only blog documents annotated with a single (investor) sentiment orientation. The 588 blog documents of Corpus B are listed in Table 57 in the Appendix.

The overall Corpus A for training and evaluating the classifier contains 319 positive (investor) sentiment orientations and 319 negative (investor) sentiment orientations. Corpus B contains 294 positive (investor) sentiment orientations and 294 negative (investor) sentiment orientations. That is, each blog document in Corpus A and Corpus B contains only

one (investor) sentiment orientation with respect to one stock that was selected randomly in case more than one were annotated. The resulting distribution of positive and negative (investor) sentiment orientations is deliberately 50-to-50 percent according to the defined sampling approach in the previous section, thus benefitting the machine learning approach for creating a classifier.

The blog documents of the corpora were sampled from multiple time periods, ranging over several years, with a peak around 2011 and 2012. This multiyear sampling is to represent diverse stock market phases of prices trending up, down, or sideways. Also, a large variety in vocabulary can be represented. See Table 3 for an overview of the covered time periods and the respective number of blog documents.

Table 3: Time period analysis of the blog documents' publication dates in Corpus A. For each half-yearly time period, the approximate price trend of the DJIA and the number of blog documents covered by the respective period are listed.

Time period	DJIA price trend	Number of blog documents
2006-01-01 – 2006-06-30	Up	1
2008-01-01 – 2008-06-30	Down	32
2008-07-01 – 2008-12-31	Down	23
2009-01-01 – 2009-06-30	Sideways	53
2009-07-01 – 2009-12-31	Up	23
2010-01-01 – 2010-06-30	Sideways	48
2010-07-01 – 2010-12-31	Up	33
2011-01-01 – 2011-06-30	Up	47
2011-07-01 – 2011-12-31	Sideways	160
2012-01-01 – 2012-06-30	Sideways	96
2012-07-01 – 2012-12-31	Sideways	102
2013-01-01 – 2013-06-30	Up	20

The length of (the body of) blog documents in the overall Corpus A is presented in Table 4 with the mean (median) number of words per blog document being 664 (579). 90% of the blog documents in the corpus contain more than 263 words. That is, the Seekingalpha blog documents are rather long documents with a substantial amount of content, leaving space for discussions and arguments. Due to the length of the blog documents, SVM is well suited as a potentially highly accurate machine learning approach for creating the investor sentiment classifier for blog documents ((Wang & Manning, 2012), and Section 2.4.3.5).

77.7% of the investor sentiment annotations in Corpus A refer to DJIA stocks (as of 2009-06-08 (S&P Dow Jones Indices LLC, 2013)). 29 of 30 DJIA stocks (as of 2009-06-08 (S&P Dow Jones Indices LLC, 2013)) and all DJIA stocks (as of 2012-09-24 (S&P Dow Jones Indices LLC, 2013)) are represented. See Table 59 in the Appendix for a complete list of the stocks to which the investor sentiment annotations refer to. In total there are 132 different stocks to which the annotations refer to. That is, the annotations cover quite a large variety of different stocks, helping to capture potential differences in vocabulary in the blog documents. Note that 2 of the annotations refer to the S&P 500 stock index (i.e., not a single stock).

Table 4: Body length analysis of blog documents in Corpus A with 638 blog documents in total in terms of number of characters and number of words.

Statistical measure	Characters per blog document	Words per blog document
Mean	3,972	664
Standard deviation	2,430	400
Minimum	307	62
Maximum	26,942	4,273
10% percentile	1,597	263
20% percentile	2,045	345
30% percentile	2,589	435
40% percentile	2,965	500
50% percentile	3,459	579
60% percentile	4,009	672
70% percentile	4,753	798
80% percentile	5,526	939
90% percentile	6,644	1,110

3.1.3 Evaluation

Prior to using the corpus of investor sentiment in blog documents for training and evaluating machine learning classifiers, the quality of the annotations in the corpus is evaluated. For this, the widely-used inter-rater-agreement metric kappa (Fleiss, 1971), which extends on Cohen (1960), is employed. In this thesis, Fleiss' kappa is used to measure the extent of agreement beyond chance between multiple raters (i.e., annotators) of the same blog documents regarding the classification of the blog documents in classes positive or negative sentiment orientation. Regarding interpretation of the real-valued kappa: a kappa of zero means agreement by chance, negative kappa values mean less than chance agreement, positive kappa values mean greater than chance agreement, and kappa=1 means perfect agreement (Cohen, 1960, p.41).

To compute Fleiss' kappa, 216 investor sentiment annotations in the set of all annotated blog documents were annotated by more than one annotator. In the first document set, 162 stock-specific investor sentiments were independently annotated each by 2 annotators. In the second document set, 54 stock-specific investor sentiments were independently annotated each by 4 annotators.

Table 5 lists Fleiss' kappa for both document sets. Kappa is approximately at the same level (0.775 and 0.733) for both document sets. Kappa values greater than 0.75 were proposed to be interpreted as "[…] excellent agreement beyond chance […]" (Fleiss et al., 2003, p.604). That is, the inter-annotator-agreement is reasonably high (considering the general ambiguity in web texts (e.g., (Das & Chen, 2007))) and the overall quality of the corpora according this (interpretation of the) metric is very good. Thus, this thesis' corpora can be regarded as well suited for training and evaluating a classifier for investor sentiment in blog documents.

Table 5: Evaluation of annotations of stock-specific investor sentiment in both document sets, annotated each by multiple annotations, by means of Fleiss' kappa measure of inter-annotator-agreement.

Document set	Number of investor sentiment	Number of annotators (raters)	Fleiss' kappa
1	162	2	0.733
2	54	4	0.775

3.2 Classifier

Designing the supervised machine learning classifier for investor sentiment in blog documents means defining and configuring (1) the transformation of the textual blog document to a vector representation (see Section 2.4.3.2) according to the vector space model (see Section 2.4.3.1), and (2) the supervised machine learning (see Section 2.4.3.2) approach SVM, which trains a classifier using the vector representation of training examples from the corpus. The configurations were informed by the literature (see Table 6) with the objective of maximizing the accuracy. The following subsections detail the configuration decisions.

Table 6: Summary of parameter configurations for the classifier of investor sentiment in blog documents and the document-to-vector-of-features-transformation, which provides the input for the SVM-based classifier.

Parameter	Configuration	Informed by
Feature definition		
Text representation	{unigrams, unigrams & bigrams, unigrams & bigrams & trigrams}	(Ng et al., 2006; Sebastiani, 2002; Wang & Manning, 2012)
Tokenization	Rule-based, distinguishing words, numbers, symbols, punctuation, and spaces.	
Feature selection and feature extraction		
Feature selection	No	(Joachims, 1998; Mejova & Srinivasan, 2011)
Stop word removal	No	(Leopold & Kindermann, 2002)
Stemming or lemmatization	Lemmatization	(Mejova & Srinivasan, 2011; Mullen & Collier, 2004)
Feature weighting		
Weighting	Binary	(Pang et al., 2002; Wang & Manning, 2012)
Length normalization	No	(Wang & Manning, 2012)
SVM configuration		
Non-linear transformation of input vectors (see Section 2.4.3.4)	No	(Hsieh et al., 2008, p.408; Ng et al., 2006, p.614)
Loss function (Definition (2.26))	L2	(Wang & Manning, 2012, p.91)
C-parameter (see Section 2.4.3.4)	$C=2^i$, where $i \in \{-5, -4, \ldots, 20\}$	(Hsu et al., 2010, p.5; Joachims, 2006)

3.2.1 Document-Vector Transformation

The purpose of the document vector transformation is to transform a textual blog document into a numerical vector representation (see Section 2.4.3.2). Whereas the textual parts

represented by components of such vectors are termed either “terms” or “features” in the literature (e.g., (Sebastiani, 2002, p.10)), they are termed features in this thesis. The transformation of a textual document to a vector of features comprises (1) defining the features based on basic text parts, (2) selecting features to reduce the dimensionality of the vector space, and (3) weighting of features to assign numbers to features.

3.2.1.1 Feature Definition

In this thesis, features are the textual parts (or linguistic units) that constitute components in document vectors (e.g., (Fürnkranz, 1998, p.1; Joachims, 1998, p.138; Leopold & Kindermann, 2002, p.423)). Principally, features can be defined on several levels of a textual document: the sub-word level, the word level, the multi-word level, the semantic level, and a pragmatic level (Joachims, 2002, p.12). Text classification commonly resorts to the word level (e.g., (Joachims, 2002, p.13)). Similarly, this thesis uses: (1) a text representation on the (multi-) word level for defining features, and (2) a “tokenization” method to identify words in English language text.

Text Representation

Following text classification literature based on machine learning approaches (examples include (Fürnkranz, 1998; Mladenic & Grobelnik, 1998)), this thesis uses word n-grams according to the following definition to represent text:

Definition: A **Word N-Gram** is a sequence of n words (e.g., (Fürnkranz, 1998, p.1; Mladenic & Grobelnik, 1998, pp.145–146)), where $n \in \{1,2,3\}$.

A word 1-gram is also termed “unigram”, a word 2-gram is also termed “bigram”, and a word 3-gram is also termed “trigram” (e.g., (Fürnkranz, 1998; Ng et al., 2006)). See Table 7 for an example sentence represented by different word n-gram types.

Table 7: Word n-grams for an example sentence (“I recommend buying the stock.”).

Word n-gram type	Examples
Unigram	“I“, “recommend“, “buying“, “the“, “stock“
Bigram	“I recommend“, “recommend buying“, “buying the“, “the stock“
Trigram	“I recommend buying”, “recommend buying the”, “buying the stock”

Single words (i.e., unigrams) have been found to be more effective for machine learning *text classification* than more complicated text representations (e.g., (Dumais et al., 1998)). Therefore, unigrams are typically used as features for text classification (e.g., (Sebastiani, 2002, p.10)).

Word n-grams with $n>1$ can provide additional information to unigrams, potentially helping to improve classification. A general problem is that such word n-grams occur less often in a corpus and some word n-grams might occur in the training set but not in the test set (e.g., (Ng et al., 2006, p.617)). Thus, such word n-grams might be effective only for large corpora (e.g., (Manning et al., 2009, p.241)). Furthermore, using word n-grams with $n>1$ in addition to unigrams increases the number of features, and thus the dimensionality increases

(e.g., (Ng et al., 2006, p.615)). However, high dimensionality can be handled very well in principle by the SVM approach (Joachims, 1998). To compromise the amount of additional information for the classifier versus the dimensionality, the word n-grams under consideration in this thesis are limited to trigrams.

Regarding document level *sentiment* orientation *classification* of (movie) reviews using SVM adding bigrams to unigrams as features improved the accuracy (Wang & Manning, 2012). Regarding the same datasets, this was not the case for adding trigrams (Wang & Manning, 2012). Different results show that combining unigrams, bigrams, and trigrams as features for sentiment classification increases the accuracy vs. using unigrams only by about 2% (Ng et al., 2006). Ng et al. also found using bigrams or trigrams alone to decrease the accuracy compared to using unigrams alone (Ng et al., 2006, p.617). Based on this evidence, the accuracy of this thesis' investor sentiment classifier for blog documents was evaluated on this thesis' corpus regarding using (1) unigrams, (2) unigrams and bigrams, and (3) unigrams, bigrams, and trigrams as features (see Section 3.3.2 for respective experiments and results).

Tokenization

Tokenization refers to determining the "basic units of a document" from a character stream (Manning et al., 2009, p.19). That is, a token "[…] is an instance of a sequence of characters […] that are grouped together as a useful semantic unit […]" of a particular type (Manning et al., 2009, p.22). A straightforward approach for tokenization identifies tokens by means of spaces (e.g., (Wang & Manning, 2012)). Using a set of more complex rules, the precision of the token extraction might be higher and also different types of tokens can be differentiated: The automatic ANNIE English Tokenizer, which was used in this thesis, uses a set of rules to distinguish tokens of the following types: words, numbers, symbols (e.g., currency symbols), punctuation (e.g., ":", "(" , ")"), and spaces (Cunningham et al., 2011, pp.115–117). The identified tokens serve as unigrams and as the elements of higher n-grams.

3.2.1.2 Feature Selection and Feature Extraction

So far, textual features have been defined to be word n-grams of different length *n*. Combining different word n-grams as features increases the dimensionality of the vector space and also comes at higher computational costs. The "curse of dimensionality" ((Bellman, 1961) cited in (Cunningham, 2007, p.1)) describes the associated machine learning problem of increasingly less accurate classifiers when the number of features increases (Cunningham, 2007, pp.1-2). To decrease the number of features, methods of (1) feature selection, and (2) feature extraction have been proposed (e.g., (Sebastiani, 2002, pp.13–18)).

Feature Selection

Methods for automatic feature selection have been reviewed and evaluated, for instance, by (Yang & Pedersen, 1997). For example, features that appear infrequently in the corpus or

uninformative features according to some metric can be removed (e.g., (Mejova & Srinivasan, 2011; Yang & Pedersen, 1997)). Such methods were not used in this thesis because SVMs work well in high dimensional feature spaces and do not require feature selection (Joachims, 1998). Not requiring feature selection seems reasonable because regarding sentiment orientation classification using SVM, several feature selection methods did not improve the accuracy with respect to the baseline configuration that was not using feature selection (Mejova & Srinivasan, 2011). Not using feature selection methods includes not removing stop words in this thesis. Stop words are "extremely common words" that occur the most often in a set of documents (Manning et al., 2009, p.27). However, regarding text classification with SVMs, removing stop words is not required to improve accuracy (Leopold & Kindermann, 2002, p.442).

Feature Extraction

Using stems or lemmas instead of words in the word n-grams defined as features above can be regarded feature extraction because the original features (i.e., words) are transformed (Cunningham, 2007, p.3). A stem is heuristically derived by cutting the endings of words to reduce the number of inflections (Manning et al., 2009, p.32). A lemma is the base form of a word with inflectional endings removed based on "[…] a vocabulary and morphological analysis […]" (Manning et al., 2009, p.32). For instance, the stem and lemma of "buys" and "buying" might be both "buy". However, the lemma of "saw" (as a noun) is "saw" and the lemma of "saw" (as a verb) is "see", whereas the stem might be just "s" (Manning et al., 2009, p.32). Via both methods, the number of distinct words in a document or corpus is reduced, and thus the (potential) dimensionality of the vector space shrinks (e.g., (Forman, 2003, p.1292; Leopold & Kindermann, 2002)). Regarding sentiment orientation classification using SVMs, stems did not increase the accuracy (Mejova & Srinivasan, 2011). Regarding the positive/negative sentiment classification using SVMs, lemmas have been shown to increase the classification accuracy slightly (Mullen & Collier, 2004). Thus, lemmas were used instead of words in the word n-grams defined above as features.

To derive lemmas, the GATE Morphological Analyzer was used in this thesis, requiring tokens and part-of-speech tags as input (Cunningham et al., 2011, pp.455–458). Part-of-speech (POS) tagging refers to identifying the lexical syntactic category (e.g., verb, noun, etc.) of each word (e.g., (Hepple, 2000, p.278)). The ANNIE POS Tagger (Cunningham et al., 2011, pp.121–122) was used in this thesis to implement POS tagging according ((Hepple, 2000) cited in (Cunningham et al., 2011, p.121)). This POS tagger requires also the information of sentence boundaries (Cunningham et al., 2011, p.121). Sentence splitting was implemented by the ANNIE Sentence Splitter (Cunningham et al., 2011, pp.119–120). The POS tags were only used to derive lemmas and did not serve as input for machine learning.

3.2.1.3 Feature Weighting

To transform textual features in a numerical vector according to the vector space model (see Section 2.4.3.1), a numerical weight is assigned to each feature. For determining the weight

$w \in \mathbb{R}$ of a feature (representing its presence or importance), several models exist (e.g., (Salton & Yang, 1973) for a comparison). These models originate from the field of Information Retrieval (e.g., (Manning et al., 2009)) and use the term "term" instead of "features" (used in text classification (e.g., (Joachims, 1998; Sebastiani, 2002))). The terms are interchangeable (e.g., (Sebastiani, 2002, p.10)) and some of the most common weighting models are adapted to refer to "features" in the following.

The Boolean model uses binary weights to determine for each feature (or term) whether it occurs at all in a document or not (e.g., (Baeza-Yates & Ribeiro-Neto, 1999, pp.25–27)):

Definition: Boolean Model (e.g., (Baeza-Yates & Ribeiro-Neto, 1999, pp.25–27)):

$$w = \begin{cases} 1, if\ the\ feature\ occurs\ at\ least\ once\ in\ the\ document \\ 0, otherwise \end{cases} \quad (3.1)$$

In the "bag of words" model, the weight of each feature (or term) is determined by the frequency of occurrence of that feature (or term) (e.g., (Manning et al., 2009, p.117)) with the original idea of using the frequency of features (or terms) as weights dating back at least to ((Luhn, 1957), also cited in (Manning et al., 2009, p.133)). Thus, features (or terms) that occur frequently in one document are assigned a high weight:

Definition: Bag of Words Model (e.g., (Manning et al., 2009, p.117):

$$w = ff(ft, bd) \quad (3.2)$$

where

ff $\in \{0,1,2,3,\ldots\}$: Feature (or term) frequency, i.e., the number of occurrences of a feature (or term) *ft* in a blog document *bd* (adapted from (Salton et al., 1975, p.615)).

Another model "[…] places greater emphasis on the value of a term as a means of distinguishing one document from another than on its value as an indication of the content of the document itself." (Sparck Jones, 1972, p.18). Thus, features (or terms) that occur only in a small number of documents are assigned high weights:

Definition: Inverse Document Frequency (adapted from (Manning et al., 2009, p.118) with the original idea dating back at least to Sparck Jones (1972, pp.17-18), who uses a slightly different definition):

$$w = idf(ft, C) = log_{10} \frac{N}{df(ft,C)} \quad (3.3)$$

Variables in Definition (3.3) are defined as follows:

df(ft,C)$\in\{1,2,3,\ldots\}$: Document frequency, i.e., the number of blog documents *bd* in the corpus *C* that contain the feature (or term) *ft* (adapted from (Manning et al., 2009, p.118)).

C: A set of *N* documents (adapted from (Manning et al., 2009, p.118).

idf$\in\mathbb{R}$: Inverse document frequency.

Combining, i.e., multiplying, *ff* and *idf* leads to the feature frequency – inverse document frequency (*ff-idf*) weighting model (known from IR as *tf-idf*) for features (or terms) (e.g., (Salton et al., 1975; Salton & Yang, 1973) for some early mentions):

Definition: Feature Frequency – Inverse Document Frequency (adapted from (Salton et al., 1975, p.615)):

$$w = ffidf(ft, bd, C) = ff(ft, bd) \cdot idf(ft, C) \tag{3.4}$$

where

ff: Feature (or term) frequency, as defined above.

idf: Inverse document frequency, as defined above.

ffidf$\in\mathbb{R}$: Feature frequency inverse document frequency.

In effect, *ff-idf* (*tf-idf*) assigns large weights to features (or terms) that occur often in single documents and that occur rarely in the whole corpus (Salton et al., 1975, p.615). The normalized *ff-idf* (*tf-idf*) weighting is widely used for machine learning text classification (Sebastiani, 2002, pp.11–12).

Normalization addresses a problem arising with the vectors of weighted features when considering documents of differing length (e.g., (Leopold & Kindermann, 2002, p.428; Salton & Buckley, 1988, p.517)). A long document would contain many words as features, resulting in high feature frequency numbers. A short document may contain the same words but only fewer of them, resulting in low feature frequency numbers. Assuming both documents to belong to the same class, the long document might be more likely (at least under the bag of words model) to be classified correctly. To make **document vectors** of feature weights comparable, they can be **normalized** by the Euclidian length (e.g., (Salton & Buckley, 1988, p.517)). For normalization, each weight of a vector has to be divided by the Euclidian length of the vector (e.g., (Salton & Buckley, 1988, p.518)):

Definition: Euclidian Length of a Document Vector (e.g., (Manning et al., 2009, p.121; Salton & Buckley, 1988, p.518)):

$$el = \sqrt{\sum_{i=1}^{n} w_i^2} \tag{3.5}$$

where

el: Euclidian length of the document vector (e.g., (Manning et al., 2009, p.121)).

$w_i \in \mathbb{R}$: The *i*-th component of a document vector with *n* components in total (e.g., (Manning et al., 2009, p.121)).

For selecting the best feature weighting model regarding *sentiment* orientation classification using SVMs with respect to maximizing the accuracy, evidence suggests the Boolean model to be a good choice for feature weighting (e.g., (Pang et al., 2002; Wang & Manning, 2012)). Thus, in this thesis, the Boolean model was used for feature weighting.

Normalized Boolean weighting was considered for SVM text classification (Joachims, 2002, p.20) and was used for sentiment orientation classification with SVM (Pang & Lee, 2004; Pang et al., 2002, p.82). However, it is unclear what the benefit is in terms of accuracy for normalizing Boolean weights. For instance, Wang and Manning (2012) did not normalize vectors with Boolean weighted features. Wang and Manning (2012) report an accuracy of 86.25% (using SVM, unigrams, Boolean feature weighting, space-based tokenization, and C=0.1 – see Section 2.4.3.4 for an explanation of the *C*-parameter) on the same corpus as Pang & Lee (2004). Pang & Lee report 87.15% accuracy (using SVM, unigrams, and *normalized* Boolean feature weighting) on the same corpus (Pang & Lee, 2004). However, the tokenization method of Wang and Manning (2012) is crude and the corresponding method of Pang and Lee (2004) is not reported, as well as the used *C*-parameter. Due to the unclear benefit, the simple approach of Wang and Manning (2012) was followed in this thesis and length normalization of document vectors was not used.

3.2.2 Support Vector Machine Configuration

SVM was used as a "learning machine" (see Figure 9) to generate the classifier for the investor sentiment document scores of blog documents in vector form using the training examples of this thesis' corpus. The arguments for using SVM have been discussed in Section 2.4.3.5. The configuration of the SVM algorithm to maximize the accuracy of the classifier comprises (1) a transformation function for the input document vector, (2) the loss function, and (3) the *C*-parameter.

The transformation of the input vector (see Section 2.4.3.4) using a non-linear function increases the dimensionality and allows for better separability of multiple vectors in classes ((Schölkopf & Smola, 2002, pp.200–201), also citing (Cover, 1965)). However, regarding document text classification, the dimensionality is already high (e.g., (Joachims, 1998)).

Thus, regarding (topical) document text classification, *not* non-linearly, i.e., linearly, transforming was found to be slightly better (e.g., (Yang & Liu, 1999, p.47)). Regarding the classification of the sentiment orientation of documents, non-linear transformations did not improve results compared to a linear SVM (e.g., (Ng et al., 2006, p.614)). Thus, input vectors were not transformed and "linear SVM" (e.g., (Hsieh et al., 2008, p.408)) was used in this thesis.

The loss function (see Definition (2.26)) is part of the SVM-optimization problem with a soft margin (see Definition (2.25)), which was used in this thesis to generate a classifier for the investor sentiment document scores of blog documents. The objective in the optimization is to minimize the cumulative "losses" due to false classifications (see Section 2.4.3.4). The losses are typically either summed up directly (L1-SVM) or the squared losses are summed up (L2-SVM) (see Definition (2.26)). Regarding sentiment orientation classification, the L2-SVM has been found to work better than L1-SVM (Wang & Manning, 2012, p.91). Based on this evidence, L2-SVM was used in this thesis.

The *C*-parameter (see Section 2.4.3.4) in the SVM-optimization problem with a soft margin (see Definition (2.25)) balances the objectives of (1) minimizing the training error, and (2) increasing the generalization ability of the classifier (see Section 2.4.3.4). The larger the value of the *C*-parameter, the higher the weight is on the first objective (see Section 2.4.3.4). If this objective generates a classifier with lower generalization ability, the error on the test set might be larger (see Section 2.4.3.2). There are only rules of thumb for choosing the *C*-parameter with respect to maximizing the accuracy of the classifier on the test set. For instance, evaluating sequences of $C=2^i$, where $i \in \{-5, -3, \ldots, 15\}$, was proposed (Hsu et al., 2010, p.5). Joachims experimented with *C* values ranging from 100 to 1,000,000 with optimal values for most used data sets lying between 10,000 and 50,000 (Joachims, 2006). Subsuming both proposed ranges for *C*, this thesis' classifier's accuracy was evaluated with respect to configuring the *C*-parameter to values of $C=2^i$, where $i \in \{-5, -4, \ldots, 20\}$. Respective experiments and results are reported in Section 3.3.2.

3.3 Evaluation

The configurations of the document-vector-transformation and of the SVM learning machine, determined in the last section, are subject to evaluation of the accuracy regarding the classification of the sentiment orientation in this section. This section reports on results of an exploratory experimental analysis regarding the best configuration of two parameters, i.e., regarding the text representation and the *C*-parameter of the SVM learning machine, in terms of high accuracy using this thesis' corpus and a cross-validation approach.

3.3.1 Cross-Validation Approach

A cross-validation approach was used for evaluating the classifier of the sentiment orientation of investor sentiment in terms of accuracy and for conducting experiments with respect to

parameter configurations. The general problem with respect to evaluating a supervised machine learning classifier is that examples of human annotated (i.e., labeled) blog documents are required for both: (1) training of the classifier, and (2) evaluating the classifier. These annotated examples are provided by this thesis' corpus designed in Section 3.1. That is, the number of annotated blog documents is limited to the size of the corpus. Because training and evaluation of the classifier must be conducted on separate sets of annotated blog documents, a strategy to annotation-efficiently and accuracy-effectively use these annotated blog documents must be devised.

The *k*-fold cross-validation approach randomly divides the set of annotated blog documents of a corpus in non-overlapping, equally sized *k* folds (e.g., (Kohavi, 1995)). Stratified cross-validation additionally requires each fold to contain the same number of positively vs. negatively labeled blog documents (Kohavi, 1995). Training and evaluation are conducted *k* times: each of the *k* folds is used as a test set once while all other folds are used as training set (e.g., (Kohavi, 1995)). The training set is used to train the supervised machine learning classifier (see Section 2.4.3.2) (e.g., (Kohavi, 1995)). The classifier is then applied to classify the blog documents in the test set (e.g., (Kohavi, 1995)). The fold that the test set is assigned to is then rotated among all folds (e.g., (Kohavi, 1995)). Finally, the classification accuracy estimate can be computed over all correctly classified blog documents in *k* test sets (e.g., (Kohavi, 1995)). In this thesis, k=10 was used, which is common practice in sentiment classification (e.g., (Ng et al., 2006; Pang & Lee, 2004; Wang & Manning, 2012)) and also stratification of the folds was used as proposed by Kohavi (1995). Note that the composition of the folds was conducted only *once* as defined in this section. That is, each experiment (described in the next section) was based on the same folds, i.e., containing the same blog documents.

3.3.2 Experiments and Results

Using the cross-validation approach and this thesis' corpora, experiments for choosing the parameter configurations of the text representation (see Section 3.2.1.1) and the *C*-parameter (see Section 3.2.2) were conducted with the objective of maximizing the accuracy relative to the accuracy of a baseline configuration SVM-classifier. For machine learning experiments, a slightly modified version of the corpora presented in Section 3.1 was used with all company mentions in the body and title of the blog documents exchanged for a neutral "[comp]"-string. This treatment is to make the classifier not use company-words to relate to a sentiment orientation class and rather abstract away from specific companies to potentially increase the generalization ability. For each experiment, a combination of the title and the body of each blog document was used.

Baseline

The baseline configuration comprises the parameter configurations derived in Section 3.2 and summarized in Table 6. Unigrams (i.e., the simplest configuration) were used as baseline

text representation. *C*=1 was used as baseline *C*-parameter configuration. This configuration is the default configuration in the GATE implementation of SVM (Cunningham et al., 2011, p.371), which was used in this thesis. The accuracy of the described baseline configuration using Corpus A is 76.2%. The baseline accuracy is with respect to Corpus A, which contains a mix of blog documents with (one or multiple stocks and) one sentiment orientation and with (multiple stocks and) multiple sentiment orientations of which one was selected randomly for training and evaluation. Such a mix also occurs in out-of-sample blog documents. Thus, the accuracy on Corpus A can be assumed to be a good estimation of the out-of-sample accuracy.

Effect of Blog Documents with Multiple Sentiment Orientations

Two experiments were conducted using a classifier trained on Corpus B, containing blog documents annotated with *one* sentiment orientation only, to study the effect of blog documents with *multiple* sentiment orientations – by comparing the results of the trained classifiers to the results of the baseline classifier. The baseline classifier uses Corpus A, which basically extends Corpus B with blog documents annotated with multiple sentiment orientations. In the first experiment, the blog documents with multiple sentiment orientations were found *not* to decrease the accuracy when training and evaluating on Corpus A instead of training with blog documents of only one sentiment orientation (using Corpus B) and evaluating on Corpus A. In the second experiment, training and evaluating on blog documents with one sentiment orientation (on Corpus B) only yielded a slightly higher than baseline accuracy. Table 8 provides an overview of the results of the experiments.

Table 8: Effects of documents annotated with one vs. multiple sentiment orientations on training and evaluating a classifier for the sentiment orientation.

Experiment	Baseline: Train and test with multiple sent. orientations	Train with one sentiment orientation	Train and test with one sentiment orientation
Training corpus	A	B	B
Testing corpus	A	A	B
Hypothesis	n/a	Accuracy is higher than baseline.	Accuracy is higher than baseline.
Accuracy	76.2%	75.5%	77.0%
Support for the hypothesis	n/a	No support	**Support**

Discussing the experiments in detail in the following, it seems reasonable to hypothesize that training a classifier on Corpus A, containing blog documents with *multiple* sentiment orientations, *reduces* the accuracy due to "distractions" by vocabulary referring not to the target class of a training example. This setup corresponds to the baseline experiment. An alternative formulation of the hypothesis is that training a classifier only on blog documents that were annotated with *one* sentiment orientation (i.e., from Corpus B) yields a *higher* than baseline accuracy. To test the hypothesis, the baseline configuration was used to train a

classifier using Corpus B and evaluate on Corpus A to be able to compare results to the baseline result.

To conduct the experiment, each of the ten test sets of Corpus B, which was used in the 10-fold cross-validation, was enriched with five randomly chosen blog documents that are part of Corpus A but not of Corpus B. The assignment was random but static for the whole cross-validation. 47 of the 50 blog documents that were used in total for enriching the test sets had been annotated with multiple sentiment orientations (of which one was selected for evaluation) – see Table 58 in the Appendix for a list. The remaining three blog documents (see Table 57) had been annotated with one sentiment orientation but are not part of Corpus A for the reason of having the same number of negative vs. positive sentiment orientations in the corpus (see Section 3.1.2). These three blog documents were added to the test sets as well, to make all test sets in total contain the same (number of) blog documents like Corpus A to be able to compare the cross-validation result to the baseline's one. The resulting accuracy of the experiment is 75.5%.

The experiment's accuracy is a bit lower than the baseline accuracy. This result provides an indication for rejecting the hypothesis. That is, no indication was found for a positive (negative) effect of blog documents with one (multiple) sentiment orientation on the accuracy. A possible reason might be the smaller number of training blog documents in Corpus B compared to Corpus A, indicating that the number of training examples is more important with respect to increasing the accuracy. Consequently, Corpus B was not used in the experiments for determining the parameter configurations below. Rather, Corpus A was used for training and evaluation like in the baseline experiment.

Finally, to get an indication of the effect of blog documents with multiple sentiment orientations on the *evaluation*, a second experiment of training and evaluating a classifier on Corpus B, which consists only of blog documents of one sentiment orientation, was conducted. The respective hypothesis is that the accuracy is higher than in the baseline experiment (of training and evaluating on Corpus A) because Corpus B represents the natural configuration for document level classification. The resulting accuracy of 77.0% seems to indicate support for the hypothesis – although the results are not directly comparable because the evaluation corpora are not identical. However, the baseline accuracy, which evaluates a one-sentiment-orientation-per-document-classifier on a corpus containing multiple sentiment orientations, is only slightly lower. This result can be interpreted as a supporting argument for this thesis' document level classification approach, which does not differentiate possible multiple sentiment orientations regarding multiple stocks on one blog document.

Choosing the Text Representation

In Section 3.2.1.1, text representations that are common and suitable for machine learning sentiment classification were presented and discussed. As a baseline text representation unigrams were used, which are a typical choice (e.g., (Sebastiani, 2002, p.10)). Adding higher order n-grams (i.e., bigrams and trigrams) can provide additional information but also

increases the number of features (see Section 3.2.1.1). To choose the best text representation regarding the overall Corpus A, experiments of altering only the text representation of the baseline configuration were conducted. The results are listed in Table 9.

Table 9: Accuracy of the classifier using a specific text representation.

Unigrams (baseline)	Unigrams & bigrams	Unigrams & bigrams & trigrams
76.2%	79.2%	77.0%

Clearly, adding bigrams to unigrams improved the accuracy with respect to the baseline configuration appreciably. Further adding trigrams did not improve the accuracy with respect to using unigrams and bigrams. Thus, unigrams & bigrams were used as text representation for this thesis' document-to-vector-transformation.

Choosing the C-Parameter

The *C*-parameter can be used for tuning the accuracy of the SVM training algorithm (see Section 2.4.3.4). In Section 3.2.2, suitable configurations of the *C*-parameter were derived from the literature. In an exploratory analysis, the baseline configuration in combination with the unigram & bigram text representation was used to evaluate each of the *C*-parameter configurations on Corpus A. Resulting accuracies are listed in Table 10. In the explored *C*-parameter range no increase or decrease of the resulting accuracy with respect to the default configuration of C=1 was observed. Thus, C=1 was used in the following.

The resulting final classifier's level of accuracy surpasses the accuracy of classifiers of investor sentiment in relevant related work (e.g., <70% in Das & Chen (2007) and 75.1% in O'Hare et al. (2009)). In contrast to many other approaches (see Section 2.6), the classification accuracy was evaluated and made transparent. Using the best parameter configurations in terms of accuracy reported in this section, a classifier was trained on the overall Corpus A (using all training examples of annotated blog documents consisting of title and body with specific company mentions exchanged for a "[comp]"-string). This classifier serves in Section 4 for classifying the (sentiment orientation of) investor sentiment of large datasets of investment blog documents. These investor sentiments are to be validated in a portfolio simulation.

Table 10: Accuracy of the classifier using a specific *C*-parameter.

i	$C=2^i$	Accuracy
–5	0.03125	79.2%
–4	0.0625	79.2%
–3	0.125	79.2%
–2	0.25	79.2%
–1	0.5	79.2%
0	1	79.2%
1	2	79.2%
2	4	79.2%
3	8	79.2%
4	16	79.2%
5	32	79.2%
6	64	79.2%
7	128	79.2%
8	256	79.2%
9	512	79.2%
10	1,024	79.2%
11	2,048	79.2%
12	4,096	79.2%
13	8,192	79.2%
14	16,384	79.2%
15	32,768	79.2%
16	65,536	79.2%
17	131,072	79.2%
18	262,144	79.2%
19	524,288	79.2%
20	1,048,576	79.2%

4 Validation by a Portfolio Simulation

This section reports on the validation of the classifier of the (investor) sentiment orientation in investment blog documents (reported on in the preceding section) regarding its usefulness for investing in specific stock portfolios by means of a portfolio simulation. The portfolio simulation uses historic investor sentiment data to select stocks in a portfolio. Thus, the predictive relationship of investor sentiment to abnormal stock returns can be analyzed on the portfolio level. This section presents the datasets of investor sentiment classified from blog documents, the design of the portfolio simulation, hypotheses regarding the output of the simulation, the methods used for testing the hypotheses, and the results obtained.

4.1 Datasets

To validate investor sentiment in blog documents, a large amount of blog documents referring to the same stocks and sampled from a continuous and long time period is required, with blog documents to be classified according to their (investor) sentiment orientation. This section describes the acquisition of the datasets from two blog platforms and discusses their properties on the monthly aggregation level, which was used for the portfolio simulation described in Section 4.2.

4.1.1 Data Acquisition

The sets of investment blog documents were sourced from two blog platforms (see Section 2.3.2): (1) Seekingalpha, and (2) Blogspot. The aim was to retrieve blog documents from these blog platforms referring each to a specific stock from the U.S. stock market index Dow Jones Industrial Average (DJIA). The DJIA comprises 30 large stocks (S&P Dow Jones Indices LLC, 2014). This kind of stocks usually gets a lot of coverage in traditional media and blogs, benefitting the portfolio simulation of this thesis with a potentially large number of investor sentiment documents per stock. Regarding Seekingalpha, support is provided by Fotak (2007, p.14), who found the blog documents to refer to mostly large firms (with an average market capitalization of U.S.$35 billion) in a 500 blog document sample from 2006. Furthermore, experiments in the scope of this thesis for retrieving blog documents referring to S&P 500 stocks not contained in the DJIA and for S&P 600 small capitalization stocks had returned only few blog documents.

In this thesis, the DJIA constituents of 2009-06-08 are used (see Table 61). These constituents remained part of the DJIA until 2012 (S&P Dow Jones Indices LLC, 2013). The 5-year time period of this thesis' datasets ranges from 2007 until 2011. This means that a few stocks in this thesis' portfolio simulation were not part of the DJIA in 2007, 2008, and 2009.

Blog documents could be retrieved for 29 stocks (described in Section 4.1.2) of the 30 DJIA stocks from both blog platforms.

The retrieval of blog documents from the blog platform websites of Seekingalpha and Blogspot was conducted via a web service of Google Search (see Appendix A.4). For Seekingalpha, searching for blog documents was restricted to the URI http://seekingalpha.com/article to retrieve only full article blog documents, discussing stocks and other financial instruments. Furthermore, the web service allowed specifying search queries for blog documents referring to a specific stock and dating to a specific day the blog document was crawled and indexed by Google. This date is assumed to be a conservative estimate of the publication date of a blog document. The stock was specified in the query by a search term listed in Table 61 in the Appendix. After retrieval, the existence of a reference to a specific DJIA stock in a blog document was verified by requiring at least one mention of several stock-specific labels (see Table 61 in the Appendix).

The text of the blog documents (i.e., the title and the body) was automatically extracted from the retrieved HTML web documents (see Appendix A.4). The text of each blog document was subject to the document-vector-transformation described in Section 3.2.1 and an SVM-based classifier (see Section 3.2.2 for the configuration), which uses the vector representation. The classifier for the sentiment orientation of investor sentiment in the blog documents was created, i.e., trained, using the corpus described in Section 3.1.2 . That is, the combination of title and body of all blog documents of Corpus A (with specific company-mentions exchanged for a "[comp]"-string to not learn company-specific sentiment orientation classifications) and the respective document level annotations of the sentiment orientation were used as training examples. The implementation of the software to train and apply the classifier is described in the Appendix A.3.

Note that a classifier trained only on Seekingalpha blog documents was used also for classifying Blogspot blog documents. To this respect, the vocabulary of *investment* blog documents from Seekingalpha and Blogspot is assumed to be similar to a large degree. Thus, the classification accuracy of the classifier on Blogspot blog documents is assumed to be comparable to the one on Seekingalpha blog documents.

4.1.2 Description and Analysis

The description and analysis of the acquired datasets is with respect to (investor) sentiment orientations classified in blog documents sourced from the blog platforms of (1) Seekingalpha, and (2) Blogspot.

4.1.2.1 Seekingalpha

The Seekingalpha dataset comprises 77,539 document level investor sentiment (document score) referring to 29 large U.S. stocks of the DJIA (as of 2009-06-08). Table 11 lists the number of investor sentiment (document score) for each stock in each calendar year, ranging from 2007 to 2011. Regarding the company names in Table 11 and other tables, E.I. du Pont

de Nemours and Company is abbreviated “Du Pont”, International Business Machines Corporation is abbreviated as “IBM Corporation”, and United Technologies Corporation is abbreviated as “United Technologies Corp.”. The total number of investor sentiment (document score) per year increases in the time period from 2007 until 2011. For 2007, the number of investor sentiment (document score) is relatively low, totaling to 4,496. The increase of the number of investor sentiment (document score) could be due to two potential reasons: (1) the growth of the Seekingalpha platform in terms of authors and blog documents, and (2) a potential recency bias by the web search engine used, which might have returned more blog documents for more recent dates.

Table 11: Number of investor sentiments in the Seekingalpha dataset on the document level.

Stock	2007	2008	2009	2010	2011	Total
3M Company	11	27	54	64	131	287
Alcoa Inc.	77	147	348	361	539	1,472
American Express Company	83	364	580	292	447	1,766
AT&T Inc.	267	827	974	1,063	1,724	4,855
Bank of America Corporation	320	1,451	3,770	2,807	3,129	11,477
The Boeing Company	119	488	511	494	117	1,729
Caterpillar Inc.	12	33	60	79	244	428
Chevron Corporation	69	219	372	442	184	1,286
Cisco Systems Inc.	191	423	725	748	1,490	3,577
The Coca-Cola Company	40	53	74	149	191	507
Du Pont	13	11	7	14	102	147
Exxon Mobil Corporation	125	336	470	565	1,235	2,731
General Electric Company	172	568	838	483	256	2,317
Hewlett-Packard Company	189	507	730	891	1,554	3,871
The Home Depot Inc.	181	260	313	336	112	1,202
Intel Corporation	252	767	1,173	1,322	359	3,873
IBM Corporation	263	839	1,197	1,045	266	3,610
Johnson & Johnson	170	379	658	745	236	2,188
J.P. Morgan Chase & Company	211	1,314	147	2,052	3,334	7,058
Kraft Foods Inc.	41	59	105	127	311	643
McDonald's Corporation	121	241	408	475	1,058	,2303
Merck & Company Inc.	115	206	300	312	154	1,087
Microsoft Corporation	701	2,342	2,414	1,907	596	7,960
Pfizer Inc.	169	489	709	752	204	2,323
The Procter & Gamble Company	72	196	356	427	149	1,200
United Technologies Corp.	48	112	152	146	55	513
Verizon Communications Inc.	49	43	76	150	419	737
Wal-Mart Stores Inc.	373	1,058	1,319	1,223	1,775	5,748
The Walt Disney Company	42	90	115	130	267	644
TOTAL	4,496	13,849	18,955	19,601	20,638	77,539

Considering individual stocks, the number of investor sentiment (document score) per stock in total varies widely and ranges from 147 (Du Pont) to 11,477 (Bank of America). Furthermore, the number of investor sentiment (document score) of some stocks in some years is relatively low (i.e., below 20) – compared to other years for the same stock. A low number of investor sentiment (document score) is observed for Du Pont stock in the time period of 2007 until 2010, and 3M stock and Caterpillar stock in 2007. The above reasons

for data sparseness might to some degree also apply to these stocks. Furthermore, technical reasons during the retrieval of the blog documents with respect to specific stocks might apply.

Distributional Overview

Figure 14 provides a distributional overview of all stocks' monthly investor sentiment indexes by yearly box plots.

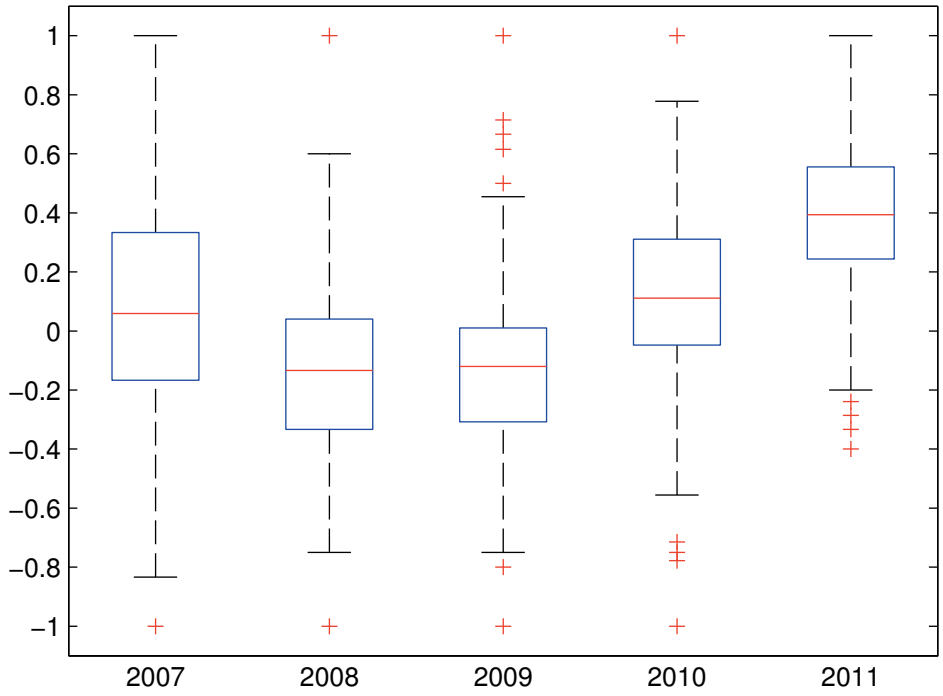


Figure 14: Box plots of Seekingalpha investor sentiment indexes. The investor sentiment indexes refer to all stocks in the dataset. Each box represents 50% of monthly investor sentiment indexes per year over the period 2007 - 2011. The horizontal line, cutting a box in two halves, is the 50% percentile. The upper (lower) line of a box is the 75% (25%) percentile. The "whiskers" represent 99.3% of the investor sentiment indexes in a year (assuming them to be normally distributed). The crosses represent outliers beyond the whiskers.

The median of the 2007 box plot is slightly above zero, falls below zero in the box plot of 2008 and increases in the box plots of subsequent years. That is, the distribution of the monthly investor sentiment indexes changes over time and 2008 is the most negative year in the study. The view on 2008 being the most negative year is also expressed in the time series plot of the monthly investor sentiment index market in Figure 15. The skewness of the yearly distributions (indicated by the location of the median in the box) seems to be low with some skew to the negative side of investor sentiment indexes for 2009 and some positive skew for 2010. The height of the whiskers and the box is largest for 2007, indicating a large range of investor sentiment indexes covered, including many index values towards the 1 and –1 ends of the scale. The range of the majority of investor sentiment indexes is smaller for the next years. The smaller range of investor sentiment indexes correlates with the larger number of investor sentiment (document scores) that make up the investor sentiment indexes. That is,

the larger number of investor sentiment (document scores) correlates with an increase of the central tendency of the investor sentiment index values.

Mean of Investor Sentiment Indexes

The mean of the monthly investor sentiment index (see Definition (2.10)) for the stocks per year is presented in Table 12. The overall mean of the whole dataset is 0.073, which can be interpreted as a slightly positive sentiment orientation. However, instead of this interpretation, the change of the mean investor sentiment index from one year to the next and the difference in the level of the mean investor sentiment index for the same year for different stocks seems more interesting. The yearly mean of monthly investor sentiment indexes of all stocks decreases for the time period from 2007 to 2008 and increases for the time period from 2008 to 2011. The dip with respect to 2008 seems plausible because the prices of stocks in the U.S. stock market substantially decreased in 2008 (see Figure 15).

Table 12: Mean investor sentiment index in the Seekingalpha dataset on a monthly basis.

Stock	2007	2008	2009	2010	2011	Total
3M Company	0.167	0.038	0.239	0.267	0.411	0.224
Alcoa Inc.	0.058	–0.109	–0.343	–0.166	0.238	–0.064
American Express Company	0.165	–0.302	–0.363	–0.167	0.510	–0.031
AT&T Inc.	0.046	–0.069	–0.153	0.096	0.341	0.052
Bank of America Corporation	–0.097	–0.435	–0.507	–0.243	0.035	–0.249
The Boeing Company	0.078	–0.113	–0.304	–0.025	0.086	–0.056
Caterpillar Inc.	–0.042	0.003	0.041	0.212	0.660	0.175
Chevron Corporation	0.238	–0.123	–0.110	0.180	0.504	0.138
Cisco Systems Inc.	0.238	–0.086	–0.086	0.198	0.316	0.116
The Coca-Cola Company	0.267	0.033	0.218	0.543	0.413	0.295
Du Pont	–0.139	0.000	0.000	0.167	0.429	0.091
Exxon Mobil Corporation	0.166	–0.186	–0.068	0.231	0.505	0.130
General Electric Company	0.254	0.001	–0.139	0.131	0.293	0.108
Hewlett-Packard Company	–0.058	–0.044	–0.093	0.199	0.353	0.071
The Home Depot Inc.	–0.303	–0.330	–0.336	–0.104	0.409	–0.133
Intel Corporation	0.153	–0.093	–0.175	0.073	0.348	0.061
IBM Corporation	0.098	–0.057	–0.135	0.136	0.264	0.061
Johnson & Johnson	0.207	–0.034	0.116	0.263	0.579	0.226
J.P. Morgan Chase & Company	–0.243	–0.489	–0.046	–0.279	0.043	–0.203
Kraft Foods Inc.	0.192	–0.140	0.036	0.383	0.673	0.228
McDonald's Corporation	0.163	–0.128	–0.098	0.200	0.457	0.119
Merck & Company Inc.	0.124	–0.199	0.000	0.317	0.416	0.132
Microsoft Corporation	0.150	–0.129	–0.148	0.140	0.370	0.077
Pfizer Inc.	0.267	–0.152	–0.075	0.212	0.377	0.126
The Procter & Gamble Company	0.242	–0.056	–0.074	0.187	0.492	0.158
United Technologies Corp.	–0.080	0.032	–0.006	–0.066	0.522	0.080
Verizon Communications Inc.	0.041	–0.011	–0.095	0.103	0.498	0.107
Wal-Mart Stores Inc.	0.091	–0.252	–0.229	0.027	0.292	–0.014
The Walt Disney Company	–0.114	0.045	–0.237	0.394	0.354	0.088
TOTAL	0.080	–0.117	–0.109	0.124	0.386	0.073

NOTES: the yearly TOTAL value was calculated over the investor sentiment index data of all stocks. The total value per stock was calculated over all investor sentiment index data per stock from 2007-2011. The overall total value was calculated over all investor sentiment index data in the whole dataset, covering all stocks and all time periods.

Considering individual stocks, stocks in the financial sector (i.e., American Express, Bank of America, and J.P. Morgan Chase) were among the stocks with the lowest level of their mean investor sentiment index in 2008. These mean investor sentiment index levels are also quite remarkable in absolute terms: the mean investor sentiment index for J.P. Morgan Chase and Bank of America are below –0.4. This low level seems plausible because the financial sector was deeply involved in the financial crisis of 2008 (e.g., (Allen & Carletti, 2010)). In 2009, beside financial stocks, aluminum producer Alcoa (e.g., (Alcoa, 2010, p.43)) and "home improvement retailer" (Home Depot, 2010, p.1) Home Depot were among the companies with their stocks observing lowest levels of the mean investor sentiment index. The low mean investor sentiment index level of non-financial stocks plausibly seems related to the fundamentals of these companies. The sales of Alcoa decreased from 26.9 billion U.S.$ in 2008 to 18.4 billion U.S.$ in 2009 and in 2009 Alcoa observed a net loss of more than 1 billion U.S.$ (Alcoa, 2010, p.43). The retailer Home Depot suffered from a decrease of net sales by 7.2% and a decrease of the gross profit by 6.6% in the fiscal year 2009, ending 2010-01-31 (Home Depot, 2010, p.19).

In contrast to the stocks observing a negative mean investor sentiment index, the mean investor sentiment index of Coca-Cola Company is positive for each year from 2007 until 2011. Presumably, this is because sales and profit of Coca-Cola Company were affected much less by the financial and economic crisis of 2008 and later (e.g., (Coca-Cola Company, 2010)).

Variability of Investor Sentiment Indexes

The range of levels of the monthly investor sentiment index values per stock per year is explored by the standard deviation and is presented in Table 13. The overall standard deviation of all monthly investor sentiment indexes in the dataset is 0.406. The standard deviation of the investor sentiment index is relatively high for stocks with a relatively small number of investor sentiment document scores (e.g., for stocks of 3M, Du Pont, Caterpillar, Kraft Foods, and Walt Disney). The high standard deviation seems plausible because a small number of investor sentiment document scores, which are aggregated into the investor sentiment index (see Definition (2.10)), correlates with a lower central tendency of the investor sentiment index variable (see above).

Table 13: Standard deviations of monthly investor sentiment index (Seekingalpha).

Stock	2007	2008	2009	2010	2011	Total
3M Company	0.718	0.730	0.466	0.311	0.267	0.531
Alcoa Inc.	0.322	0.394	0.245	0.318	0.289	0.366
American Express Company	0.400	0.307	0.188	0.247	0.109	0.420
AT&T Inc.	0.297	0.148	0.151	0.118	0.075	0.240
Bank of America Corporation	0.344	0.093	0.110	0.095	0.151	0.271
The Boeing Company	0.290	0.274	0.301	0.192	0.325	0.307
Caterpillar Inc.	0.753	0.718	0.795	0.531	0.125	0.665
Chevron Corporation	0.487	0.387	0.237	0.196	0.228	0.394
Cisco Systems Inc.	0.369	0.233	0.141	0.240	0.137	0.288
The Coca-Cola Company	0.535	0.585	0.384	0.294	0.204	0.446
Du Pont	0.731	0.739	0.603	0.937	0.408	0.709
Exxon Mobil Corporation	0.364	0.163	0.206	0.160	0.108	0.322
General Electric Company	0.303	0.261	0.199	0.168	0.212	0.277
Hewlett-Packard Company	0.243	0.148	0.211	0.171	0.152	0.254
The Home Depot Inc.	0.317	0.175	0.253	0.267	0.229	0.376
Intel Corporation	0.198	0.163	0.135	0.088	0.178	0.240
IBM Corporation	0.286	0.184	0.161	0.124	0.273	0.252
Johnson & Johnson	0.253	0.265	0.156	0.140	0.214	0.289
J.P. Morgan Chase & Company	0.322	0.151	0.159	0.120	0.128	0.264
Kraft Foods Inc.	0.635	0.448	0.441	0.395	0.155	0.513
McDonald's Corporation	0.332	0.222	0.188	0.219	0.105	0.307
Merck & Company Inc.	0.476	0.178	0.271	0.246	0.370	0.385
Microsoft Corporation	0.149	0.130	0.111	0.134	0.167	0.238
Pfizer Inc.	0.332	0.162	0.224	0.110	0.241	0.300
The Procter & Gamble Company	0.417	0.293	0.255	0.213	0.237	0.352
United Technologies Corp.	0.708	0.309	0.306	0.351	0.393	0.481
Verizon Communications Inc.	0.499	0.777	0.250	0.381	0.154	0.495
Wal-Mart Stores Inc.	0.198	0.149	0.123	0.109	0.079	0.245
The Walt Disney Company	0.680	0.381	0.447	0.419	0.169	0.502
TOTAL	0.455	0.384	0.337	0.347	0.262	0.406

NOTES: see notes below Table 12.

Relating Investor Sentiment to Returns

The relationship between investor sentiment (document scores) in the Seekingalpha dataset to stock returns is studied on the "market" level in the following. This analysis is to get a first impression of the relationship and does not yet deal with effects of investor sentiment on abnormal returns. Figure 15 plots the contemporaneous monthly time series of (1) the investor sentiment index market (see Definition (2.11)), aggregating the investor sentiment document scores with respect to all 29 stocks represented in the dataset per month, and (2) cumulative log returns of the DJIA stock market index (of which all of the 29 stocks were constituents starting from 2009-06-08 and most of them also before). The visual impression clearly indicates a considerable amount of contemporaneous correlation between the two variables.

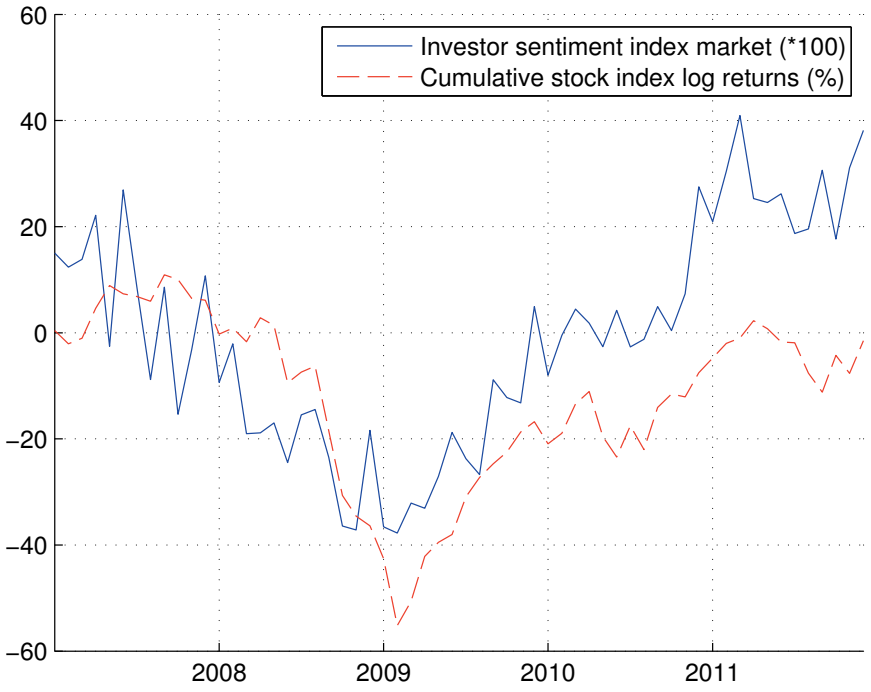


Figure 15: Seekingalpha investor sentiment index market vs. cumulative DJIA returns based on monthly frequency data over the period 2007 – 2011.

To further investigate the relationship, a scatter plot of monthly investor sentiment index market values and DJIA log returns is provided in Figure 16.

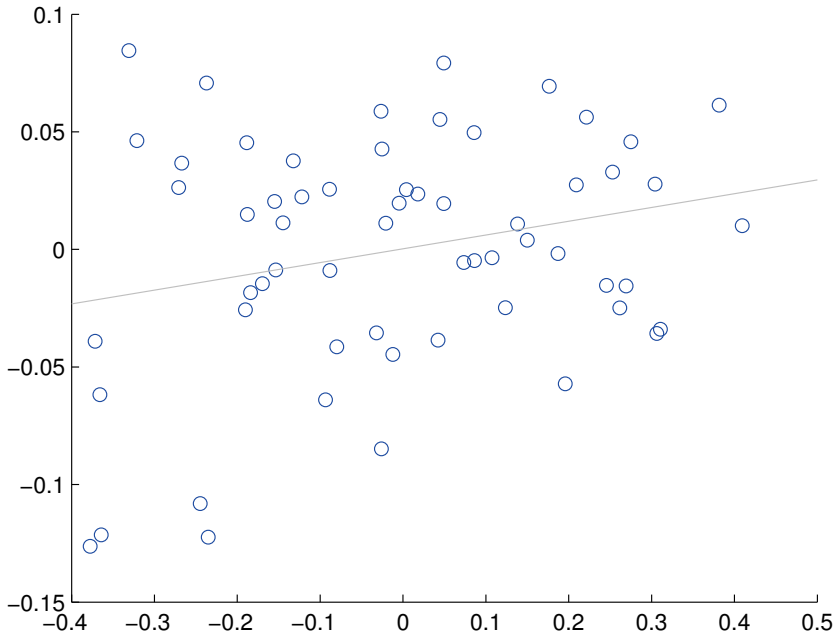


Figure 16: Scatter plot of DJIA log returns and Seekingalpha investor sentiment index market *sim* using monthly data. Returns are on the vertical axis and *sim* is on the horizontal axis. The line is the ordinary least squares regression line $r(t)=0.00027 + 0.05855 \cdot sim_t$ with $R^2=0.061$ (the fraction of variations in *r* that is explained by the regression (see Definition (4.5))). The *t*-statistic of the *sim*-coefficient is 1.9392.

The linear regression line in Figure 16 indicates a positive linear relationship, i.e., an increase in the monthly investor sentiment index market correlates to some degree with an increase in DJIA log returns. The slope is statistically significantly different from zero at the 10% level. However, the pairs are scattered wildly around the regression line and the R^2 is rather small. At least some evidence exists for a relationship on this level. Note that this does not tell much about the effect of stock level investor sentiment document on abnormal returns. This effect is subject to investigation in the next sections.

4.1.2.2 Blogspot

The Blogspot dataset comprises 198,844 investor sentiment (document scores). That is, the Blogspot dataset is substantially larger than the Seekingalpha dataset. Because Blogspot is (in contrast to Seekingalpha) a large blog platform with unrestricted access to create blogs and content (i.e., general topic), this seems plausible.

Table 14: Number of investor sentiments in the Blogspot dataset on the document level.

Stock	2007	2008	2009	2010	2011	Total
3M Company	34	61	85	133	241	554
Alcoa Inc.	2	1	3	0	1,177	1,183
American Express Company	673	1,379	1,870	2,087	3,794	9,803
AT&T Inc.	92	90	78	87	3,640	3,987
Bank of America Corporation	977	1,972	3,597	3,824	6,027	16,397
The Boeing Company	21	24	21	28	5,453	5,547
Caterpillar Inc.	29	34	136	212	407	818
Chevron Corporation	1	2	3	4	753	763
Cisco Systems Inc.	226	320	549	967	2,151	4,213
The Coca-Cola Company	16	15	20	23	1,344	1,418
Du Pont	34	39	43	72	168	356
Exxon Mobil Corporation	541	748	859	1,354	3,325	6,827
General Electric Company	15	11	12	13	3,869	3,920
Hewlett-Packard Company	585	894	1,396	1,868	3,283	8,026
The Home Depot Inc.	24	28	40	38	6,575	6,705
Intel Corporation	14	8	6	9	6,322	6,359
IBM Corporation	26	19	14	18	5,744	5,821
Johnson & Johnson	2	1	2	5	2,442	2,452
J.P. Morgan Chase & Company	341	1,151	1,626	2,147	4,636	9,901
Kraft Foods Inc.	117	217	533	703	1,503	3,073
McDonald's Corporation	2,649	3,908	5,022	6,513	6,943	25,035
Merck & Company Inc.	2	6	4	0	862	874
Microsoft Corporation	121	94	71	69	7,433	7,788
Pfizer Inc.	6	2	4	3	2,881	2,896
The Procter & Gamble Company	11	17	40	22	4,000	4,090
United Technologies Corp.	0	0	0	0	550	550
Verizon Communications Inc.	54	51	109	183	688	1,085
Wal-Mart Stores Inc.	4,186	5,714	6,982	6,957	7,533	31,372
The Walt Disney Company	2,628	4,446	6,047	6,990	6,920	27,031
TOTAL	13,427	21,252	29,172	34,329	100,664	198,844

Like in the Seekingalpha dataset, the total number of investor sentiment document scores per year increases for the time period ranging from 2007 until 2011, with more than half of the investor sentiment document scores referring to 2011 (see Table 14). Furthermore, the

total number of investor sentiment document scores varies across the stocks, ranging from 356 (for Du Pont) to 31,372 (for Wal-Mart Stores). Also, for many stocks the yearly number of investor sentiment document scores is much higher for 2011 compared to the years before, applying to Alcoa, AT&T, Boeing, Chevron, Coca-Cola, General Electric, Home Depot, Intel, IBM, Johnson & Johnson, Merck, Microsoft, Pfizer, Procter & Gamble, and United Technologies. Possible reasons for data sparseness before 2011 are: (1) less published blog documents when going back in time, (2) a bias of the web search engine that returns more recent blog documents, and (3) technical problems in the retrieval process. For a few stocks there is even no investor sentiment (document score) at all in some years in the Blogspot dataset. These gaps are likely to be due to technical reasons in the retrieval process.

Distributional Overview

The distributional overview of the monthly investor sentiment indexes in the Blogspot dataset is provided by the yearly box plots for 2007 – 2011 in Figure 17. For each year observed, the median and the boxes are on the negative side, indicating that (at least) 75% of the monthly investor sentiment indexes have a negative sentiment orientation. However, in years 2007 – 2010, the boxes show a skew towards the positive side – while still being negative. For years 2007 – 2010, about 25% of the monthly investor sentiment indexes have a positive sentiment orientation. It seems questionable whether the mostly negative monthly investor sentiment indexes relate to the fundamentals or future abnormal returns of the stocks. But still, this work investigates the possibility that the relative ranks among the levels of monthly investor sentiment indexes relate to future abnormal returns.

Comparing box plots of years 2007, 2008, and 2009, the box plots (i.e., the height and location of the box, the location of the median, and the height of the whiskers) look similar, indicating homogeneous distributions that do not vary over these periods. This observation is in contrast to the Seekingalpha dataset. Also, the question arises whether the overall sentiment (with respect to the 29 stocks covered) is represented well in the time period of 2007 – 2009. For 2010, a substantial increase of the median is observed with respect to 2009. This increase is similar to the Seekingalpha dataset. It also seems reasonable considering the development of the stock market (see Figure 18). The box and the whiskers in the 2011 box plot are much narrower compared to the other box plots in the years before. This observation is similar to the Seekingalpha dataset and might be due to the much larger number of investor sentiment document scores available in 2011.

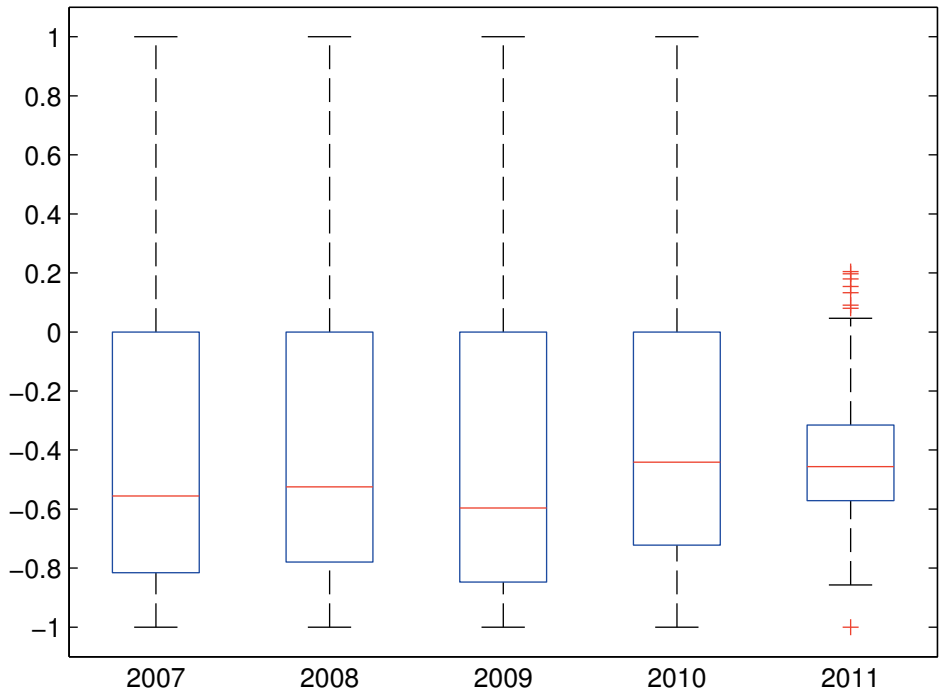


Figure 17: Box plots of Blogspot investor sentiment indexes. The investor sentiment indexes refer to all stocks in the dataset. Each box represents 50% of monthly investor sentiment indexes per year over the period 2007 - 2011. The horizontal line, cutting a box in two halves, is the 50% percentile. The upper (lower) line of a box is the 75% (25%) percentile. The "whiskers" represent 99.3% of the investor sentiment indexes in a year (assuming them to be normally distributed). The crosses represent outliers beyond the whiskers.

Mean of Investor Sentiment Indexes

The mean of the monthly investor sentiment index for the stocks per year is presented in Table 15. The mean for the whole dataset is –0.445, implying an overall negative investor sentiment – unlike in the Seekingalpha dataset. The low mean of the monthly investor sentiment index could be interpreted as authors of blog documents in Blogspot having a more negative bias than Seekingalpha blog document authors. Regarding the yearly overall mean investor sentiment index in the Blogspot dataset, the mean investor sentiment index values of 2007, 2008, and 2011 are close to the mean of the whole dataset. For 2009, the mean investor sentiment index is slightly more negative and for 2010 it is slightly more positive. Only the more positive mean investor sentiment index for 2010 seems to be related to the market price development (see Figure 18). Thus, the change from year to year in the mean investor sentiment index in the Blogspot dataset seems rather implausible.

Regarding individual stocks, like in the Seekingalpha dataset, financial sector stocks (e.g., Bank of America and J.P. Morgan Chase) had a very negative mean investor sentiment index in 2008. Also, retailer Home Depot had a very negative mean investor sentiment index in 2009. In the previous section on Seekingalpha, arguments why this is plausible were discussed. However, other stocks' levels of mean investor sentiment index diverge from the Seekingalpha ones. For instance, Alcoa had a more positive mean investor sentiment index in 2009. Furthermore, Coca-Cola Company and Procter & Gamble Company were associated

with a much more negative mean investor sentiment index from 2007 until 2011. It is questionable, whether these levels plausibly reflect the investor sentiment of Blogspot blog authors because the dataset contains only a low number of investor sentiment document scores for the time period of 2007 until 2010 for all three stocks. However, also for some stocks for which the dataset contains more investor sentiment document scores, such as McDonald's and AT&T, the mean investor sentiment indexes are (very) negative during all five years observed. The high amount of negativity seems to be rather unrelated to the fundamentals of these stocks and possibly may relate to some negative consumer portraits of these companies or of some of their products or services.

Table 15: Mean investor sentiment index in the Blogspot dataset on a monthly basis.

Stock	2007	2008	2009	2010	2011	Total
3M Company	–0.550	–0.617	–0.250	–0.217	–0.280	–0.383
Alcoa Inc.	0.000	0.083	–0.083	0.000	–0.378	–0.076
American Express Company	–0.593	–0.676	–0.700	–0.518	–0.418	–0.581
AT&T Inc.	–0.513	–0.640	–0.857	–0.682	–0.532	–0.645
Bank of America Corporation	–0.675	–0.714	–0.681	–0.488	–0.638	–0.639
The Boeing Company	–0.625	–0.361	–0.717	–0.486	–0.502	–0.538
Caterpillar Inc.	–0.319	–0.172	–0.351	–0.147	–0.201	–0.238
Chevron Corporation	–0.083	0.000	–0.250	–0.250	–0.201	–0.157
Cisco Systems Inc.	–0.350	–0.396	–0.453	–0.144	–0.077	–0.284
The Coca-Cola Company	–0.583	–0.444	–0.653	–0.544	–0.341	–0.513
Du Pont	–0.578	–0.022	–0.172	–0.566	–0.489	–0.365
Exxon Mobil Corporation	–0.489	–0.535	–0.441	–0.340	–0.327	–0.427
General Electric Company	–0.528	–0.444	–0.667	–0.417	–0.434	–0.498
Hewlett-Packard Company	–0.393	–0.449	–0.416	–0.285	–0.261	–0.361
The Home Depot Inc.	–0.726	–0.696	–0.903	–0.530	–0.742	–0.719
Intel Corporation	–0.417	–0.333	–0.333	–0.417	–0.313	–0.363
IBM Corporation	–0.556	–0.569	–0.458	–0.556	–0.485	–0.525
Johnson & Johnson	–0.083	–0.083	0.000	–0.250	–0.411	–0.166
J.P. Morgan Chase & Company	–0.483	–0.758	–0.709	–0.629	–0.624	–0.641
Kraft Foods Inc.	–0.528	–0.580	–0.552	–0.350	–0.404	–0.483
McDonald's Corporation	–0.822	–0.813	–0.839	–0.771	–0.770	–0.803
Merck & Company Inc.	–0.083	–0.333	–0.250	0.000	–0.336	–0.200
Microsoft Corporation	–0.746	–0.618	–0.838	–0.605	–0.535	–0.668
Pfizer Inc.	–0.333	–0.167	–0.250	–0.250	–0.453	–0.291
The Procter & Gamble Company	–0.250	–0.778	–0.708	–0.578	–0.401	–0.543
United Technologies Corp.	0.000	0.000	0.000	0.000	–0.350	–0.070
Verizon Communications Inc.	–0.229	–0.394	–0.306	–0.406	–0.350	–0.337
Wal-Mart Stores Inc.	–0.762	–0.717	–0.783	–0.699	–0.730	–0.738
The Walt Disney Company	–0.689	–0.658	–0.708	–0.656	–0.630	–0.668
TOTAL	–0.448	–0.444	–0.494	–0.406	–0.435	–0.445

NOTES: see notes below Table 12.

Variability of Investor Sentiment Indexes

Regarding the variability of the monthly investor sentiment index values of the stocks in this thesis' Blogspot dataset, Table 16 presents the yearly and total standard deviations.

Table 16: Standard deviations of monthly investor sentiment index (Blogspot).

Stock	2007	2008	2009	2010	2011	Total
3M Company	0.433	0.375	0.388	0.519	0.217	0.420
Alcoa Inc.	0.426	0.289	0.515	0.000	0.331	0.381
American Express Company	0.099	0.068	0.084	0.047	0.060	0.127
AT&T Inc.	0.504	0.222	0.207	0.315	0.061	0.315
Bank of America Corporation	0.103	0.049	0.081	0.067	0.080	0.110
The Boeing Company	0.569	0.502	0.447	0.500	0.067	0.455
Caterpillar Inc.	0.760	0.629	0.381	0.323	0.224	0.494
Chevron Corporation	0.289	0.426	0.452	0.452	0.163	0.375
Cisco Systems Inc.	0.180	0.213	0.173	0.303	0.116	0.249
The Coca-Cola Company	0.669	0.499	0.452	0.774	0.134	0.543
Du Pont	0.468	0.657	0.791	0.344	0.273	0.571
Exxon Mobil Corporation	0.146	0.148	0.117	0.122	0.098	0.148
General Electric Company	0.502	0.499	0.492	0.534	0.088	0.449
Hewlett-Packard Company	0.144	0.112	0.102	0.084	0.131	0.135
The Home Depot Inc.	0.445	0.438	0.230	0.514	0.076	0.384
Intel Corporation	0.793	0.492	0.492	0.669	0.074	0.542
IBM Corporation	0.604	0.474	0.498	0.641	0.055	0.485
Johnson & Johnson	0.289	0.289	0.426	0.622	0.122	0.402
J.P. Morgan Chase & Company	0.255	0.099	0.103	0.093	0.085	0.167
Kraft Foods Inc.	0.268	0.199	0.090	0.159	0.084	0.191
McDonald's Corporation	0.036	0.023	0.035	0.025	0.039	0.042
Merck & Company Inc.	0.289	0.651	0.452	0.000	0.167	0.396
Microsoft Corporation	0.165	0.310	0.217	0.271	0.048	0.241
Pfizer Inc.	0.492	0.389	0.452	0.452	0.075	0.400
The Procter & Gamble Company	0.754	0.410	0.450	0.637	0.080	0.539
United Technologies Corp.	0.000	0.000	0.000	0.000	0.171	0.159
Verizon Communications Inc.	0.449	0.397	0.337	0.283	0.235	0.343
Wal-Mart Stores Inc.	0.048	0.046	0.034	0.035	0.050	0.052
The Walt Disney Company	0.043	0.050	0.044	0.042	0.046	0.051
TOTAL	0.467	0.438	0.432	0.430	0.213	0.407

NOTES: see notes below Table 12.

The overall standard deviation of the whole dataset is 0.407. Thus, the standard deviation is almost the same like for the Seekingalpha dataset. Concerning individual stocks, stocks with a relatively high standard deviation greater than 0.5 over the full 2007 – 2011 time period (i.e., Coca Cola, Du Pont, Intel, and Procter & Gamble Company) come with a much lower standard deviation per year for 2007, 2008, 2009, and 2010. This effect correlates with a much lower number of investor sentiment document scores in the respective years of 2007, 2008, 2009, and 2010 compared to 2011 (see Table 14). Thus, the high standard deviation for some stocks could be explained by a small number of possibly diverging investor sentiment document scores that results in a low central tendency of the respective investor sentiment index variable.

Relating Investor Sentiment to Returns

Graphically relating investor sentiment from the Blogspot dataset to stock returns on the market level, Figure 18 displays the contemporaneous monthly frequency investor sentiment index market time series (see Definition (2.11)) and the cumulative DJIA log returns. The DJIA is assumed to represent the U.S. stock market and covers 30 stocks of which most are

contained in the investor sentiment index market aggregate of stock-specific investor sentiment document scores. Figure 18 reveals a rather flat (i.e., invariant) investor sentiment index market time series throughout 2007 and 2008. In contrast to the Seekingalpha investor sentiment index market time series (see Figure 15), the Blogspot one is completely on the negative side. However, the relative change of the investor sentiment index market, starting in 2009, seems reasonable because it correlates to some degree with the relative changes in the cumulative DJIA log return time series.

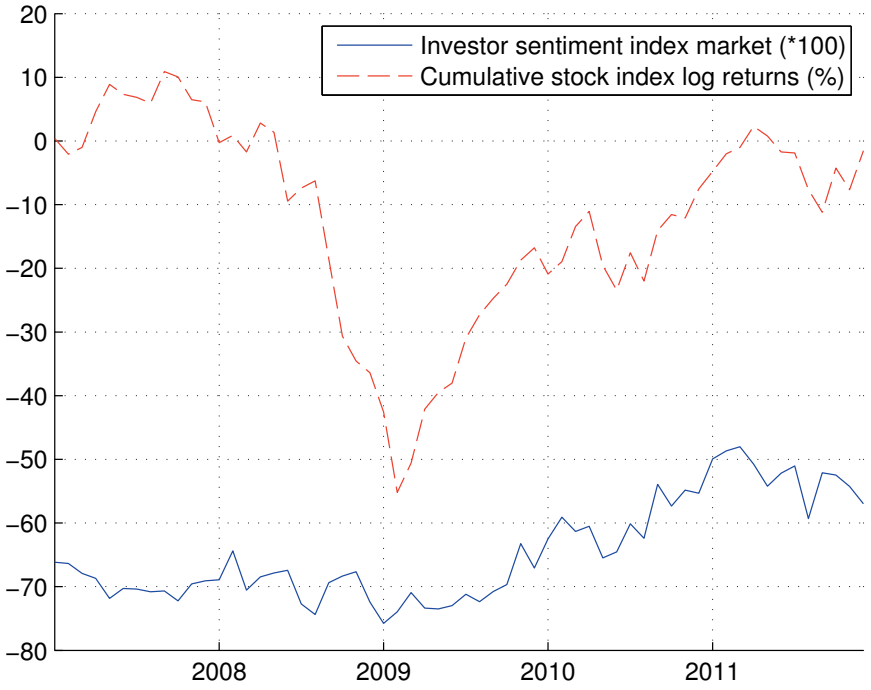


Figure 18: Blogspot investor sentiment index market vs. cumulative DJIA returns based on monthly frequency data over the period 2007 – 2011.

To further investigate the relationship, a scatter plot of monthly investor sentiment index market values from the Blogspot dataset and DJIA log returns is provided in Figure 19. The regression line indicates a positive contemporaneous relationship between the two variables like in the Seekingalpha dataset. However, the relationship is weaker because the coefficient of the investor sentiment index market is statistically not significantly different from zero and the R^2 is even lower. Note again that these findings do not tell much about the effect of investor sentiment document from the Blogspot dataset on future abnormal returns. This effect is investigated in the next sections.

A possible reason for the weak contemporaneous relationship between the investor sentiment index market and DJIA log returns revealed by the scatter plot might be the inclusion of 2007 and 2008 data, in which the investor sentiment index market was rather invariant. The invariance might be due to data sparseness in some stocks. Furthermore, the weak relationship might be genuinely due to authors' investor sentiments or due to errors in (1) the classification of the sentiment orientation, and (2) the identification of investor sentiment document. Both issues seem especially likely for the Blogspot dataset. First, the

classifier is not specific for (the vocabulary in) Blogspot blog documents, which might result in a lower accuracy compared to evaluation results on the Seekingalpha corpus. Second, the Blogspot dataset might contain many blog documents that actually do not have an investment focus. Consequently, the quality of the aggregated investor sentiment index of the Blogspot dataset, which was used in the portfolio simulation, can vary from year to year in the sampling time period and it could be lower compared to the Seekingalpha dataset.

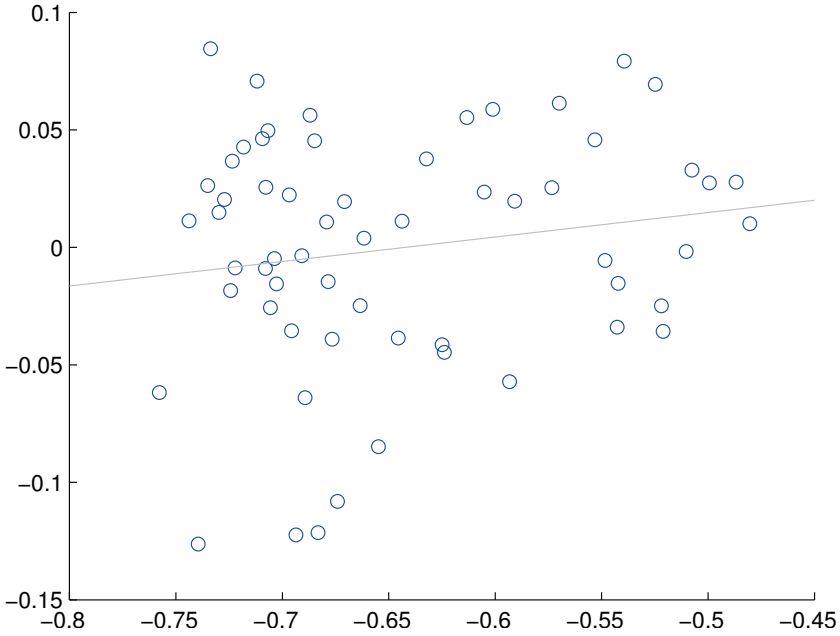


Figure 19: Scatter plot of DJIA log returns and Blogspot investor sentiment index market *sim* using monthly data. Returns are on the vertical axis and sim is on the horizontal axis. The line is the ordinary least squares regression line $r(t)=0.0669 + 0.1042 \cdot sim_t$ with $R^2=0.027$ (the fraction of variations in *r* that is explained by the regression (see Definition (4.5))). The *t*-statistic of the *sim*-coefficient is 1.2683.

With respect to the portfolio simulation, no data transformations were attempted to counter potential problems due to partial data sparseness or varying data quality. However, investor sentiment document scores were aggregated into monthly investor sentiment indexes. Despite observed partial data sparseness, the raw form of the Blogspot dataset can still be valuable and it was used to investigate this thesis' research question. If support for these hypotheses can be found on this basis, this is considered to be conservative because in future work better quality datasets might be obtained (i.e., retrieved), especially with respect to Blogspot and other general-topic blog platforms.

4.2 Design and Hypotheses

This section reports on the design of the portfolio simulation, which uses the datasets of investor sentiment document scores, described in the preceding section, to validate the automatically classified sentiment orientations in these datasets regarding the usefulness for investors. Concerning this validation several hypotheses based on behavioral finance theory and methods for testing the hypotheses are presented.

4.2.1 Portfolio Simulation Design

The portfolio simulation was designed in similar ways to the validations in related investor sentiment literature (e.g., (Chen et al., 2014; Tetlock et al., 2008; Zhang & Skiena, 2010)). Similar to mutual fund performance evaluation (e.g., (Carhart, 1997; Kothari & Warner, 2001)), monthly data frequency was used. That is, this study is not interested in short term (i.e., daily) changes of investor sentiment indexes. Daily data frequency would be rather noisy with respect to this study's datasets because it would be grounded on too few investor sentiment document scores per period. The monthly data frequency of this study is in contrast to many related studies (see Section 2.5) that typically use daily frequency investor sentiment data to make investment decisions. However, the monthly frequency should be more related to real-life investing of either funds or private investors. In total, this study is based on a 60 month time period. Table 17 provides an overview of this study's portfolio simulation, the parameter settings and robustness checks of the parameters. Details are discussed subsequently.

A stock selection strategy serves as the core of the portfolio simulation design. Stock selection is a form of active portfolio management and means to select a number of stocks in a portfolio (Grinold & Kahn, 2000). The design of this study's stock selection strategy is based on the cross-sectional momentum stock selection strategy (Jegadeesh & Titman, 1993). The underlying cross-sectional momentum effect refers to the effect that a portfolio of stocks selected based on highest (lowest) past returns continues to yield higher (lower) returns in the next period on average (Jegadeesh & Titman, 1993). Based on the cross-sectional momentum effect, the cross-sectional momentum strategy is a long-short strategy that selects the highest ranked stocks in a long portfolio and the lowest ranked stocks in a short portfolio (Jegadeesh & Titman, 1993).

In this thesis' stock selection strategy, monthly investor sentiment indexes with respect to the 29 DJIA stocks in this study's datasets (see Section 4.1) were used as ranking and selection criterion instead of using past period returns in the cross-sectional momentum strategy. In each (monthly) period, all stocks were ranked by their monthly investor sentiment index level. The j stocks ranked highest were selected in a long portfolio and the j stocks ranked lowest were selected in a short portfolio. This study defines $j \in \{1,2,\ldots,14\}$ as a parameter of the portfolio simulation with j=5 being the (arbitrary) baseline setting. For simplicity, each stock was weighted equally and all positions were held for one period. In each period, all positions were rebalanced to equal weights. The design of holding the same number of stocks in each monthly period and not having overlapping holding periods should have prevented some problems of biased results and statistical inference problems, documented in respective studies with long-term ($\geq$ 1 year) and clustered stock holdings (e.g., (Barber & Lyon, 1997; Fama, 1998; Lyon et al., 1999)).

In this study's portfolio simulation, positions were opened on the close of the last bank working day in the U.S. of a month. The positions were closed on the close of the last bank

working day in the U.S. of the next month. The aggregation of the investor sentiment indexes of a stock was designed such that all investor sentiment document scores referring to that stock were aggregated with assumed publication dates (i.e., the date Google indexed the respective blog document, see Appendix A.4) being in the range of one day before a position was opened, ranging back to the previous month's last bank working day. In case there was no investor sentiment document score available in a period for a certain stock to be aggregated into an investor sentiment index, such a stock was not considered to be selected in a portfolio. An exception might have occurred when the simulation design requires selecting a number of stocks per period that exceeded the number of stocks in a certain period for which investor sentiment document scores were available. Due to data non-availability, this situation might have occurred for the Blogspot dataset (see Section 4.1.2.2).

To study the informational value of investor sentiment from blogs and to test respective hypotheses (defined in the next subsection) no transaction costs were assumed in this study's portfolio simulations to start with. However, for reasons of practical relevance also the effect of transaction costs on the portfolio simulation results was checked. To this respect, 50 bps roundtrip transaction costs were assumed of which 25 bps became due on opening a position and 25 bps became due on closing a position. 25 bps can be regarded realistic and reasonable relative transaction costs based on a survey that found 26 bps median transaction costs among brokers (Berkowitz et al., 1988, p.104) and 31 bps (26 bps) average trading costs for buying (selling) large cap stocks in the 1990s (Keim & Madhavan, 1998). Because recent transaction costs might be lower for large institutions, 25 bps could be regarded a conservative estimate. However, private investor's actual transaction costs might be considerably higher (Barber & Odean, 2000). Because *all* positions were (re-)balanced to equal weights in *each* period in this study's portfolio simulation, two transactions were assumed for each position in basically each period: closing the position and (re-)opening (with new weights or another stock). Because the number of positions in each period was constant, 50 bps portfolio level transaction costs were assumed for simplicity in each period of the investor sentiment-based portfolio simulation.

The above design is a long-short stock selection strategy that holds the same number of equally-weighted long vs. short positions. Thus, it aims to profit from upward price developments *and* downward price developments in the respectively selected stocks. Because the cash gained from selling short stocks can be used to buy the long position stocks, the strategy is principally self-financing when ignoring transaction costs (Michaud, 1993, p.44). In principle, the strategy can be designed to be independent of market movements with the risk free interest rate serving as the return-benchmark (Michaud, 1993). However, in practice most portfolios of long-short strategies entail (some) market risk (Michaud, 1993, p.47). Thus, other benchmarks are discussed below. Beside the portfolio of all long and short positions, this work separately also studies the portfolios of (1) only the long positions, and (2) only the short positions.

This study's portfolio simulation is based on the monthly calendar-time portfolio method (Jaffe, 1974) similar to (Loughran & Ritter, 1995, pp.44–45; Lyon et al., 1999, p.193; Mitchell & Stafford, 2000, pp.308–309). This method is recommended for detecting (monthly mean) abnormal returns in terms of alpha (Lyon et al., 1999; Mitchell & Stafford, 2000). It "[…] is robust to the most serious statistical problems." (Mitchell & Stafford, 2000, p.291). Foremost, cross-sectional correlation of abnormal returns of the stocks in a portfolio is accounted for (Lyon et al., 1999, p.198; Mitchell & Stafford, 2000, p.288). Second, the problem of a bad model of theoretically expected returns is rather small (Lyon et al., 1999, p.198; Mitchell & Stafford, 2000, p.288). In case of a bad model, the estimated level of alpha could be partly due to mispricing and partly due to misspecification of the model (Mitchell & Stafford, 2000, pp.292,309,324-325). Indications that a portion of such misspecification might exist in the Fama & French (1992, 1993) model (see Section 2.1.2.2) is given in ((Fama & French, 1993) cited in (Mitchell & Stafford, 2000, pp.292,309,325)). However, the bad model problem primarily affects small stocks (Fama, 1998), which were not used in this study. Still, to improve on the model of Fama & French (1992, 1993), the model of Carhart (1997) was used, which adds another explanatory variable (see Section 2.1.2.3).

The power of the calendar-time portfolio method to detect abnormal returns was found to be "sufficient" (Mitchell & Stafford, 2000, pp.321–323). That is, not detecting abnormal returns does not necessarily mean they do not exist (see Section 4.2.3 for an elaboration).

A drawback of the calendar-time portfolio method is that the resulting abnormal return measurement does not resemble investor experience (Lyon et al., 1999, p.166). Therefore, a simple but intuitive measure of abnormal return similarly to (Barber & Odean, 2000) was also used by measuring the total return of the simulated portfolio over the simulation period in excess of a benchmark portfolio's return over the same period. The major advantage of this measure is that it does not require a theoretical asset pricing model. Furthermore, a risk-adjusted comparison of the benchmark portfolio and the simulated portfolio was conducted by the respective Sharpe ratios (Sharpe, 1966) of both portfolios. The major advantage of the Sharpe ratio is that it is the best known investment performance measure ((Modigliani & Modigliani, 1997) cited in (Eling & Schuhmacher, 2007, p.2645)). The Sharpe ratio is appropriate in case portfolio returns are distributed normally (e.g., (Eling & Schuhmacher, 2007)). However, Sharpe ratio-based relative rankings of hedge funds that observed non-normally distributed returns were shown to be identical to rankings obtained from 12 alternative performance measures in most cases (Eling & Schuhmacher, 2007). Thus, the Sharpe ratio was found to be an adequate performance measure in the hedge fund context (Eling & Schuhmacher, 2007). The long-short stock selection strategy used in the study of this thesis is closely related to hedge fund strategies (Fung & Hsieh, 2013). Furthermore, the return distributions of the portfolios to be compared should have similar characteristics because the stocks to be selected were sourced all from the same universe of 29 DJIA stocks and the frequency of selection was identical. Thus, reporting the Sharpe ratio for informal portfolio performance comparisons appears to be valid (– complementary formal hypothesis

testing approaches based on estimations of abnormal returns are presented in Section 4.2.3). A higher Sharpe ratio of the simulated portfolio compared to a benchmark portfolio would indicate some added value by the information of investor sentiment in blog documents.

The first benchmark portfolio was given by the "normal portfolio" (Feibel, 2003, p.115) generated by the *passive* version of the underlying investment strategy in this study's portfolio simulation. That is, all 29 DJIA stocks in the portfolio simulation's stock universe were contained equal-weighted in the normal portfolio with long positions that remained unchanged over the whole simulation time period. Thus, the normal portfolio represents a passive investor's portfolio who does not rebalance and who does not possess the information of investor sentiment from blogs. Thus, the normal portfolio is also termed the buy-and-hold portfolio in this study. Consequently, there should be no biases due to newly listed stocks or rebalancing in the normal portfolio (Lyon et al., 1999). The normal portfolio was used as benchmark portfolio for the long-short stock selection strategy because this strategy practically entails market risk, invalidating the risk-free rate as benchmark (see above). In contrast to the simulation of the long-short stock selection strategy, the buy-and-hold portfolio simulation generates buy transactions only on opening all positions at the beginning of the simulation and sell-transactions on closing all the positions at the end of the simulation period. That is, 25 bps transaction costs were assumed in the first period and 25 bps transaction costs in the last period on the portfolio level for the buy-and-hold portfolio.

The excess return with respect to the normal portfolio can be regarded as either a kind of an "own benchmark abnormal return" (Barber & Odean, 2000, p.783) or a market-adjusted abnormal return (also used by (Barber & Odean, 2000)). The abnormal return is relative to the return of the normal portfolio, consisting of 29 DJIA stocks that serve as a proxy for the U.S. stock market. Thus, this study regards the mean (median) monthly return of the simulated portfolio in excess of the buy-and-hold/normal portfolio a measure of mean (median) monthly abnormal return.

The second benchmark portfolio was generated by the cross-sectional price momentum strategy (described above), on which this study's stock selection strategy using investor sentiment from blog documents is based on. Thus, it is the natural *active* strategy for generating a benchmark portfolio. To make the estimates of alpha and the portfolio returns of the strategies comparable, the cross-sectional price momentum strategy was configured in the same way like this study's stock selection strategy except for informing it by the stocks' monthly log returns. That is, the stocks were ranked by the return in the last month and held for one month in the subsequent month.

Table 17: Portfolio simulation design overview.

Simulation parameter	Setting	Robustness check
Fixed parameters		
Type of active investing	Stock selection	–
Strategy	Long-short strategy	Long only, short only
Stock universe	29 large DJIA stocks	–
Position weighting	Equal-weighting with rebalancing in each period	–
Period length	One month	–
Number of periods to form investor sentiment indexes	One period	–
Number of holding periods	One period	–
Type of investor sentiment	Monthly investor sentiment index level	–
Number of simulation periods	60	–
Evaluation metric	Portfolio level estimate of alpha of Carhart's (1997) model	Excess return vs. the buy-and-hold portfolio return
Transaction costs	None	50 bps per period
Variable parameters		
Number of stocks to select in long and short (sub-)portfolios	Baseline: 5	{1,2,3,4,6,7,8,9,10, 11,12,13,14}

4.2.2 Hypotheses

For answering the research question of this thesis (see Section 1.2), testable hypotheses regarding long term effects on stock portfolios on the monthly scale are provided. Investor sentiment as the independent variable was measured by the investor sentiment index (see Definition (2.10), while its effects on the portfolio level were measured in terms of (1) an estimate of the alpha in Carhart's (1997) model (see Definition (2.5)), and (2) excess returns versus the buy-and-hold benchmark portfolio's returns. The estimated alpha of the simulated portfolio measures mean monthly abnormal returns. Alternatively, the mean or median monthly excess return (vs. the buy-and-hold portfolio's returns) is another measure of abnormal returns (see the preceding subsection).

The set of hypotheses to be provided is based on predictions of behavioral finance theory and especially noise trader theory (see Section 2.2.2). These theories account for fundamental psychological biases that affect investors' decision making (see Section 2.2.1, and, e.g., (Hirshleifer, 2001)). Theoretical models of investor sentiment relate these biases to mispricing effects (e.g., (Barberis et al., 1998; Daniel et al., 1998)). That is, investor sentiment can be defined to relate to abnormal returns (see Section 2.2.2). According to noise trader theory, trading on investor sentiment is correlated, cannot be easily arbitraged away, and can exist over longer time periods (see Section 2.2.2, and, e.g., (Shleifer & Summers, 1990; Shleifer & Vishny, 1997)). This long term effect has also been evidenced by means of a noise trader model (De Long et al., 1991). Therefore, this thesis focuses on studying effects of investor sentiment on abnormal returns on the monthly time horizon.

All hypotheses are based on the highest and lowest ranks of the level of the investor sentiment index in a monthly period for selecting stocks in a portfolio as detailed in the

previous section. That is, in each period, the current period's investor sentiment indexes of all stocks in the simulation were used to choose the next period's stocks held in the simulated portfolio. The hypotheses imply the rank of the level of the investor sentiment index to predict abnormal returns (in line with this thesis' definition of investor sentiment in Section 2.2.2) and posit effects on the portfolio level on the monthly mean or median of abnormal returns.

The highest and lowest ranks are considered together and separately in the first three hypotheses. First, a combined portfolio of long and short positions is considered to measure the effect of the ranked investor sentiment index.

Hypothesis (H1): Using the investor sentiment index for selecting the highest (lowest) ranked stocks in a long (short) portfolio results in (H1.1) positive alpha and (H1.2) positive mean or median excess returns.

The next hypothesis considers only long positions based on the highest ranks of the investor sentiment index:

Hypothesis (H2): Using the investor sentiment index for selecting the highest ranked stocks in a long portfolio results in (H2.1) positive alpha and (H2.2) positive mean or median excess returns.

The next hypothesis considers only short positions based on the lowest ranks of the investor sentiment index:

Hypothesis (H3): Using the investor sentiment index for selecting the lowest ranked stocks in a short portfolio results in (H3.1) positive alpha and (H3.2) positive mean or median excess returns (after shorting).

The hypotheses formulated before can in principle be studied with any number of stocks that are selected in a portfolio. The upper bound is determined by the size of the stock universe (of the simulation). To study the relative effect of a differing number of stocks selected in a portfolio, the next hypothesis is formulated based on the intuition that the highest and lowest ranked stocks should have the largest effect on abnormal returns and selecting more stocks necessarily includes stocks on a bad rank, possibly blurring the simulation results negatively on the portfolio level.

Hypothesis (H4): ***Selecting N stocks*** *in a long portfolio and a short portfolio using the investor sentiment index results in (H4.1) higher alpha and (H4.2) positive mean or median of paired differences of excess returns* ***than selecting N+1 stocks****.*

The parameter N in hypothesis H4 (and also in H5, and in H6) refers to the number of stocks to select in each period in a long portfolio using the highest ranked stocks *and* at the same time to the number of stocks to select in a short portfolio using the lowest ranked stocks. The range of N in the above hypothesis is specified in Table 17. The result was evaluated on the combined long-short portfolio, containing $2 \cdot N$ stocks. The posited effects were measured over the whole simulation period in terms of mean or median abnormal returns.

The "higher alpha"-formulation in hypothesis H4 (and in H5, and in H6) is equivalent to the longer formulation of "positive difference of the alphas of the two portfolios", which is used in Section 4.2.3.1 on testing the hypotheses. The "paired differences" of two time series of excess returns in H4 (and in H5, and in H6) is formally defined in Definition (4.18). This measure compares abnormal returns (technically, excess returns vs. the buy-and-hold portfolio's returns) of two portfolios and is equivalent to comparing total returns of the two simulated portfolios directly because the buy-and-hold portfolio's returns are deducted from both portfolios' returns.

The next hypothesis postulates investor sentiment indexes generated from blog documents of the Seekingalpha platform to have a higher effect on abnormal returns of a stock portfolio compared to investor sentiment from blog documents of the Blogspot blog platform. Arguments for this hypothesis have been provided in Section 2.3.2: Seekingalpha is a platform of investment-related blogs only from mostly semi-professional authors (see Section 2.3.2). Because there are editorial rules, Seekingalpha hosts high quality opinionated content (see Section 2.3.2). Because the platform has a large number of readers, the potential price impact is high (see Section 2.3.2). In contrast, Blogspot is not constrained to investment topics, has no editorial rules, and practically anyone can set up a blog on Blogspot (see Section 2.3.2). Thus, the quality of the content, the number of readers, and the potential price impact are presumably lower compared to Seekingalpha.

Hypothesis (H5): ***Using the Seekingalpha investor sentiment index*** *for selecting N stocks in a long portfolio and a short portfolio results in (H5.1) higher alpha and (H5.2) positive mean or median of paired differences of excess returns* ***than using the Blogspot investor sentiment index****.*

The last hypothesis postulates the effect of investor sentiment from blogs on abnormal returns of a stock portfolio to be not just driven by the cross-sectional price momentum effect (Jegadeesh & Titman, 1993, 2001, 2011). The momentum effect is compatible with behavioral explanations due to underreaction to news (Hong et al., 2000). Such misbehavior is considered to be part of investor sentiment (e.g., (Barberis et al., 1998)). In contrast to investor sentiment from blog documents, the price data for the cross-sectional momentum strategy can be collected at much lower costs (in terms of money, time, and computational resources). Thus, this study is interested in whether investor sentiment from blog documents contains more predictive value regarding abnormal returns. The stock selection strategy design is based on the cross-sectional momentum strategy with identical configuration (see Section 4.2.1) except the information used for ranking stocks, making the results comparable. On this basis, the following hypothesis is being formulated:

Hypothesis (H6): ***Using the investor sentiment index*** *for selecting N stocks a long portfolio and a short portfolio results in (H6.1) higher alpha and (H6.2) positive mean or median of paired differences of excess returns* ***than using the stocks' returns****.*

Table 18 summarizes the hypotheses.

Table 18: Overview of hypotheses and sub-hypotheses (abbreviated as H and Sub H).

H	Treatment	Sub H	Posited effect
	Part 1: Analyzing the portfolio simulation results: *"Using the investor sentiment index (in each period) for selecting …*		*… results (over the whole simulation, on the portfolio level) in …*
H1	*… highest (lowest) ranked stocks in a long (short) portfolio …*	H1.1	*… positive alpha."*
		H1.2	*… positive mean or median excess return."*
H2	*…highest ranked stocks in a long portfolio...*	H2.1	*… positive alpha."*
		H2.2	*… positive mean or median excess return."*
H3	*…lowest ranked stocks in a short portfolio…*	H3.1	*… positive alpha."*
		H3.2	*… positive mean or median excess return."*
	Part 2: Comparing different simulation configurations		*… results (over the whole simulation, on the portfolio level) in …*
H4	*"**Selecting N stocks** in a long portfolio and a short portfolio using the investor sentiment index…**than selecting N+1 stocks**."*	H4.1	*… higher alpha …*
		H4.2	*… positive mean or median of paired differences of excess returns …*
H5	*"**Using the Seekingalpha investor sentiment index** for selecting N stocks in a long portfolio and a short portfolio … **than using the Blogspot investor sentiment index**."*	H5.1	*… higher alpha …*
		H5.2	*… positive mean or median of paired differences of excess returns …*
H6	*"**Using the investor sentiment index** for selecting N stocks in a long portfolio and a short portfolio … **than using the stocks' returns**."*	H6.1	*… higher alpha …*
		H6.2	*… positive mean or median of paired differences of excess returns …*

4.2.3 Testing of the Hypotheses

The hypotheses formulated in the preceding subsection are with respect to stock portfolio level abnormal returns in terms of (1) alpha, and (2) mean/median of excess returns versus the buy-and-hold benchmark portfolio's returns. This section discusses respective methods for testing the hypotheses.

4.2.3.1 Alpha-based Tests

Alpha-based tests of hypotheses require making inferences from the portfolio simulation results about a portfolio's mean monthly abnormal returns in terms of alpha. To estimate the alpha, Carhart's (1997) four-factor model (see Section 2.1.2.3) was used, which can be considered state-of-the-art with respect to multifactor models of normal (or theoretically expected) returns, and which was also used in related studies for estimating mean abnormal returns (e.g., (Tetlock et al., 2008)). To estimate the alpha of Carhart's (1997) model on the portfolio level, the regression model of Definition (2.5) in Section 2.1.2.3 was used, with the overall portfolio (consisting of all positions) log returns of one period being defined as:

$$r_P(t) = ln\left(\sum_{i=1}^{n}\left(\frac{1}{n}\cdot e^{pos_i(t)\cdot r_i(t)}\right)\right) + ln(1-tr) \tag{4.1}$$

Variables in Definition (4.1) are defined as follows:

$r_P(t)$: Overall portfolio log return in period t.

$r_i(t)$: Log return of stock i in period t (see Definition (2.3)).

$pos_i(t) \in \{-1,1\}$: Indicates whether a stock i is in the long (i.e., pos=1) or in the short sub-portfolio (i.e., pos=–1) in period t.

n: Number of stocks in the overall portfolio.

tr: Overall portfolio level relative transaction costs (see Section 4.2.1), if any.

t: Discrete time.

For estimating Carhart's model, monthly frequency market data (see Appendix A.5) and the ordinary least squares (OLS) regression method were used. More specifically, the OLS regression method was used to estimate the coefficients (α and $\beta_1, \ldots, \beta_4$) of the model with the estimated coefficients termed a, $b_1, \ldots, b_4$ to distinguish them from the true coefficients.

With respect to the error term values and independent variables of the time series regression model, the **assumptions** required for consistent OLS estimation and hypothesis testing regarding the regression coefficients were made in this thesis (e.g., (Wooldridge, 2013, pp.337–343,372-376,391-392)):

(1) The data-generarting process of independent variables and the dependent variable is a stochastic process and follows Carhart's (1997) model and is stationary and weekly dependent ((Wooldridge, 2013, pp.337,372)).

(2) No independent variable is constant or an exact linear combination of the other independent variables (e.g., (Wooldridge, 2013, pp.338,373)).

(3) The expected value of each error term value $\varepsilon(t)$, given the independent variables for t, is zero (e.g., (Wooldridge, 2013, pp.338,373)).

(4) The variance of each error term value $\varepsilon(t)$, given the independent variables at t, is identical (e.g., (Wooldridge, 2013, pp.340,341,375)). This property is also known as homoskedasticity (e.g., (Wooldridge, 2013, pp.340,375)).

(5) The error term values $\varepsilon(t_1)$ *and* $\varepsilon(t_2)$, given the independent variables for t_1 and t_2, are uncorrelated for all $t_1 \neq t_2$ (e.g., (Wooldridge, 2013, pp.341,375)).

Regarding assumption (1), the aspect of stationarity and weak dependence of the data generating process should be valid, considering that all independent variables and the dependent variable constitute some kind of returns. Because the calculation of returns is a form of first differencing of price time series, the process should be weakly dependent and stationary (e.g., (Wooldridge, 2013, p.384)).

Another aspect of assumption (1), as well as assumptions (2) and (3) refer to correct model specification. Assumption (2) would be violated if independent variables would be *perfectly linearly* dependent or if an independent variable would be constant (Wooldridge,

2013, pp.338,373). This violation should have been prevented by the design of Carhart's (1997) model used in this study.

Assumption (3) would be violated primarily, if an important explanatory independent variable would be missing in the model (e.g., (Wooldridge, 2013, pp.338-339)), leading to an inconsistent estimate of alpha (e.g., (Wooldridge, 2013, pp.373-374)). Because the actual alpha (α) is unknown, the correctness of the assumption was not checked and this study relies on the elaborate four-independent-variable model of Carhart (1997).

Given the above assumptions and based on the central limit theorem (e.g., (Wooldridge, 2013, p.761)), the estimated regression coefficients (using OLS) are asymptotically normally distributed and standard statistical hypothesis testing procedures can be used (e.g., (Wooldridge, 2013, pp.376,391)). For hypothesis testing, a method robust to violations of assumptions (4) and (5) was used. The method is discussed at the end of this subsection, after discussing the concepts for model estimation and statistical inference subsequently.

The objective of the OLS-estimation is to minimize the sum of squared values of the residuals $\hat{\varepsilon}(t)$ (e.g., (Wooldridge, 2013, p.73)), i.e., the sample counterparts of the error term values (e.g., (Wooldridge, 2013, p.53)). With respect to this work's study, OLS minimizes the (sum over all periods of squared) differences between the desired output, i.e., the portfolio return in excess of the risk free rate (which the model of Carhart (1997) aims to explain) and the *actual output* of the model, i.e., the **estimate of the portfolio return** in excess of the risk free rate:

$$\widehat{r_{P-F}}(t) = a + b_1\big(r_I(t) - r_F(t)\big) + b_2 SMB(t) + b_3 HML(t) + b_4 MOM(t) \qquad (4.2)$$

where all variables are defined as in Definition (2.5) and

$\widehat{r_{P-F}}(t)$: Estimate of the portfolio log return in period t in excess of the risk free rate in period t.

a: Estimate of alpha (α) in the model of Carhart (1997).

$b_1, \ldots, b_4$: Estimates of the coefficients $\beta_1, \ldots, \beta_4$ in Carhart's model.

The **residuals** can now be calculated as:

$$\hat{\varepsilon}(t) = r_P(t) - r_F(t) - \widehat{r_{P-F}}(t) \qquad (4.3)$$

where

$r_P(t)$: Empirical portfolio log return in period t.

$r_F(t)$: Empirical risk free log return in period t.

$\widehat{r_{P-F}}(t)$: Estimate of the portfolio log return in period t in excess of the risk free rate in period t.

Formally, the **term to be minimized** by OLS on estimating the coefficients a, b1, …, b4 is (e.g., (Wooldridge, 2013, pp.69,803)):

$$\sum_{t=1}^{n}\left(\hat{\varepsilon}(t)\right)^2 \tag{4.4}$$

where

$\hat{\varepsilon}(t)$: Residual of the model's estimated output value in period t.

n: Number of observations (i.e., number of periods in the portfolio simulation).

The **coefficient of determination, R^2,** measures the quality of the estimated model in terms of the fraction of the explained variation of the dependent variable (e.g., (Wooldridge, 2013, pp.36,76)). Subsequently, R^2 is defined with respect to the model to be estimated, where variables are defined as above (adapted from (Wooldridge, 2013, p.76)):

$$R^2 = 1 - \frac{\sum_{t=1}^{n}\left(\hat{\varepsilon}(t)\right)^2}{\sum_{t_1=1}^{n}\left(\left(r_P(t_1) - r_F(t_1)\right) - \frac{\sum_{t_2=1}^{n}\left(r_P(t_2) - r_F(t_2)\right)}{n}\right)^2} \tag{4.5}$$

To compare the quality of models with a differing number of independent variables, the **adjusted R^2** ((Theil, 1961) cited in (Hossain & Bhatti, 2003, p.93)) can be used (e.g., (Wooldridge, 2013, p.194) for the formula).

Given the above assumptions, the estimate of alpha in Carhart's (1997) model, a, is an asymptotically normally distributed random variable with mean α and variance var_a (e.g., (Wooldridge, 2013, pp.376,112):

$$a \sim N(\alpha, var_a) \tag{4.6}$$

The estimate of the square root of the variance, i.e., the standard deviation, of a is commonly referred to as "**standard error**" (e.g., (Wooldridge, 2013, pp.96,97), denoted by $se(a)$. The standard error is the precision of the estimate a of alpha (e.g., (Wooldridge, 2013, p.97). The method used in this thesis for calculating the standard error is referred to at the end of this subsection. Based on the standard error, the following **test statistic** measures how many (estimated) standard deviations the estimated alpha a is away from a hypothesized true value of alpha, α_0 (e.g., (Wooldridge, 2013, p.122)):

$$ts = \frac{a - \alpha_0}{se(a)} \tag{4.7}$$

Variables in Definition (4.7) are defined as follows:

ts: *t*-statistic value.

a: Estimate of alpha (α) in Carhart's (1997) model.

α_0: Hypothesized true value of alpha (α) in Carhart's (1997) model.

se(a): Standard error of *a*.

With respect to this study's portfolio simulations, the test statistic allows inferencing whether the true alpha is positive (as postulated in hypotheses H1.1, H2.1, H3.1 in Section 4.2.2), i.e., $\alpha>0$ by stating the following opposing **null hypothesis** (e.g., (Wooldridge, 2013, pp.113-120,774)):

$$\mathrm{H0}^{1.1,2.1,3.1}\colon \alpha \leq \alpha_0 \qquad \text{where } \alpha_0=0 \tag{4.8}$$

The test statistic then reduces to:

$$ts = \frac{a}{se(a)} \tag{4.9}$$

Assuming the null hypothesis to be true, the test statistic is asymptotically standard-normal distributed (e.g., (Wooldridge, 2013, pp.376,777)). However, for a sample size up to 60 (like in this study), the *t*-distribution is traditionally assumed (e.g., (Wooldridge, 2013, p.777)), thus (adapted from (Wooldridge, 2013, p.113)):

$$ts \sim td(n-k) \tag{4.10}$$

where

ts: *t*-statistic value.

td(n–k): Student's *t*-distribution with *n–k* degrees of freedom.

n: Number of periods in the portfolio simulation.

k: Number of coefficients of the estimated regression model including alpha, i.e., in case of Carhart's (1997) model, $k=5$.

Given the distribution of the test statistic under the null hypothesis, one can determine the probability of observing a given (or a more extreme) test statistic value. If the test statistic value is large, it is unlikely to observe such a value if the null hypothesis was true and the null hypothesis can be rejected (e.g., (Wooldridge, 2013, pp.115-116)). Because this study is actually interested in evidence regarding positive alpha values in the hypotheses H1.1, H2.1, and H3.1 (defined in Section 4.2.2), positive test statistic values were assumed to enter a positive significance test (in case they were negative, it was rejected anyways): To formally reject the null hypothesis $\mathrm{H0}^{1.1,2.1,3.1}$, a condition based on a **critical value** of the test statistic *ts* at a certain **significance level** *sl* is defined first (adapted from (Wooldridge, 2013, pp.96,115-120,774-778)):

$$ts > td(n-k, 1-sl) \tag{4.11}$$

where

ts: *t*-statistic value.

td(n–k,1–sl): The *(1–sl)*-centile of the *t*-distribution with *n–k* degrees of freedom defines the critical value of the *t*-statistic (e.g., (Wooldridge, 2013, pp.96,115-120,774-778)).

sl: Significance level, e.g., 0.1 (10%).

n: Number of observations (e.g., (Wooldridge, 2013, p.96)), i.e., number of periods in the portfolio simulation.

k: Number of coefficients of the estimated regression model (e.g., (Wooldridge, 2013, p.96)), i.e., in case of Carhart's (1997) model, *k*=5.

Regarding rejecting $H0^{1.1,2.1,3.1}$, this study requires a significance level of 10%. In case, the test statistic exceeded the critical value, the null hypothesis $H0^{1.1,2.1,3.1}$ was rejected at the defined significance level (e.g., (Wooldridge, 2013, p.116)). See Figure 20 for an example of a probability density function of the *t*-distribution that corresponds to this study's portfolio simulation with 60 periods and an estimate of alpha by Carhart's model (i.e., with 5 coefficients).

The probability of observing a *t*-statistic value greater than the critical value is 10% or less in the example presented in Figure 20. The critical value of 1.297 in the example was determined by the inverse cumulative distribution function of the *t*-distribution with 55 degrees of freedom for 90% probability. When the *t*-statistic exceeds the critical value, the hypothesized $\alpha \leq 0$ (i.e., the null hypothesis $H0^{1.1,2.1,3.1}$) can be rejected at the significance level of 10%. Furthermore, the respective **alternative hypothesis** (i.e., H1.1, or H2.1, or H3.1 – see Section 4.2.2 for the long form) is implicitly accepted:

$$\mathrm{H1.1, H2.1, H3.1}: \alpha > 0 \tag{4.12}$$

Regarding hypotheses H4.1, H5.1, and H6.1 (see Section 4.2.2), which respectively **compare two alphas** (α_A and α_B) estimated both by Carhart's (1997) model on two different portfolio return samples (*A* and *B*) (generated by two different stock selection strategies or datasets), the statistical hypothesis test requires the following null hypothesis:

$$H0^{4.1,5.1,6.1}: \alpha_A \leq \alpha_B \tag{4.13}$$

Definition (4.13) is equivalent to:

$$H0^{4.1,5.1,6.1}: \alpha_A - \alpha_B \leq 0 \tag{4.14}$$

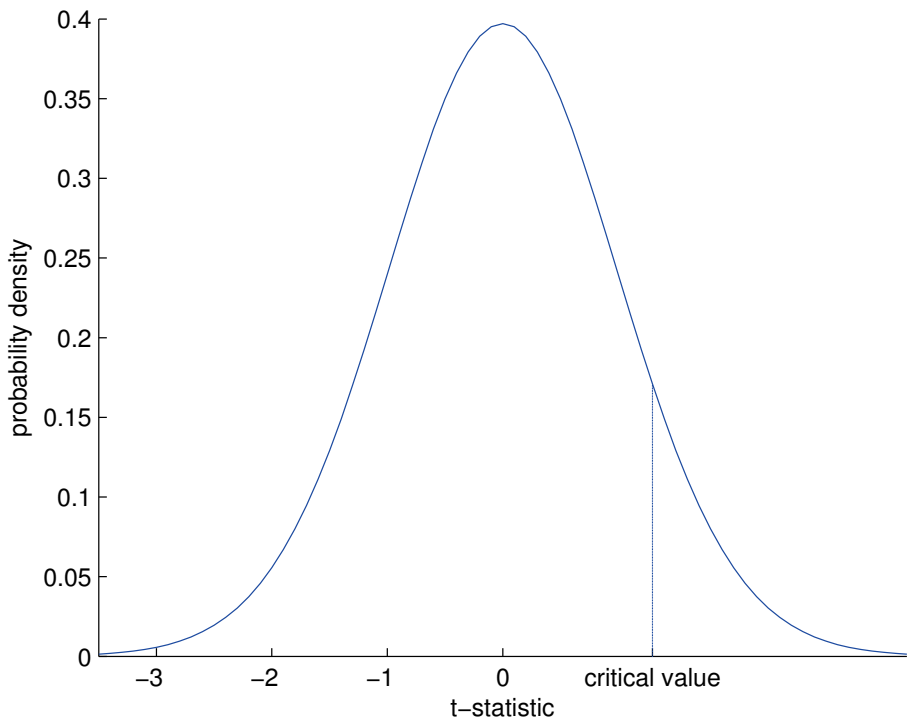


Figure 20: Probability density function of the *t*-distribution with 55 degrees of freedom. The critical value of the *t*-statistic for the (right-hand sided) significance level of 10% is 1.297. That is, the area under the curve on the right hand side of the critical value makes up 10% of the total area. The probability of observing a *t*-statistic greater than the critical value is <10%.

To estimate the difference of alphas $\alpha_A - \alpha_B$ (denoted as α_{A-B} below), a pooled regression method ((Gujarati, 1970a, 1970b) cited in (Gujarati & Porter, 2009, pp.285–288)) was used, based on the Carhart (1997) model:

$$\begin{aligned} r_{P-F}(t) = \alpha_B + ß_{B,1}\big(r_I(t) - r_F(t)\big) + ß_{B,2} \cdot SMB(t) + ß_{B,3} \cdot HML(t) + ß_{B,4} \cdot MOM(t) \\ + \alpha_{A-B} \cdot g(t) + ß_{A-B,1}\big(r_I(t) - r_F(t)\big) \cdot g(t) + ß_{A-B,2} \cdot SMB(t) \cdot g(t) \\ + ß_{A-B,3} \cdot HML(t) \cdot g(t) + ß_{A-B,4} \cdot MOM(t) \cdot g(t) + \varepsilon(t) \end{aligned} \quad (4.15)$$

where

$$g(t) = \begin{cases} 1 & if\ period\ t\ data\ refers\ to\ sample\ A \\ 0 & if\ period\ t\ data\ refers\ to\ sample\ B \end{cases}$$

and

$r_{P-F}(t)$: Estimate of the portfolio log return in period t in excess of the risk free rate in period t.

α_B: Alpha coefficient in the model of Carhart (1997) regarding sample B.

$ß_{B,1},\ldots,ß_{B,4}$: Coefficients in Carhart's model for sample B.

α_{A-B}: Difference between Carhart-alphas in samples A vs. B.

$ß_{A-B,1},\ldots,ß_{A-B,4}$: Difference of coefficients in Carhart's model for sample A vs. sample B.

r_I, r_F, t, SMB, HML, MOM, ε: see Definition (2.5).

Similar to Definition (4.9) the following test statistic was used for testing the null hypothesis $H0^{4.1,5.1,6.1}$:

$$ts = \frac{a_{A-B}}{se(a_{A-B})} \tag{4.16}$$

where

ts: *t*-statistic value.

a_{A-B}: Estimate of the difference between Carhart-alphas (α) in samples *A* vs. *B*.

$se(a_{A-B})$: Standard error of a_{A-B}.

Under the null hypothesis $H0^{4.1,5.1,6.1}$ Definition (4.10) holds. Testing the null hypothesis using the defined test statistic requires extending assumption (4), formulated above: The variance of each error term value $\varepsilon(t)$, given the independent variables at *t*, is identical across sample *A* and *B* (Gujarati & Porter, 2009, p.286; Wooldridge, 2013, pp.340,341,375). If condition (4.11) was met (with n=120 and k=10) for a significance level of 10%, the null hypothesis $H0^{4.1,5.1,6.1}$ was rejected and the respective opposing alternative hypothesis was accepted:

$$\text{H4.1, H5.1, H6.1: } \alpha_A > \alpha_B \tag{4.17}$$

The alternative hypotheses H1.1, H2.1, H3.1, H4.1, H5.1, and H6.1 are the actual hypotheses that this study is interested in. Regarding statistically testing the corresponding null-hypotheses, two **types of errors** are possible: (**type I**) rejecting the null-hypothesis in case it is true (e.g., (Wooldridge, 2013, pp.771-774)) or (**type II**) not rejecting the null-hypothesis in case it is false (e.g., (Wooldridge, 2013, pp.771-774)). The maximum *tolerable* probability of a type I error is expressed by the **significance level** (e.g., (Wooldridge, 2013, p.773)). "The smallest significance level at which the null hypothesis can be rejected […]" is determined by the *p*-value (Wooldridge, 2013, p.848). Thus, the ***p*-value** is the *actual* probability of a type I error (e.g., (Wooldridge, 2013, pp.773,778-781)). Furthermore, the *p*-value is the probability of observing a given or greater than given test statistic value under the null hypothesis (e.g., (Wooldridge, 2013, p.779)). The **power of a statistical test** is the probability of the opposing event of a type II error (e.g., (Wooldridge, 2013, p.773)).

The power to detect positive (mean monthly) abnormal returns in terms of positive alpha is limited on using a right-tailed *t*-test regarding multi-factor models' alpha (Kothari & Warner, 2001). Using simulated funds in which they gradually introduced abnormal returns, they found the Carhart model to have a 96% (86%) identification rate of alpha greater than zero at the 5% (1%) significance level to require 5% annual abnormal returns in a scenario of a fund's portfolio selected from large capitalization stocks (Kothari & Warner, 2001, pp.2002–2003). For lower annual abnormal returns, the identification rate (of true positives) would be much lower (Kothari & Warner, 2001). That is, even if in this study's portfolio

simulations no evidence for (positive) alpha was found, this does not mean that (positive) alpha does not exist.

Type I errors are regarded to be more important with respect to this study's hypothesis tests than type II errors. To this respect, Kothari and Warner (2001) show that multi-factor alpha measures of (mean) abnormal returns are well specified in general. However, using simulated funds in which they gradually introduced abnormal returns, they found Carhart's model to falsely identify alpha greater than zero at the 5% (1%) significance level in 10% (4%) of cases in a scenario of a fund's portfolio selected from large capitalization stocks (Kothari & Warner, 2001, pp.2002–2003).

Regardless of the validity of assumptions (4) and (5) required for hypothesis testing, the coefficients of Carhart's (1997) model and the variant of Definition (4.15) were always estimated by OLS with estimates being consistent also in the cases of heteroscedasticity and autocorrelation of error terms (e.g., (Wooldridge, 2013, pp.391-392)).

For estimating the standard errors, the heteroskedasticity and autocorrelation-consistent method of Newey & West (NW) (Newey & West, 1987) was used. That is, the validity of the NW method does not depend on assumption (4) or assumption (5). The estimated standard errors are the central ingredient in the *t*-statistics defined in (4.9) and in (4.16). On the basis of the *t*-statistic values, the hypotheses formulated in Section 4.2.2 were tested as described above. This approach can be considered valid if the sample size is large (e.g., (Wooldridge, 2013, p.391)). The sample size of this thesis' study is 60 regarding estimations of alpha and 120 regarding estimations of alpha differences. Thus, it should be large enough. Still, the asymptotic validity of hypothesis test results has to be kept in mind. Furthermore, the measures of the quality of the estimated model, R^2 and adjusted R^2, are not affected by violations of assumption (4) (Wooldridge, 2013, pp.258,259). The measures are also not affected by violations of assumption (5) in case of stationary and weakly dependent data (Wooldridge, 2013, p.400).

4.2.3.2 Benchmark-based Tests

Benchmark-based tests of the hypotheses H1.2, H2.2, H3.2, H4.2, H5.2, and H6.2, formulated in Section 4.2.2, are with respect to the mean or the median of the simulation's portfolio log returns in excess of log returns of a benchmark portfolio. Whereas H1.2, H2.2, and H3.2 are about the existence of positive excess log returns (vs. the buy-and-hold passive benchmark portfolio's log returns), H4.2, H5.2, and H6.2 compare excess log returns of the simulated portfolio using investor sentiment (indexes) against another (active benchmark) portfolio's excess log returns.

Paired samples tests were considered for the hypothesis tests because the respective two samples of portfolio returns being compared cannot be assumed to be independent. That is, regarding H1.2, H2.2, and H3.2 the simulated portfolio (using investor sentiment indexes) is a subset of the benchmark portfolio. Regarding H4.2, the portfolios to be compared are

identical apart from one stock. Regarding H5.2 and H6.2, the portfolios to be compared might contain also the same stocks to a large degree over many periods, depending on the datasets used for selecting these stocks. For testing the null hypotheses of H1.2, H2.2, H3.2, H4.2, H5.2, and H6.2 two types of tests were considered: (1) a parametric *t*-test, and (2) a nonparametric test.

The null hypothesis of the paired samples *t*-test (e.g., (Downing & Clark, 2010, pp.315–316)) regarding the alternative hypotheses H1.2, H2.2, and H3.2 is: the mean of paired differences of the time series values of the simulated portfolio's log returns (using investor sentiment indexes) and the (buy-and-hold) benchmark portfolio's log returns is zero or negative. The **paired differences** of time series values from two samples required for this test can be defined generically as:

$$pd(t) = x_A(t) - x_B(t) \tag{4.18}$$

where

pd: Paired difference.

$x_A(t)$: Value *x* of sample *A* in period *t*.

$x_B(t)$: Value *x* of sample *B* in period *t*.

With respect to H1.2, H2.2, and H3.2, the paired differences of the time series values of the simulated (investor sentiment-based) portfolio's returns versus the buy-and-hold portfolio's returns result in the excess return time series against which the original hypotheses H1, H2, and H3 were formulated (in Section 4.2.2). Thus, the formulation of the null hypothesis of H1.2, H2.2, and H3.2, of the mean of the excess return time series being zero or negative is equivalent to the one based on paired differences. The paired samples *t*-test of the null hypothesis assumes the excess return time series to be normally distributed (e.g., (Downing & Clark, 2010, pp.315–316)). To verify this assumption, the Anderson-Darling test (Anderson & Darling, 1952, 1954) and the Jarque-Bera test (Jarque & Bera, 1987) were used. The version of the test of Anderson & Darling described in (D'Agostino & Stephens, 1986) was used. The two tests are complementary to some degree because they have different power to detect different non-normal distributions (Yazici & Yolacan, 2007). Whereas the power of the test of Anderson and Darling is very high for many non-normal distributions for sample sizes of 50, the power of the test of Jarque and Bera can be lower and increases substantially with sample size (Yazici & Yolacan, 2007). The null hypothesis of normality of these tests was rejected at the 10% significance level. In case the assumption of a normal distribution was indicated to hold, the *p*-value of the paired samples *t*-test of the null hypothesis was derived based on the cumulative distribution function of the *t*-distribution with 59 degrees of freedom (following (Downing & Clark, 2010, p.316)). For rejecting the null hypothesis of the paired samples *t*-test, the *p*-value was required to indicate a 10% significance level (or better).

As a nonparametric alternative to the paired samples *t*-test, the paired samples variant of the signed rank test (SRT) of Wilcoxon (Wilcoxon, 1945) was used with the midrank method to treat ties, i.e., observations with identical magnitude (e.g., (Gibbons & Chakraborti, 2011, pp.193-194)). Wilcoxon's signed rank test does not require the normality assumption of the *t*-test (e.g., (Gibbons & Chakraborti, 2011, pp.157,210-211)). To determine the statistical significance by the p-value of Wilcoxon's test statistic, an approximation based on the standardized test statistic (corrected for ties), i.e., the z-score, which is asymptotically standard-normally distributed (e.g., (Gibbons & Chakraborti, 2011, pp.196,202,212)), was used using a continuity correction (e.g., (Gibbons & Chakraborti, 2011, pp.202,212)) and a correction for ties (e.g., (Gibbons & Chakraborti, 2011, pp.202-203)). A right-tailed test was performed with respect to H1.2, H2.2, and H3.2, testing the null hypothesis of the median (of paired) difference(s) of the simulated portfolio's log returns (using investor sentiment indexes) and the benchmark portfolio's log returns (i.e., excess log returns) being zero or negative. The null hypothesis was rejected for *p*-values indicating a 10% significance level (or better).

For hypotheses H4.2, H5.2, and H6.2, the log returns of the investor sentiment index-based portfolio simulations in excess of the buy-and-hold portfolio's log returns needed to be compared to other portfolios' log returns in excess of the buy-and-hold portfolio's log returns. Regarding (the alternative) hypothesis H4.2, this study tested the null hypothesis of the paired differences (see Definition (4.18)) of investor sentiment index-based portfolio's excess returns with N stocks selected in each period (i.e., sample A) and investor sentiment index-based portfolio's excess returns with N+1 stocks selected in each period (i.e., sample B) to be zero or negative on the monthly mean or median. Regarding (the alternative) hypothesis H5.2, this study tested the null hypothesis of the paired differences of the investor sentiment index-based portfolio's excess returns using the Seekingalpha dataset (i.e., sample A) and investor sentiment index-based portfolio's excess returns using the Blogspot dataset (i.e., sample B) to be zero or negative on the monthly mean or median. Regarding (the alternative) hypothesis H6.2, this study tested the null hypothesis of the paired differences of the investor sentiment-based portfolio's excess returns (i.e., sample A) and the active benchmark price momentum portfolio's excess returns (i.e., sample B) to be zero or negative on the monthly mean or median. For the test of the null hypotheses of H4.2, H5.2, and H6.2 regarding the monthly mean, the paired samples *t*-test on the time series of paired differences was considered as described above. For the test of the null hypotheses of H4.2, H5.2, and H6.2 regarding the monthly median, Wilcoxon's signed rank test on the time series of paired differences was used as described above.

4.3 Results

For each hypothesis formulated in Section 4.2.2, results of the portfolio simulation (presented in Section 4.2.1) and results of a hypothesis test using (1) the Seekingalpha dataset, and (2) the Blogspot dataset are reported. Regarding H1, H2, and H3, results of tests for the existence

of positive abnormal returns in terms of (1) alpha, and (2) mean/median returns of the simulated portfolio in excess of a passive benchmark portfolio's returns are reported. For the portfolio simulation regarding H1, H2, and H3, the baseline configuration (see Table 17) of choosing five stocks in a long or short portfolio was used. Regarding H4, H5, and H6, results of tests for the difference of abnormal returns of the simulated portfolio using investor sentiment indexes relative to an active benchmark portfolio's abnormal returns are reported. The simulations assume (1) no transaction costs, and (2) 50 bps transaction costs.

4.3.1 Hypothesis H1: Effects on a Long & Short Portfolio

This section reports results of the long-and-short-portfolio simulation using the baseline configuration of selecting five stocks per period in the long and short sub-portfolios. Furthermore, results of tests of hypotheses H1.1 and H1.2 on (1) the Seekingalpha dataset, and (2) the Blogspot dataset are reported. Robustness checks of H1 with respect to different portfolio sizes are reported in Section 4.3.4 in conjunction with relative effects of altering the baseline configuration in terms of the number of stocks selected in a long or short sub-portfolio.

4.3.1.1 Seekingalpha

Ignoring transaction costs, the estimated alpha of the portfolio simulation on the Seekingalpha dataset indicates positive mean monthly abnormal returns of 0.83% of the combined long and short sub-portfolios, each of which held in each period the 5 highest ranked stocks vs. the 5 lowest ranked stocks using the level of the investor sentiment index in the previous period. The result is statistically significant at the 1% level. Consequently, the lower bound of the 90% confidence interval of the alpha estimate is 0.49%, and thus also positive. That is, the null hypothesis of a negative or zero alpha was rejected. Thus, the evidence supports H1.1 for the Seekingalpha dataset, when ignoring transaction costs. Also the goodness of fit of the regression is high as suggested by the *F*-test and a reasonable level of R^2. See Table 19 for details.

Adjusting the simulated portfolio returns for transaction costs (as discussed in Section 4.2.1) resulted in a reduced estimate of alpha of 0.33% (see Table 19). Whereas alpha was still estimated to be positive, the statistical evidence does not allow rejecting the null hypothesis of a negative or zero alpha.

Table 19: H1.1 *t*-test results for Seekingalpha regarding the long-short portfolio.

Coefficients	Null hypothesis	90% CI bound	Estimate	Statistic	p	Reject null?
With transaction costs						
a	$\alpha \leq 0$	–0.0001	0.0033	1.2457	0.1091	No
Without transaction costs						
a	$\alpha \leq 0$	0.0049	0.0083	3.1541	0.0013***	Yes
b_1			–0.1080	–1.2517		
b_2			–0.0261	–0.1727		
b_3			–0.6834	–2.7627		
b_4			–0.0538	–1.1578		
Goodness of fit						
R^2			0.4851			
Adj. R^2			0.4476			
F-test	All $ß_i$ coefficients are zero			12.9523	<0.0001***	Yes

NOTES: "Coefficients" refers to the coefficients in Carhart's (1997) model. "Estimate" is the OLS-estimate of the respective coefficient: a = estimate of α; b_1 = estimate of the $ß_1$ coefficient with respect to (wrt.) the $r_I – r_F$ factor; b_2 = estimate of the $ß_2$ coefficient wrt. the SMB factor; b_3 = estimate of the $ß_3$ coefficient wrt. the HML factor; b_4 = estimate of the $ß_4$ coefficient wrt. the MOM factor; The estimates of coefficients b_1, …, b_4 and goodness of fit are basically the same for simulations with vs. without transaction costs because transaction costs were assumed 50 bps per period on the portfolio level. The rest of the columns refer to testing the stated null hypotheses in the column "Null hypothesis". The *p*-value is wrt. a (right-tailed *t*-)test regarding the null hypothesis of $\alpha \leq 0$ and the respective null hypothesis of the the *F*-test. "Reject null" refers to whether the respective null hypothesis is rejected given the evidence at the specified significance level (i.e., 10% for the test regarding α and 1% for the *F*-test). Significance levels are denoted as follows: *** = 1%, ** = 5%, * = 10%. Hypothesis test results wrt. α are based on a *t*-test using Newey & West (1987) standard errors. Based on the estimate of the standard error of the estimate of α, the *t*-statistic, the *p*-value, and the CI (i.e., confidence interval lower bound of the coefficient estimate) were calculated. The upper bound CI is $+\infty$. The CI and *p*-values are reported only for α, against which a hypothesis was formulated. The column "Statistic" reports the *t*-statistic for all coefficients and the *F*-statistic for the *F*-test. The number of observations of the dependent and independent variables in Carhart's (1997) model was 60.

This result is corroborated by the plot in Figure 21 of the cumulative portfolio log returns of the simulated (combined long and short sub-) portfolio exceeding the returns of the buy-and-hold benchmark portfolio substantially. Because the buy-and-hold portfolio returns closely resemble the ones of U.S. stock market indexes such as the DJIA (not shown), the simulated portfolio also outperforms these indexes on a cumulated return basis. However, when accounting for transaction costs in the simulated portfolio, the outperformance is not that clear. Note that the buy-and-hold portfolio return time series was adjusted for transaction costs (in the first and last period of the simulation).

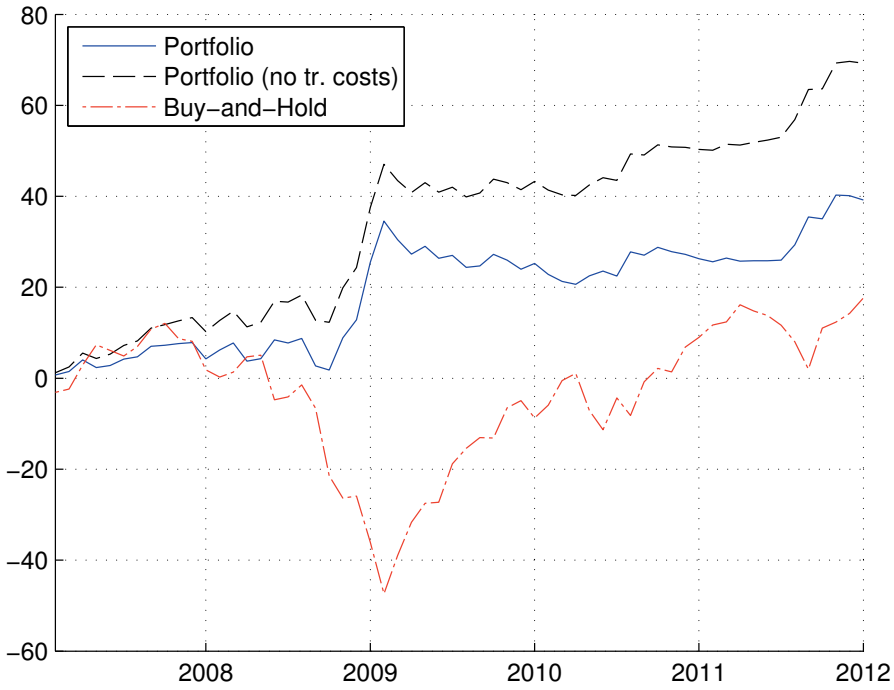


Figure 21: Plot of the returns of the Seekingalpha long-short portfolio (adjusted for/without transaction costs) vs. the buy-and-hold portfolio. The returns are cumulative log returns in %.

Table 20 allows analyzing results more closely. The mean monthly excess return is 0.85% (0.36%) ignoring (accounting for) transaction costs. The level of mean monthly excess return is closely related to the estimates of alpha, supporting the findings above. Because the median of the monthly excess returns is negative, it suggests at least a few monthly periods with large positive excess returns to exist. The level of risk of the simulated portfolio (estimated by the standard deviation of the annual returns) is much lower compared to the one of the buy-and-hold portfolio. Consequently, the Sharpe ratio is considerably higher. Even when accounting for transaction costs, the simulated portfolio is more favorable to this respect. However, excess returns of 39% in 2008 basically produce the difference. Excess returns in other years of the simulation time period except 2011 are negative. Obviously, shorting low investor sentiment stocks and going long on high investor sentiment stocks during the 2008 financial crisis resulted in net positive returns. That is, the possibly negative returns of the long positions (in a generally negative market environment with most stock prices falling) were exceeded by the positive returns of the short positions. Afterwards, during the generally positive market environment of 2009, presumably all stocks had positive returns. Thus, combining the equally weighted long and short portfolios resulted in only slightly positive returns. That is, the simulated portfolio returns were kind of independent from the overall market movements – as intended by the design of the long-short strategy (see Section 4.2.1). The drawback is that in 2009 and in 2010 the simulated portfolio did not profit to the full extent from the general trend of stock prices increasing. To differentiate effects of investor sentiment on the long and short portfolios, effects with respect to these portfolios were studied separately and are reported in Section 4.3.2 and in Section 4.3.3.

Table 20: Excess returns and performance measures (Seekingalpha long-short portfolio).

	Without transaction costs			With transaction costs		
	Portfolio	**BaH**	**Excess**	**Portfolio**	**BaH**	**Excess**
Annual return (%)						
2007	13.35	8.34	5.01	7.85	8.09	–0.25
2008	10.99	–34.02	45.01	4.98	–34.02	39.00
2009	17.13	21.01	–3.87	11.13	21.01	–9.88
2010	9.27	11.74	–2.47	3.27	11.74	–8.48
2011	18.92	7.41	11.51	12.91	7.41	5.50
2012	–0.43	3.68	–4.10	–0.93	3.43	–4.35
Total	69.24	18.16	51.08	39.20	17.66	21.55
Median/mean return (%)						
Monthly median	0.79	0.89	–0.48	0.29	0.89	–0.99
Monthly mean	1.15	0.30	0.85	0.65	0.29	0.36
Performance measures						
Ann. ret. std. (%)	6.89	19.08		5.14	19.06	
Sharpe ratio	0.34	0.04		0.18	0.04	

NOTES: The performance of the portfolio simulation is reported in terms of annual and total returns, monthly median/mean returns, annual return standard deviation in percent ("Ann ret. std. (%)") as a measure of risk, and the *monthly* Sharpe (1966) ratio calculated using risk free returns described in Appendix A.5. Note that the Sharpe ratio was shown to be adequate for relative rankings also in case of non-normally distributed portfolio returns (Eling & Schuhmacher, 2007). All returns are percentage portfolio log returns. The monthly returns of the portfolio simulation range from February 2007 until January 2012 because the investor sentiment datasets, which were used to select stocks in portfolios, cover the time period from January 2007 until December 2011. Thus, the 2012 annual return stems from January only. Excess returns were benchmarked against the buy-and-hold ("BaH") portfolio, which held equal-weighted long positions without rebalancing in all 29 DJIA stocks from the stock universe of the portfolio simulation. Accounting for transaction costs meant deducting 50 bps per period for the investor sentiment-based portfolio simulation and 25 bps in the first and 25 bps in the last period of the BaH-simulation on the portfolio level.

Finally, Table 21 reports on test results on the existence of negative or zero median monthly excess returns (of the simulated portfolio vs. the buy-and-hold portfolio). The normality assumption of a possible *t*-test was rejected by both tests reported. Thus, only Wilcoxon's signed rank test (SRT) was used, suggesting that the null hypothesis of zero or negative median excess returns cannot be rejected.

Table 21: Test results for positive excess returns (Seekingalpha long-short portfolio).

Test	Null hypothesis	Without transaction costs			With transaction costs		
		Statistic	p	Reject?	Statistic	p	Reject?
Verification of the normality assumption of the *t*-test							
AD-test	Normal distr. of excess returns	1.1216	0.0056	Yes	1.1170	0.0057	Yes
JB-test		13.978	0.0079	Yes	13.920	0.0075	Yes
Nonparametric test							
Wilcox. SRT	Median(exc. ret.)≤0	0.2098	0.4169	No	–0.372	0.6450	No

NOTES: The assumption of the *t*-test of a normal distribution of excess log returns was tested by the AD-test (i.e., Anderson & Darling, 1952, 1954; D'Agostino & Stephens, 1986) and the JB-test (i.e., Jarque & Bera, 1987). The respective test statistics are reported in the "Statistic" column. Rejecting the null hypothesis of normality required a 10% significance level (or better). In case of normality, the parametric *t*-test would have tested the null hypothesis of zero or negative mean excess log returns ("exc.ret") of the simulated portfolio. Excess log returns were calculated as paired difference time series of the simulated portfolio's returns using investor sentiment indexes vs. the buy-and-hold portfolio's returns. In case of non-normality, Wilcoxon's (1945) signed rank test ("Wilcox. SRT") offers an alternative to the *t*-test for non-parametrically testing the null hypothesis of zero or negative median excess log returns of the simulated portfolio. The reported "Statistic" is a *z*-score regarding the SRT. The null hypothesis of the SRT was rejected at a *p*-value indicating a 10% or better significance level. Accounting for transaction costs meant deducting 50 bps per period for the investor sentiment-based portfolio simulation and 25 bps in the first and 25 bps in the last period of the buy-and-hold-simulation on the portfolio level.

Summarizing, the alpha-based evidence clearly supports hypothesis H1.1 in the case of ignoring transaction costs for the Seekingalpha dataset. When accounting for transaction costs, H1.1 cannot be accepted. The benchmark-portfolio-based evidence does not support H1.2 (either ignoring or accounting for transaction costs).

4.3.1.2 Blogspot

Ignoring transaction costs, the estimated alpha of the portfolio simulation on the Blogspot dataset indicates positive mean monthly abnormal returns of 0.21% of the combined long and short portfolio. However, the result is statistically not significant (at the 10% significance level). Adjusting the simulated portfolio returns for transaction costs resulted in a negative estimate of alpha. See Table 22 for details. That is, in either case, the null hypothesis of a negative or zero alpha was *not* rejected. Thus, the evidence does not support H1.1 for the Blogspot dataset.

The goodness of fit of the estimated model (see Table 22) is much worse than the goodness of fit of the regressed model on the Seekingalpha dataset. Also, the null hypothesis of the F-test was not rejected. Thus, the alternative benchmark-portfolio-based analysis of abnormal performance is considered next.

Table 22: H1.1 *t*-test results for Blogspot regarding the long-short portfolio.

Coefficients	Null hypothesis	90% CI bound	Estimate	Statistic	p	Reject null?
With transaction costs						
a	$\alpha \leq 0$	–0.0063	–0.0029	–1.1142	0.8650	No
Without transaction costs						
a	$\alpha \leq 0$	–0.0013	0.0021	0.8146	0.2094	No
b_1			0.0868	1.6835		
b_2			–0.0641	–0.7295		
b_3			–0.1489	–1.5027		
b_4			0.0323	0.6310		
Goodness of fit						
R^2			0.0533			
Adj. R^2			–0.0156			
F-test	All $ß_i$ coefficients are zero			0.7740	0.5468	No

NOTES: see notes below Table 19.

Figure 22 shows the cumulative log returns of the simulated (long-short) portfolio – ignoring transaction costs – to exceed the cumulative log returns of the buy-and-hold portfolio in most periods. When accounting for transaction costs, the buy-and-hold portfolio's cumulative log returns exceed the ones of the simulated portfolio beginning in 2010.

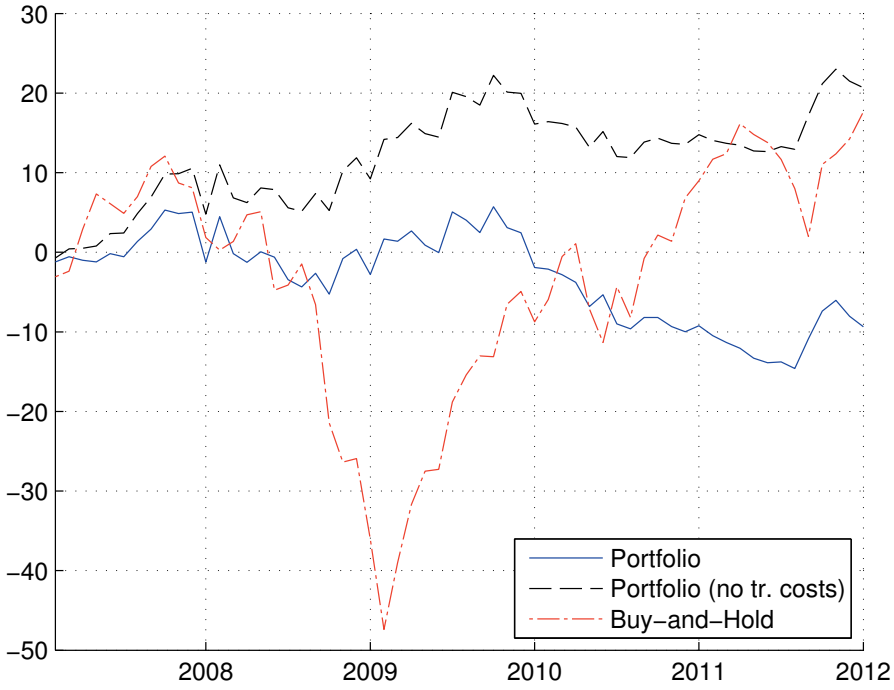


Figure 22: Plot of the returns of the Blogspot long-short portfolio (adjusted for/without transaction costs) vs. the buy-and-hold portfolio. The returns are cumulative log returns in %.

Ignoring transaction costs, the Sharpe ratio is higher, the total return is higher, and the standard deviation of the annual return is lower for the simulated long-short portfolio compared to the buy-and-hold portfolio (see Table 23 for details). The mean monthly excess

log return is positive at a low level of 0.04%. This level is well below the alpha estimate of 0.21%, and thus the more conservative estimate.

Table 23: Excess returns and performance measures (Blogspot long-short portfolio).

	Without transaction costs			With transaction costs		
	Portfolio	**BaH**	**Excess**	**Portfolio**	**BaH**	**Excess**
Annual return (%)						
2007	10.55	8.34	2.21	5.04	8.09	–3.05
2008	1.32	–34.02	35.34	–4.69	–34.02	29.34
2009	8.09	21.01	–12.92	2.08	21.01	–18.92
2010	–6.41	11.74	–18.16	–12.42	11.74	–24.17
2011	7.95	7.41	0.55	1.95	7.41	–5.46
2012	–0.84	3.68	–4.51	–1.34	3.43	–4.76
Total	20.67	18.16	2.51	–9.37	17.66	–27.03
Median/mean return (%)						
Monthly median	–0.11	0.89	–0.35	–0.61	0.89	–0.85
Monthly mean	0.34	0.30	0.04	–0.16	0.29	–0.45
Performance measures						
Ann. ret. std. (%)	6.52	19.08		6.28	19.06	
Sharpe ratio	0.10	0.04		–0.11	0.04	

NOTES: see notes below Table 20.

A paired samples *t*-test could bring light to the question of the existence of positive abnormal returns. The *t*-test is about the null hypothesis stating that the mean of the paired differences of the time series of the simulated portfolio's log returns and the buy-and-hold portfolio's log returns (i.e., excess log returns) are zero or negative. However, the normal distribution assumption of the *t*-test is not valid as suggested by the AD-test and the JB-test (with $p<0.04$ in all cases). Thus, Wilcoxon's signed rank test was used. The test does not reject its null hypothesis with the *p*-value of the test for the simulation ignoring (accounting for) transaction costs being 0.72 (0.89). The null hypothesis of the test states that the median of the paired differences of the time series of the simulated portfolio's log returns and the buy-and-hold portfolio's log returns (i.e., excess log returns) are zero or negative. Table 24 provides all test results.

Table 24: Test results for positive excess returns (Blogspot long-short portfolio).

Test	**Null hypothesis**	**Without transaction costs**			**With transaction costs**		
		Statistic	**p**	**Reject?**	**Statistic**	**p**	**Reject?**
Verification of the normality assumption of the *t*-test							
AD-test	Normal distr. of excess returns	0.8621	0.0249	Yes	0.8596	0.0253	Yes
JB-test		6.5379	0.0337	Yes	6.5041	0.0328	Yes
Nonparametric test							
Wilcox. SRT	Median(exc. ret.)≤0	–0.578	0.7183	No	–1.233	0.8912	No

NOTES: see notes below Table 21.

Summarizing, there is no evidence rejecting the null hypothesis of zero or negative alpha or for rejecting the null hypothesis of the median of monthly excess (log) returns being zero or negative for the long-short portfolio simulation on the Blogspot dataset either ignoring or

adjusting for transaction costs. Therefore, H1.1 and H1.2 are not accepted for the Blogspot dataset.

4.3.2 Hypothesis H2: Effects on a Long Portfolio

This section reports results of the long portfolio simulation and tests of hypotheses H2.1 and H2.2 on (1) the Seekingalpha dataset, and (2) the Blogspot dataset. The long portfolio held in each period the five highest ranked stocks (according the baseline configuration) using the level of the investor sentiment index in the previous period.

4.3.2.1 Seekingalpha

Ignoring (accounting for) transaction costs, the estimated alpha of the long portfolio simulation on the Seekingalpha dataset indicates positive mean monthly abnormal returns of 0.79% (0.29%). The hypothesis test result (rejecting zero or negative alpha) is statistically significant at the 1% (10%) level when ignoring (accounting for) transaction costs. See Table 25 for details. Thus, the evidence supports H2.1 for the Seekingalpha dataset, irrespective of transaction costs. This result was achieved despite the level of estimated alpha being slightly lower compared to the level of estimated alpha of the long-short portfolio simulation. However, the drawback of the long only portfolio is revealed below in Table 26.

Table 25: H2.1 *t*-test results for Seekingalpha regarding the long portfolio.

Coefficients	Null hypothesis	90% CI bound	Estimate	Statistic	p	Reject null?
With transaction costs						
A	$\alpha \leq 0$	0.0000	0.0029	1.3081	0.0981*	Yes
Without transaction costs						
A	$\alpha \leq 0$	0.0050	0.0079	3.5647	0.0004***	Yes
b_1			1.0309	15.6552		
b_2			–0.3917	–2.9789		
b_3			–0.1448	–1.3897		
b_4			0.0095	0.2136		
Goodness of fit						
R^2			0.8375			
Adj. R^2			0.8257			
F-test	All $ß_i$ coefficients are zero			70.8512	<0.0001***	Yes

NOTES: see notes below Table 19.

The goodness of fit of the regression is high as suggested by the *F*-test and the adjusted R^2 (see Table 25). In fact, the goodness of fit is substantially higher than the one for the long-short portfolio regression. The higher fit might be related to a higher correlation of the simulated portfolio's cumulative log returns with the buy-and-hold portfolio's cumulative log returns (see Figure 23), which is closely related to the market proxy variable in Carhart's (1997) model, which is supposed to explain the simulated portfolio's log returns (in excess of the risk free rate).

The plot in Figure 23 shows the cumulated portfolio log returns of the simulated long portfolio exceeding the buy-and-hold portfolio's cumulated log returns. However, when

accounting for transaction costs in the simulated portfolio, the outperformance is much smaller and basically only observable in the second half of 2010 and in 2011.

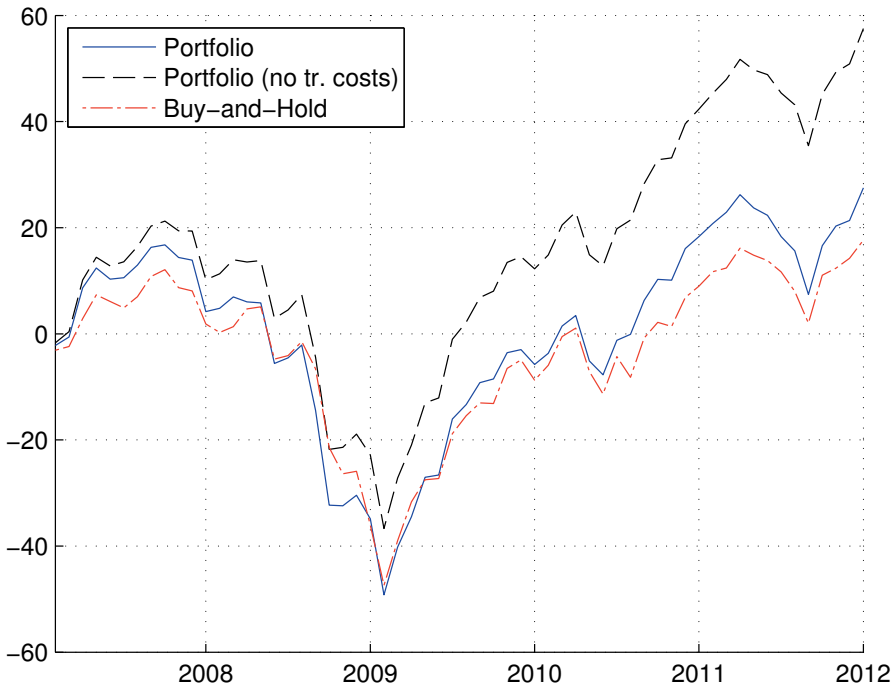


Figure 23: Plot of the returns of the Seekingalpha long portfolio (adjusted for/without transaction costs) vs. the buy-and-hold portfolio. The returns are cumulative log returns in %.

Table 26 allows analyzing simulation results more closely.

Table 26: Excess returns and performance measures (Seekingalpha long portfolio).

	Without transaction costs			With transaction costs		
	Portfolio	**BaH**	**Excess**	**Portfolio**	**BaH**	**Excess**
Annual return (%)						
2007	19.39	8.34	11.05	13.88	8.09	5.79
2008	–38.29	–34.02	–4.27	–44.30	–34.02	–10.28
2009	33.44	21.01	12.44	27.43	21.01	6.43
2010	25.05	11.74	13.30	19.04	11.74	7.29
2011	11.30	7.41	3.90	5.30	7.41	–2.11
2012	6.64	3.68	2.96	6.14	3.43	2.71
Total	57.53	18.16	39.37	27.49	17.66	9.83
Median/mean return (%)						
Monthly median	1.60	0.89	0.63	1.10	0.89	0.13
Monthly mean	0.96	0.30	0.66	0.46	0.29	0.16
Performance measures						
Ann. ret. std. (%)	25.33	19.08		25.34	19.06	
Sharpe ratio	0.15	0.04		0.06	0.04	

NOTES: see notes below Table 20.

In Table 26 the mean monthly and median monthly excess returns are positive and the Sharpe ratio is higher for the simulated long portfolio compared to the buy-and-hold portfolio. When accounting for transaction costs, the Sharpe ratio favors the simulated portfolio only slightly. Because the simulation holds long positions in each period, it suffers

from the negative overall stock market environment in 2008. Ignoring 2008 and transaction costs, every year observed positive excess returns. This observation indicates a consistently positive investment value of high level investor sentiment index values from Seekingalpha. However, the Sharpe ratio is substantially lower compared to the combined long-short portfolio simulation's one. Thus, the short portfolio simulation is studied separately in Section 4.3.3.

Finally, Table 27 shows Wilcoxon's signed rank test rejecting the null hypothesis of zero or negative median excess returns when ignoring transaction costs. The null hypothesis was not rejected when accounting for transaction costs.

Table 27: Test results for positive excess returns (Seekingalpha long portfolio).

Test	Null hypothesis	Without transaction costs			With transaction costs		
		Statistic	p	Reject?	Statistic	p	Reject?
Verification of the normality assumption of the *t*-test							
AD-test	Normal distr. of excess returns	0.8978	0.0203	Yes	0.8758	0.0230	Yes
JB-test		9.9172	0.0165	Yes	9.5406	0.0168	Yes
Nonparametric test							
Wilcox. SRT	Median(exc. ret.)≤0	2.6907	0.0036	Yes	0.6515	0.2574	No

NOTES: see notes below Table 21.

Summarizing, the alpha-based evidence clearly supports hypothesis H2.1 with respect to the Seekingalpha dataset irrespective of transaction costs. The benchmark-portfolio-based evidence also highly rejects the null hypothesis of H2.2 (regarding median excess returns) at the 1% significance level when ignoring transaction costs.

4.3.2.2 Blogspot

The estimated alpha of the portfolio simulation on the Blogspot dataset indicates negative mean monthly abnormal returns. Also the confidence interval lower bound is negative (ignoring or accounting for transactions). See Table 28 for details.

Table 28: H2.1 *t*-test results for Blogspot regarding the long portfolio.

Coefficients	Null hypothesis	90% CI bound	Estimate	Statistic	p	Reject null?
With transaction costs						
a	$\alpha \leq 0$	–0.0106	–0.0064	–1.9877	0.9741	No
Without transaction costs						
a	$\alpha \leq 0$	–0.0056	–0.0014	–0.4435	0.6704	No
b_1			1.0772	20.7594		
b_2			–0.1776	–1.4018		
b_3			0.0420	0.2307		
b_4			0.0385	0.7227		
Goodness of fit						
R^2			0.8234			
Adj. R^2			0.8106			
F-test	All $ß_i$ coefficients are zero			64.1239	<0.0001***	Yes

NOTES: see notes below Table 19.

Based on results in Table 28, the null hypothesis of a negative or zero alpha is *not* rejected. Thus, the evidence does not support H2.1 for the Blogspot dataset. The goodness of fit of the OLS regression is high (see Table 28). The regression result is corroborated by Figure 24, showing the simulated portfolio's cumulative log returns to be smaller than the buy-and-hold portfolio ones' in most of the periods, even when ignoring transaction costs.

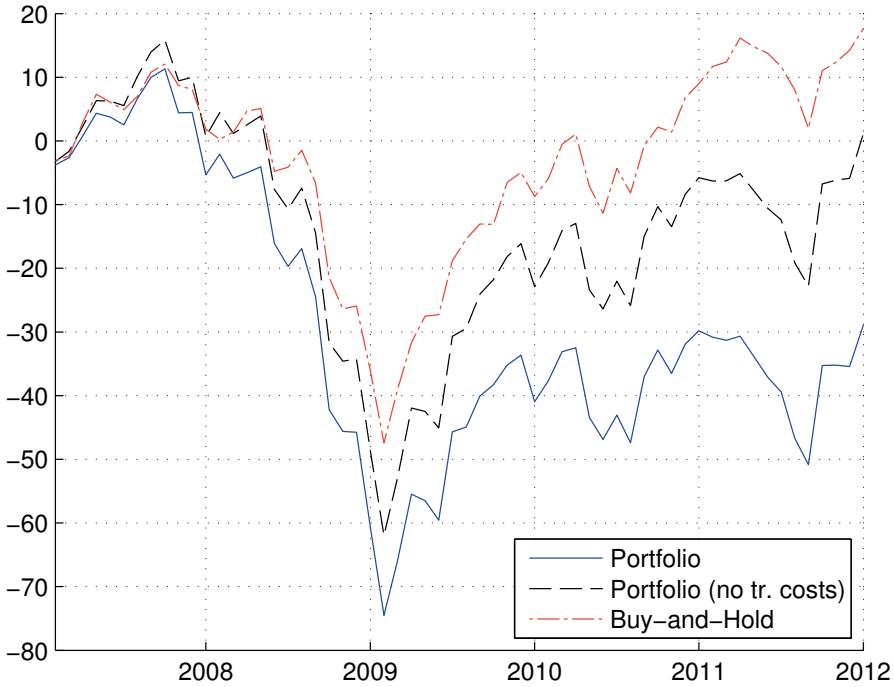


Figure 24: Plot of the returns of the Blogspot long portfolio (adjusted for/without transaction costs) vs. the buy-and-hold portfolio. The returns are cumulative log returns in %.

Table 29 also shows the inferiority of the simulated long portfolio regarding all measures.

Table 29: Excess returns and performance measures (Blogspot long portfolio).

	Without transaction costs			With transaction costs		
	Portfolio	**BaH**	**Excess**	**Portfolio**	**BaH**	**Excess**
Annual return (%)						
2007	9.98	8.34	1.64	4.47	8.09	–3.62
2008	–44.23	–34.02	–10.21	–50.24	–34.02	–16.22
2009	18.12	21.01	–2.88	12.11	21.01	–8.89
2010	7.76	11.74	–3.99	1.75	11.74	–10.00
2011	2.49	7.41	–4.92	–3.52	7.41	–10.93
2012	7.04	3.68	3.37	6.54	3.43	3.12
Total	1.16	18.16	–17.00	–28.88	17.66	–46.54
Median/mean return (%)						
Monthly median	0.57	0.89	–0.37	0.07	0.89	–0.82
Monthly mean	0.02	0.30	–0.28	–0.48	0.29	–0.78
Performance measures						
Ann. ret. std. (%)	22.36	19.08		22.85	19.06	
Sharpe ratio	–0.01	0.04		–0.09	0.04	

NOTES: see notes below Table 20.

Consequently, there is no evidence rejecting the null hypothesis of the (mean or) median of the paired differences of the simulated portfolio's log returns and the buy-and-hold portfolio's log returns (i.e., excess log returns) being zero or negative. Table 30 shows results of the nonparametric statistical hypothesis test because the normality assumption of the *t*-test was not met.

Table 30: Test results for positive excess returns (Blogspot long portfolio).

Test	Null hypothesis	Without transaction costs			With transaction costs		
		Statistic	p	Reject?	Statistic	p	Reject?
Verification of the normality assumption of the *t*-test							
AD-test	Normal distr. of excess returns	0.5200	0.1814	No	0.5282	0.1730	No
JB-test		4.1263	0.0715	Yes	4.0644	0.0732	Yes
Nonparametric test							
Wilcox. SRT	Median(exc. ret.)≤0	–1.174	0.8798	No	–2.602	0.9954	No

NOTES: see notes below Table 21.

Summarizing, there is no evidence rejecting the null hypothesis of zero or negative alpha or for rejecting the null hypothesis of the (mean or) median of monthly excess log returns being zero or negative for the long portfolio simulation on the Blogspot dataset either ignoring or adjusting for transaction costs. Therefore, H2.1 and H2.2 are not accepted for the Blogspot dataset.

4.3.3 Hypothesis H3: Effects on a Short Portfolio

This section reports results of the short-portfolio simulation and tests of hypotheses H3.1 and H3.2 on (1) the Seekingalpha dataset, and (2) the Blogspot dataset. The short portfolio held in each period the five lowest ranked stocks (according to the baseline configuration) using the level of the investor sentiment index in the previous period.

4.3.3.1 Seekingalpha

Ignoring (accounting for) transaction costs, the estimated alpha of the short portfolio simulation on the Seekingalpha dataset indicates positive (negative) mean monthly abnormal returns of 0.45% (–0.05%). However, the lower bound of the 90% confidence interval of the alpha estimate is negative. Thus, the null hypothesis of zero or negative alpha cannot be rejected at the 10% significance level. Thus, the evidence does not support H3.1 for the Seekingalpha dataset, irrespective of transaction costs.

The goodness of fit of the regression is high as suggested by the adjusted R^2. In fact, the goodness of fit is substantially higher than the one for the long-short portfolio regression. See Table 31 for details.

Table 31: H3.1 *t*-test results for Seekingalpha regarding the short portfolio.

Coefficients	Null hypothesis	90% CI bound	Estimate	Statistic	p	Reject null?
With transaction costs						
a	$\alpha \leq 0$	–0.0065	–0.0005	–0.1053	0.5417	No
Without transaction costs						
a	$\alpha \leq 0$	–0.0015	0.0045	0.9798	0.1657	No
b_1			–1.1952	–7.6167		
b_2			0.2931	1.0758		
b_3			–1.1289	–2.8317		
b_4			–0.0835	–0.9060		
Goodness of fit						
R^2			0.8200			
Adj. R^2			0.8069			
F-test	All $ß_i$ coefficients are zero			62.6543	<0.0001***	Yes

NOTES: see notes below Table 19.

Figure 25 shows the cumulated portfolio log returns of the simulated short portfolio exceeding the buy-and-hold portfolio's cumulated log returns.

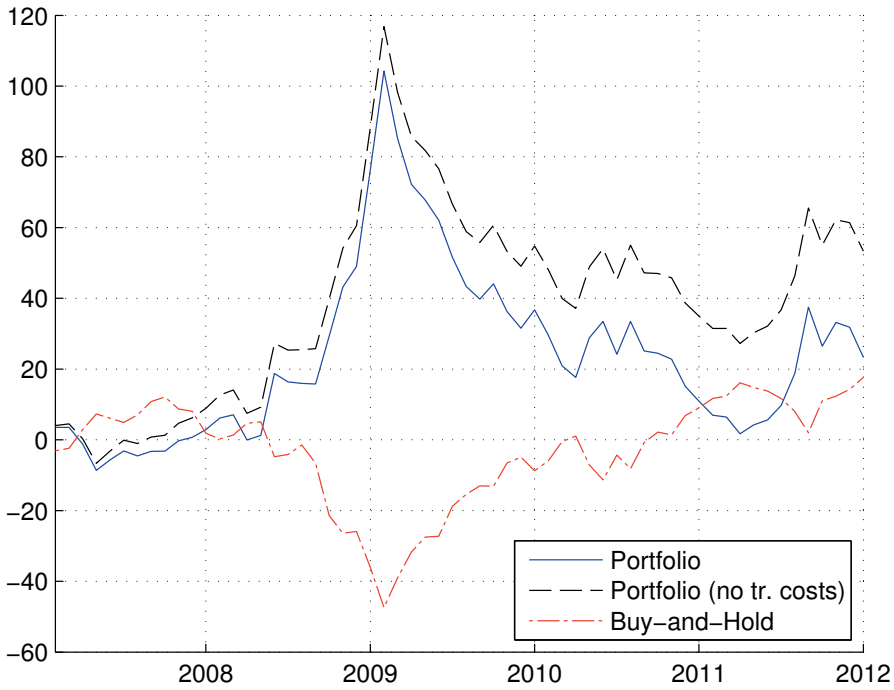


Figure 25: Plot of the returns of the Seekingalpha short portfolio (adjusted for/without transaction costs) vs. the buy-and-hold portfolio. The returns are cumulative log returns in %.

At the end of 2008 (and in the second half of 2011), the short portfolio exhibits a distinct peak (see Figure 25). That is, shorting stocks during generally negative market environments (with most stock prices falling) resulted in substantial gains. However, holding short positions during a generally positive market (with most stock prices increasing) in 2009 resulted in severe drawdowns, mostly reversing the previous gains. When accounting for transaction costs in the simulated portfolio, the distinct peaks remain. However, the

cumulated log returns fall partially below the buy-and-hold portfolio ones' in 2011. Table 32 provides further insights.

Table 32: Excess returns and performance measures (Seekingalpha short portfolio).

	Without transaction costs			With transaction costs		
	Portfolio	**BaH**	**Excess**	**Portfolio**	**BaH**	**Excess**
Annual return (%)						
2007	6.23	8.34	–2.12	0.72	8.09	–7.37
2008	54.30	–34.02	88.33	48.30	–34.02	82.32
2009	–11.47	21.01	–32.48	–17.48	21.01	–38.49
2010	–10.38	11.74	–22.13	–16.39	11.74	–28.13
2011	22.69	7.41	15.29	16.69	7.41	9.28
2012	–8.03	3.68	–11.71	–8.53	3.43	–11.96
Total	53.34	18.16	35.18	23.30	17.66	5.64
Median/mean return (%)						
Monthly median	0.22	0.89	–0.72	–0.28	0.89	–1.22
Monthly mean	0.89	0.30	0.59	0.39	0.29	0.09
Performance measures						
Ann. ret. std. (%)	25.82	19.08		25.17	19.06	
Sharpe ratio	0.09	0.04		0.03	0.04	

NOTES: see notes below Table 20.

Table 32 shows the mean monthly excess return to be positive. However, the *median* monthly excess return is negative. This divergence suggests that a few periods with large positive returns exist – which can be observed on the plot of Figure 25 and also by the >80% excess return in 2008. Combined with the long portfolio (see Section 4.3.2.1), the large peak in the cumulated return time series leads to more than negating the drawdowns of the long portfolio in 2008 (see Section 4.3.1.1). However, during 2009, the short portfolio's drawdowns also negate the long portfolio's gains (see Section 4.3.1.1). Considering the Sharpe ratio, it is in favor of the simulated short portfolio compared to the buy-and-hold portfolio – when ignoring transaction costs.

Finally, Table 33 shows Wilcoxon's signed rank test not rejecting the null hypothesis of zero or negative median excess returns irrespective of transaction costs. There is also no statistical evidence for positive mean excess returns because the normality assumption of the *t*-test was not met.

Table 33: Test results for positive excess returns (Seekingalpha short portfolio).

Test	**Null hypothesis**	**Without transaction costs**			**With transaction costs**		
		Statistic	**p**	**Reject?**	**Statistic**	**p**	**Reject?**
Verification of the normality assumption of the *t*-test							
AD-test	Normal distr. of excess returns	0.8580	0.0255	Yes	0.8550	0.0260	Yes
JB-test		8.4484	0.0214	Yes	8.4360	0.0197	Yes
Nonparametric test							
Wilcox. SRT	Median(exc. ret.)≤0	–0.328	0.6284	No	–0.600	0.7257	No

NOTES: see notes below Table 21.

Summarizing, the alpha-based evidence does not support hypothesis H3.1 with respect to the Seekingalpha dataset. Furthermore, the benchmark-portfolio-based evidence does not

support H3.2 (either ignoring or accounting for transaction costs). The investment value of the investor sentiment index values ranked lowest, which drive the selection of stocks in the short portfolio, can be observed during large stock market crises. In combination with the long portfolio, the short portfolio can prevent large drawdowns, decrease risk as measured by the annualized standard deviation of the portfolio's returns from 25.82% to 6.89% and increase the (monthly) Sharpe ratio from 0.09 to 0.34 (see Table 20, when ignoring transaction costs).

4.3.3.2 Blogspot

The estimated alpha of the portfolio simulation on the Blogspot dataset indicates positive mean monthly abnormal returns of 0.29% only when ignoring transaction costs. However, the 90% confidence interval lower bound is negative in all cases and the *p*-values are statistically not significant. See Table 34 for details. That is, in all cases, the null hypothesis of a negative or zero alpha is *not* rejected. Thus, the evidence does not support H3.1 for the Blogspot dataset. The goodness of fit of the OLS regression is high.

Table 34: H3.1 *t*-test results for Blogspot regarding the short portfolio.

Coefficients	Null hypothesis	90% CI bound	Estimate	Statistic	p	Reject null?
With transaction costs						
a	$\alpha \leq 0$	–0.0068	–0.0021	–0.5864	0.7200	No
Without transaction costs						
a	$\alpha \leq 0$	–0.0018	0.0029	0.8071	0.2116	No
b_1			–0.8695	–9.2618		
b_2			–0.0052	–0.0346		
b_3			–0.2664	–1.5128		
b_4			0.0514	0.6068		
Goodness of fit						
R^2			0.7567			
Adj. R^2			0.7390			
F-test	All $ß_i$ coefficients are zero			42.7680	<0.0001***	Yes

NOTES: see notes below Table 19.

The time series plot in Figure 26 of the simulated portfolio's log returns looks similar to the one for the Seekingalpha dataset shown in the previous subsection. However, the Blogspot-based plot has lower peaks in 2008 and in 2011. When ignoring transaction costs, the simulated portfolio's cumulative log return curve is above the one of the buy-and-hold portfolio in most periods of 2008, 2009, and 2010.

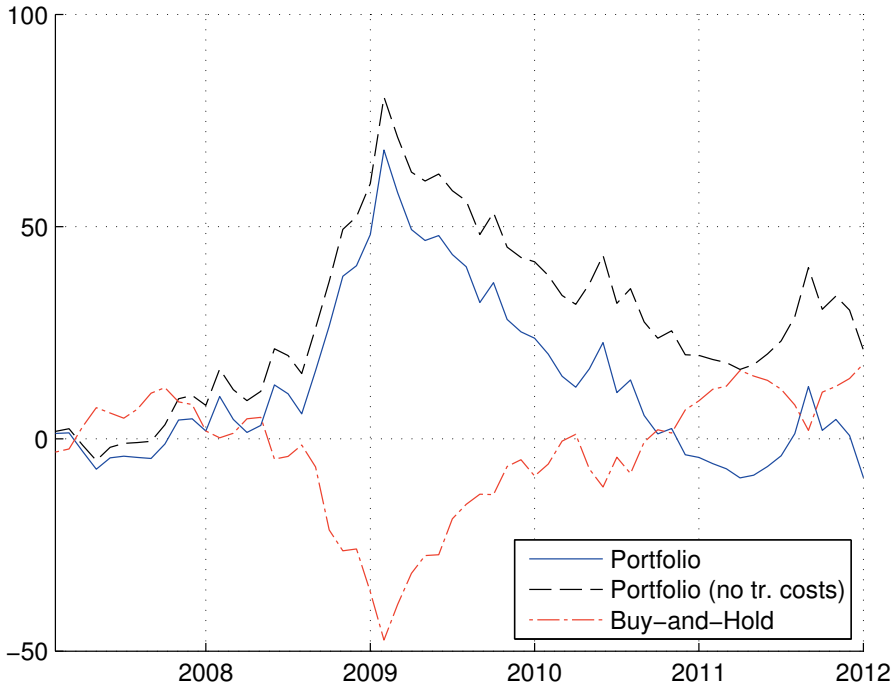


Figure 26: Plot of the returns of the Blogspot short portfolio (adjusted for/without transaction costs) vs. the buy-and-hold portfolio. The returns are cumulative log returns in %.

The monthly mean of excess returns is positive, when ignoring transaction costs (see Table 35).

Table 35: Excess returns and performance measures (Blogspot short portfolio).

	Without transaction costs			With transaction costs		
	Portfolio	**BaH**	**Excess**	**Portfolio**	**BaH**	**Excess**
Annual return (%)						
2007	10.24	8.34	1.89	4.73	8.09	–3.36
2008	42.07	–34.02	76.09	36.06	–34.02	70.08
2009	–9.57	21.01	–30.58	–15.58	21.01	–36.58
2010	–22.93	11.74	–34.67	–28.93	11.74	–40.68
2011	10.53	7.41	3.12	4.53	7.41	–2.88
2012	–9.39	3.68	–13.06	–9.89	3.43	–13.31
Total	20.95	18.16	2.79	–9.08	17.66	–26.74
Median/mean return (%)						
Monthly median	0.05	0.89	–0.74	–0.46	0.89	–1.24
Monthly mean	0.35	0.30	0.05	–0.15	0.29	–0.45
Performance measures						
Ann. ret. std. (%)	22.88	19.08		22.39	19.06	
Sharpe ratio	0.04	0.04		–0.04	0.04	

NOTES: see notes below Table 20.

Table 35 shows that the (monthly) Sharpe ratios of the simulated short portfolio and the buy-and-hold portfolio are identical (when ignoring transaction costs). The reason is that the plot in Figure 26 of the simulated short portfolio (ignoring transaction costs) looks roughly like the plot of the buy-and-hold portfolio flipped upside-down with about the same amount of overall return vs. volatility. That is, the short portfolio could have been roughly replicated

by shorting the buy-and-hold portfolio without using any investor sentiment index information generated from the Blogspot dataset.

Because the median monthly excess returns are negative, the null hypothesis of the median of the excess log returns being zero or negative cannot be rejected (see Table 36).

Table 36: Test results for positive excess returns (Blogspot short portfolio).

Test	Null hypothesis	Without transaction costs			With transaction costs		
		Statistic	p	Reject?	Statistic	p	Reject?
Verification of the normality assumption of the *t*-test							
AD-test	Normal distr. of excess returns	0.4850	0.2216	No	0.4818	0.2257	No
JB-test		3.5835	0.0887	Yes	3.5841	0.0892	Yes
Nonparametric test							
Wilcox. SRT	Median(exc. ret.)≤0	–0.460	0.6773	No	–0.769	0.7791	No

NOTES: see notes below Table 21.

Summarizing, there is no evidence rejecting the null hypothesis of zero or negative alpha or for rejecting the null hypothesis of the median of monthly excess returns being zero or negative for the short portfolio simulation on the Blogspot dataset either ignoring or adjusting for transaction costs. Therefore, H3.1 and H3.2 are not accepted for the Blogspot dataset. Still, when combining the short portfolio with the long portfolio, the (monthly) Sharpe ratio increases from 0.04 to 0.1 (see Table 23), when ignoring transaction costs.

4.3.4 Hypothesis H4: Effects in Relation to Portfolio Sizes

This section reports results of the long-short portfolio simulation in relationship to the number of stocks selected in the long and short sub-portfolios. Regarding these results hypotheses H4.1 and H4.2 were tested on (1) the Seekingalpha dataset, and (2) the Blogspot dataset. Before considering hypotheses H4.1 and H4.2, hypotheses H1.1 and H1.2 were tested to check for the existence of portfolio level abnormal returns (on the mean or median) depending on the number of stocks selected in the long and short sub-portfolios. These tests allow checking the robustness of the H1.1 and H1.2 hypotheses.

4.3.4.1 Seekingalpha

When ignoring transaction costs, the estimates of alpha and the mean monthly excess returns (versus the buy-and-hold portfolio) are positive for all $N \in \{1,2,\ldots 14\}$ where N represents the number of stocks selected in each period in a long and short (sub-)portfolio respectively. This simulation result suggests the existence of positive (mean monthly) abnormal returns not to depend on the parameter N. That is, a positively signed predictive relationship between the level of the monthly investor sentiment index of the Seekingalpha dataset and abnormal returns (measured on the portfolio level by an alpha estimate or mean excess returns) seems to exist on the mean. Regarding statistical evidence, the null hypothesis of zero or negative alpha is rejected at the 1% significance level for N=4,5,6,7,11 at the 5% significance level for N=3,4,…,14, and at the 10% significance level for N=2,3,…,14. That is, for N=2,3,…,14 the statistical evidence suggests rejecting the null hypothesis of zero or negative alpha,

supporting the alternative hypothesis H1.1. When accounting for transaction costs, positive estimates of alpha exist only for N=1,2,…,7 with no statistical evidence rejecting the null hypothesis. Table 37 shows details.

Table 37: H1.1 *t*-test results for the Seekingalpha long-short portfolio with N stocks selected in the long and short sub-portfolios.

N	α estimate	90% CI bound	*t*-statistic	p	Reject α≤0?	RF-test
With transaction costs						
1	0.0017	–0.0061	0.2796	0.3904	No	Yes
2	0.0015	–0.0044	0.3268	0.3725	No	No
3	0.0025	–0.0026	0.6313	0.2652	No	Yes
4	0.0038	–0.0003	1.2071	0.1163	No	Yes
5	0.0033	–0.0001	1.2457	0.1091	No	Yes
6	0.0008	–0.0022	0.3340	0.3698	No	Yes
7	0.0011	–0.0018	0.4938	0.3117	No	Yes
8	–0.0001	–0.0028	–0.0550	0.5218	No	Yes
9	–0.0008	–0.0032	–0.4256	0.6640	No	Yes
10	–0.0009	–0.0032	–0.4626	0.6773	No	Yes
11	–0.0009	–0.0030	–0.5580	0.7104	No	Yes
12	–0.0011	–0.0033	–0.6964	0.7555	No	Yes
13	–0.0019	–0.0039	–1.2911	0.8990	No	Yes
14	–0.0018	–0.0037	–1.2305	0.8881	No	Yes
Without transaction costs						
1	0.0067	–0.0011	1.1087	0.1362	No	Yes
2	0.0065	0.0006	1.4359	0.0783*	Yes	No
3	0.0075	0.0024	1.8991	0.0314**	Yes	Yes
4	0.0088	0.0047	2.8093	0.0034***	Yes	Yes
5	0.0083	0.0049	3.1541	0.0013***	Yes	Yes
6	0.0058	0.0028	2.5267	0.0072***	Yes	Yes
7	0.0061	0.0032	2.7363	0.0042***	Yes	Yes
8	0.0049	0.0022	2.3965	0.0100**	Yes	Yes
9	0.0042	0.0018	2.2301	0.0149**	Yes	Yes
10	0.0042	0.0018	2.2506	0.0142**	Yes	Yes
11	0.0041	0.0020	2.5041	0.0076***	Yes	Yes
12	0.0039	0.0017	2.3463	0.0113**	Yes	Yes
13	0.0031	0.0011	2.0300	0.0236**	Yes	Yes
14	0.0032	0.0013	2.2032	0.0159**	Yes	Yes

NOTES: "α estimate" is the OLS-estimate of the coefficient α in Carhart's (1997) model on the portfolio returns (in excess of the risk free rate). The overall portfolio consists in each period of the simulation of N stocks selected in both, a long and a short sub-portfolio. In the simulation with transaction costs, 50 bps were assumed per period on the portfolio level. The p-value is regarding a test with respect to (wrt.) the null hypothesis of $\alpha \leq 0$. "Reject $\alpha \leq 0$" refers to whether the null hypothesis is rejected given the evidence at the 10% significance level. Significance levels are denoted as follows: *** = 1%, ** = 5%, * = 10%. For the hypothesis test wrt. α, a right-tailed t-test with Newey & West (1987) standard errors was conducted. Based on the estimate of the standard error, the t-statistic, the p-value, and the CI (i.e., confidence interval lower bound of the coefficient estimate) were calculated. The upper bound CI is $+\infty$. The null hypothesis of the F-test of all beta coefficients of Carhart's (1997) model being jointly zero was rejected at the 1% level, indicated in the column "RF-test". The number of observations of the dependent and independent variables has been 60.

Table 38 presents performance measurements of the portfolio simulations with different N in terms of monthly mean and median of portfolio returns in excess of the buy-and-hold portfolio's returns and the (monthly) Sharpe ratio.

Table 38: H1.2 test results for the Seekingalpha long-short portfolio with *N* stocks selected in the long and short sub-portfolios.

N	Sharpe ratio	Monthly mean excess return	Monthly median excess return	Wilcoxon's SRT z-score	p	Reject null?
With transaction costs						
1	0.0335	–0.0229	–1.8463	–0.7178	0.7635	No
2	0.0622	0.0490	–0.7571	–0.5926	0.7233	No
3	0.1306	0.2966	–0.9741	–0.6147	0.7306	No
4	0.1861	0.4825	–1.1385	–0.2687	0.6059	No
5	0.1782	0.3591	–0.9852	–0.3718	0.6450	No
6	0.0988	0.0740	–0.6699	–0.6883	0.7544	No
7	0.0983	0.0544	–1.1194	–0.6221	0.7330	No
8	0.0435	–0.0973	–0.7255	–0.7178	0.7635	No
9	0.0133	–0.1644	–0.4664	–0.8135	0.7920	No
10	–0.0112	–0.2173	–0.3695	–0.7030	0.7590	No
11	–0.0287	–0.2494	–0.7699	–0.8208	0.7941	No
12	–0.0471	–0.2817	–0.7851	–0.8429	0.8004	No
13	–0.0981	–0.3683	–1.0771	–0.9460	0.8279	No
14	–0.1064	–0.3715	–1.0463	–0.9607	0.8316	No
Without transaction costs						
1	0.1320	0.4694	–1.3457	–0.2319	0.5917	No
2	0.1910	0.5412	–0.2565	–0.0773	0.5308	No
3	0.2640	0.7888	–0.4735	0.1288	0.4487	No
4	0.3241	0.9747	–0.6379	0.3276	0.3716	No
5	0.3397	0.8514	–0.4846	0.2098	0.4169	No
6	0.2839	0.5662	–0.1693	0.0037	0.4985	No
7	0.2971	0.5466	–0.6188	0.0110	0.4956	No
8	0.2707	0.3950	–0.2248	0.0920	0.4633	No
9	0.2442	0.3279	0.0342	–0.2393	0.5945	No
10	0.2207	0.2750	0.1311	–0.1656	0.5658	No
11	0.2270	0.2429	–0.2693	–0.1877	0.5745	No
12	0.2189	0.2105	–0.2844	–0.2761	0.6087	No
13	0.1823	0.1240	–0.5765	–0.3423	0.6339	No
14	0.1921	0.1208	–0.5457	–0.2761	0.6087	No

NOTES: The performance of the simulated overall portfolio (using investor sentiment indexes) with selecting *N* stocks in both, a long and a short sub-portfolio, is reported in terms of the *monthly* Sharpe (1966) ratio, monthly mean and monthly median excess log portfolio returns in percent. Note that the Sharpe ratio was calculated using risk free returns described in Appendix A.5 and that it was shown to be adequate for relative rankings also in case of non-normally distributed portfolio returns (Eling & Schuhmacher, 2007). Excess log returns were calculated as paired difference time series of the simulated overall portfolio's log returns vs. the buy-and-hold portfolio's log returns. Wilcoxon's (1945) signed rank test ("SRT") was used to non-parametrically test the null hypothesis of zero or negative median excess log returns of the simulated overall portfolio. Regarding the SRT, the *z*-score and the *p*-value are reported. The null hypothesis of the SRT was rejected (see the column "Reject null?") at a *p*-value indicating a 10% significance level (or better). Accounting for transaction costs meant deducting 50 bps per period for the simulated overall portfolio and 25 bps in the first and 25 bps in the last period for the buy-and-hold portfolio.

According to Table 38, the Sharpe ratio is greater than the 0.04 Sharpe ratio of the buy-and-hold portfolio (see Section 4.3.1) for all values of *N* when ignoring transaction costs. This outperformance indicates (theoretical) positive investment value irrespective of the setting of the parameter *N*. Also the monthly mean excess return is positive for all values of *N*. This supports the alpha-related findings (based on the alpha estimates) of the existence of positive abnormal returns on the monthly mean for all values of *N*. Regarding the *t*-test of the

null hypothesis of zero or negative mean excess returns, detailed results are not provided because the required normality of excess returns was rejected for N=1,2,…,10 by the AD-test or the JB-test at the 10% significance level. For N=11,12,13,14 the null hypothesis of zero or negative mean excess returns was not rejected at the 10% significance level (either accounting for or ignoring transaction costs). Furthermore, Wilcoxon's signed rank test (SRT) was used for testing the null hypothesis of monthly median excess returns being zero or negative. This null hypothesis cannot be rejected given the evidence (see Table 38). The null hypothesis can also be not rejected when accounting for transaction costs. Note that when accounting for transaction costs, positive monthly mean excess returns are observed for N=2,3,…,7 only.

A plot of alpha estimates and monthly mean excess returns (presented in Table 37 and Table 38) helps in evaluating hypotheses H4.1 and H4.2 of the alpha or monthly mean excess returns to increase with decreasing N. When considering Figure 27, estimates of alpha and the monthly mean excess return roughly decrease for N ranging from 4,3,2 to 1 and mostly increase for N ranging from 14,13,12,… to 4. This observation indicates some support for H4.1 (related to alpha) and H4.2 (related to monthly mean excess returns).

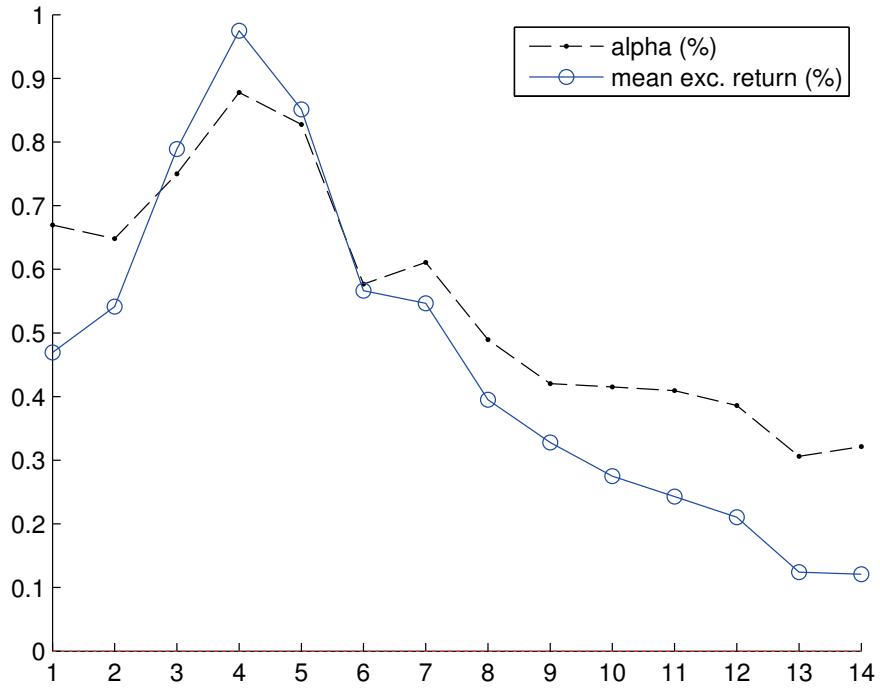


Figure 27: Plot of alpha and mean excess return for the Seekingalpha long-short portfolio, ignoring transaction costs. The outcome depends on the number of stocks selected (on the horizontal axis) in each period in the long and short portfolios. The monthly alpha (in %) was estimated by Carhart's (1997) model and returns are monthly log returns (in %).

The largest estimate of alpha (monthly mean excess return) of 0.88% (0.975%) is observed for N=4. The increase for small but increasing N was not hypothesized or anticipated. However, it seems reasonable because a certain minimum number of stocks might have to be selected to include the ones generating the highest alpha. This minimum

requirement might be due to the stock's investor sentiment index of the Seekingalpha dataset not relating perfectly to highest level of next period alpha.

Next, Figure 28 accounts for transaction costs. It shows the same relative patterns and in absolute terms positive estimates of alpha (monthly mean excess returns) to occur for N=1,2,…,7 (N=2,3,…,7).

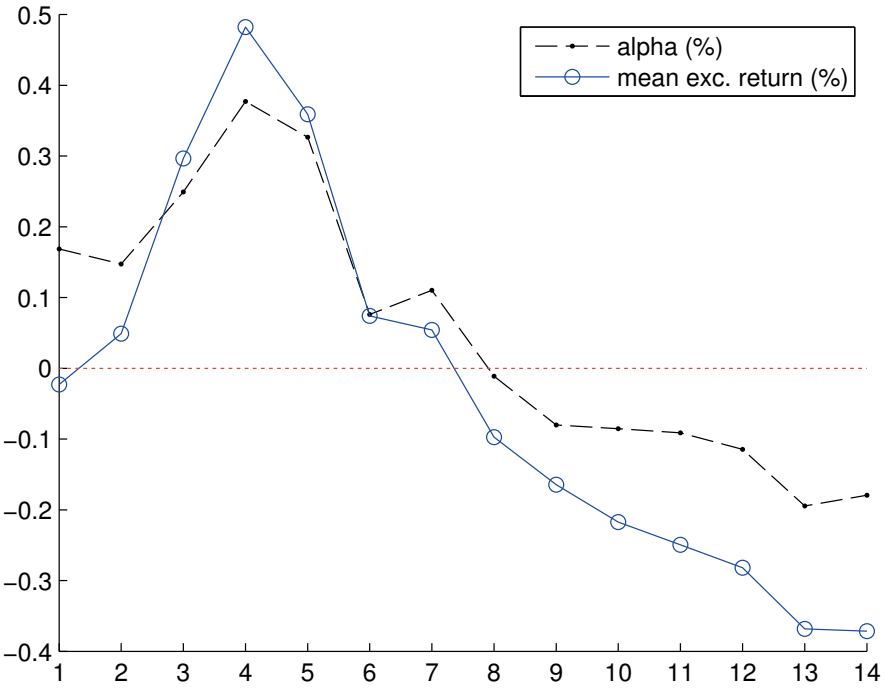


Figure 28: Plot of alpha and mean excess return for the Seekingalpha long-short portfolio, adjusting for transaction costs. The outcome depends on the number of stocks selected (on the horizontal axis) in each period in the long and short portfolios. The monthly alpha (in %) was estimated by Carhart's (1997) model and returns are monthly log returns (in %).

Regarding checking for potential statistical evidence rejecting any null hypothesis $\alpha_N - \alpha_{N+1} \leq 0$ (which is equivalent to $\alpha_N \leq \alpha_{N+1}$) of (the alternative hypothesis) H4.1 with $N \in \{1,2,\ldots,13\}$ referring to the number of stocks selected in each period in the long and short sub-portfolios, Table 39 presents results. The sign of the estimate of the difference of alphas of portfolio simulations with subsequent values of N is positive in 9 out of 13 cases, indicating some support for H4.1. However, there is no statistical evidence rejecting the null hypothesis of H4.1 because all 90% CI lower bounds of alpha difference estimates are negative. Note that the results presented in Table 39 are with respect to the portfolio simulation that ignores transaction costs and the one that accounts for transaction costs. The reason is that alpha *differences* do not change when accounting for transaction costs by the same relative amount per period on the portfolio level.

Table 39: H4.1 t-test results for the Seekingalpha long-short portfolio with *N* vs. *N*+1 stocks selected in the long and short sub-portfolios. The tests are regarding the null hypothesis of H4.1 with respect to the difference of Carhart alphas.

N	$\alpha_N - \alpha_{N+1}$ estimate	90% CI bound	*t*-statistic	p	Reject $\alpha_N \leq \alpha_{N+1}$?	RF-test
1	0.0002	–0.0095	0.0283	0.4887	No	No
2	–0.0010	–0.0088	–0.1695	0.5671	No	Yes
3	–0.0013	–0.0078	–0.2536	0.5999	No	Yes
4	0.0005	–0.0048	0.1236	0.4509	No	Yes
5	0.0025	–0.0020	0.7200	0.2365	No	Yes
6	–0.0003	–0.0045	–0.1064	0.5423	No	Yes
7	0.0012	–0.0027	0.4008	0.3447	No	Yes
8	0.0007	–0.0029	0.2485	0.4021	No	Yes
9	0.0001	–0.0033	0.0195	0.4922	No	Yes
10	0.0001	–0.0031	0.0239	0.4905	No	Yes
11	0.0002	–0.0028	0.1009	0.4599	No	Yes
12	0.0008	–0.0021	0.3590	0.3601	No	Yes
13	–0.0002	–0.0029	–0.0727	0.5289	No	Yes

NOTES: "$\alpha_N - \alpha_{N+1}$ estimate" is the OLS-estimate of the coefficient α_{A-B} (with *A* being a proxy for *N* and *B* being a proxy for *N+1*) in the model in Definition (4.15). The estimate of the coefficient α_{A-B} estimates the difference in alpha of Carhart's (1997) model for the overall portfolio simulation outputs using *N* vs. *N*+1 stocks selected in both, a long and a short sub-portfolio. The result is irrespective of transaction costs. The *p*-value is regarding a test with respect to (wrt.) the null hypothesis of $\alpha_{A-B} \leq 0$. "Reject $\alpha_N \leq \alpha_{N+1}$?" refers to whether the null hypothesis is rejected given the evidence at the 10% significance level. For the hypothesis test, a right-tailed *t*-test based on Newey & West (1987) standard errors was conducted. Based on the estimate of the standard error, the *t*-statistic, the *p*-value, and the CI (i.e., confidence interval lower bound of the α_{A-B} estimate) were calculated. The upper bound CI is $+\infty$. The null hypothesis of the *F*-test of all beta coefficients of Carhart's (1997) model being jointly zero was rejected at the 1% significance level, indicated in the column "RF-test". The number of observations of the dependent and independent variables in Carhart's model has been 120.

Related to testing H4.2, 10 of 13 means of the pairwise differences of excess returns are positive (for *N*=4,5,…,13). Details on hypothesis test results regarding the null hypothesis of H4.2 relating to the mean of the pairwise differences of excess returns are not reported because the normality assumption required for the paired samples *t*-test of this null hypothesis was violated for 9 of 13 values of *N* (for *N*=2,3,4,6,8,9,11,12,13) as indicated by the AD-test or JB-test at the 10% significance level. Among the remaining values of *N*, for *N*=5,7 the null hypothesis of a zero or negative mean difference was rejected at the 5% significance level.

Medians of excess returns (vs. the buy-and-hold portfolio) are not plotted because the difference of medians does not equal the median of paired differences, regarding which a null hypothesis of the alternative H4.2 was formulated. Thus, Table 40 presents medians of paired differences of excess returns. The table shows 8 of 13 median differences of excess returns of two portfolios with *N* and *N+1* stocks selected to be greater than zero. The Wilcoxon signed rank test (SRT) rejects the null hypothesis of a zero or negative median difference in two cases at the 5% significance level and in four cases at the 10% significance level. Note that the results presented in Table 40 are valid when accounting for or when ignoring transaction costs.

Table 40: H4.2 test results for the Seekingalpha long-short portfolio with *N* vs. *N*+1 stocks selected in the long and short sub-portfolios.

N	Median difference	Wilcoxon's SRT z-score	p	Reject null?
1	–0.0017	–0.3938	0.6532	No
2	–0.0019	–1.0122	0.8443	No
3	0.0003	–0.0994	0.5396	No
4	–0.0005	0.2024	0.4198	No
5	0.0038	2.2563	0.0120**	Yes
6	–0.0014	–0.2393	0.5945	No
7	0.0016	1.8514	0.0321**	Yes
8	0.0011	1.5202	0.0642*	Yes
9	0.0003	0.4380	0.3307	No
10	0.0001	0.4012	0.3441	No
11	0.0001	0.9239	0.1778	No
12	0.0001	1.4171	0.0782*	Yes
13	0.0000	–0.8753	0.8093	No

NOTES: The performance *difference* of the simulated overall portfolio with *N* vs. *N*+1 stocks selected in the long and short sub-portfolios is measured in terms of the median of paired differences of excess log returns of the overall portfolios. Excess log returns were calculated as paired difference time series of a simulated overall portfolio's returns using investor sentiment indexes vs. the buy-and-hold portfolio's returns. Wilcoxon's (1945) signed rank test ("SRT") was used to non-parametrically test the null hypothesis of zero or negative median of paired differences in excess log returns of the simulated overall portfolios. Regarding the SRT, the *z*-score and the *p*-value are reported. The significance level is denoted as follows: *** = 1%, ** = 5%, * = 10%. The null hypothesis of the SRT was rejected (see the column "Reject null?") at a *p*-value indicating a 10% (or better) significance level. Results are irrespective of transaction costs.

Summarizing, some indications that support H4.1 and H4.2 (by means of several positive estimates of alpha differences and positive means of excess return differences) for the Seekingalpha dataset were found – either ignoring or accounting for transaction costs. However, H4.1 was not accepted because of lack of statistical evidence. Furthermore, the null hypothesis of the mean (median) of excess return differences being negative or zero was rejected for only two (four) values of *N* of the number of stocks selected. Thus, H4.2 was accepted only for these values of *N*. Furthermore, as a side effect, the simulation results using a different number of stocks *N* in the portfolios suggest support for H1.1 and H1.2 (relating to monthly mean excess returns) across many *N*. Whereas statistical evidence was found only for H1.1 when transaction costs were ignored, these results still indicate robustness of these hypotheses.

4.3.4.2 Blogspot

Ignoring transaction costs, the estimate of alpha is positive for all $N \in \{1,2,\ldots 14\}$. Thus, the existence of positive alpha for the Blogspot dataset also does not depend on the parameter *N*. However, statistical evidence for rejecting the null hypothesis of zero or negative alpha was found only for *N*=2,3,6 at the 10% significance level. When accounting for transaction costs, a positive estimate of alpha exists only for *N*=2 (without statistical evidence rejecting the null hypothesis). Table 41 reports details.

Table 41: H1.1 *t*-test results for the Blogspot long-short portfolio with *N* stocks selected in the long and short sub-portfolios.

N	α estimate	90% CI bound	*t*-statistic	p	Reject α≤0?	RF-test
With transaction costs						
1	–0.0028	–0.0080	–0.7001	0.7566	No	Yes
2	0.0001	–0.0041	0.0449	0.4822	No	No
3	–0.0010	–0.0046	–0.3368	0.6312	No	No
4	–0.0018	–0.0052	–0.6746	0.7486	No	No
5	–0.0029	–0.0063	–1.1142	0.8650	No	No
6	–0.0007	–0.0045	–0.2238	0.5881	No	No
7	–0.0014	–0.0051	–0.4833	0.6846	No	No
8	–0.0018	–0.0055	–0.6460	0.7395	No	Yes
9	–0.0026	–0.0061	–0.9686	0.8315	No	Yes
10	–0.0025	–0.0059	–0.9299	0.8218	No	Yes
11	–0.0031	–0.0065	–1.1863	0.8797	No	Yes
12	–0.0036	–0.0068	–1.4789	0.9276	No	Yes
13	–0.0047	–0.0076	–2.1176	0.9806	No	Yes
14	–0.0047	–0.0076	–2.0517	0.9775	No	Yes
Without transaction costs						
1	0.0022	–0.0030	0.5480	0.2930	No	Yes
2	0.0052	0.0009	1.5586	0.0624	Yes	No
3	0.0041	0.0004	1.4354	0.0784	Yes	No
4	0.0032	–0.0002	1.2143	0.1149	No	No
5	0.0021	–0.0013	0.8146	0.2094	No	No
6	0.0043	0.0005	1.4698	0.0737	Yes	No
7	0.0036	–0.0001	1.2599	0.1065	No	No
8	0.0032	–0.0005	1.1203	0.1337	No	Yes
9	0.0024	–0.0011	0.8995	0.1862	No	Yes
10	0.0025	–0.0009	0.9573	0.1713	No	Yes
11	0.0019	–0.0015	0.7248	0.2358	No	Yes
12	0.0014	–0.0018	0.5780	0.2828	No	Yes
13	0.0003	–0.0026	0.1245	0.4507	No	Yes
14	0.0003	–0.0026	0.1486	0.4412	No	Yes

NOTES: see notes below Table 37.

Table 42 provides benchmark-related performance results of the portfolio simulations with different *N*. Ignoring transaction costs, the (monthly) Sharpe ratio is greater than the 0.04 Sharpe ratio of the buy-and-hold portfolio (see Section 4.3.1) for all values of *N*. Also the monthly mean excess return is positive for *N*=1,2,…,12. To this respect, the *t*-test for the null hypothesis of zero or negative mean excess returns could not be applied because normality of excess returns was violated for all values of *N* as indicated by the AD-test or JB-test at the 10% significance level. Furthermore, the monthly median excess return is negative for all values of *N*. A negative median in combination with a positive mean indicates many periods observing negative excess returns and only a few periods observing large positive excess returns. Consequently, Wilcoxon's signed rank test (SRT) of the null hypothesis of zero or negative monthly median excess returns does not reject the null hypothesis. When accounting for transaction costs, monthly mean and median excess returns are negative for all values of *N*. This observation suggests the Blogspot investor sentiment index not to have practical investment value. See Table 42 for details.

Table 42: H1.2 test results for the Blogspot long-short portfolio with *N* stocks selected in the long and short sub-portfolios. The test refers to Wilcoxon's signed rank test of the null hypothesis of H1.2 regarding the median of excess log returns of a long-short portfolio.

N	Sharpe ratio	Monthly mean excess return	Monthly median excess return	Wilcoxon's SRT z-score	p	Reject null?
With transaction costs						
1	–0.0231	–0.3035	–0.6627	–1.2699	0.8979	No
2	0.0241	–0.1153	–1.3154	–1.0049	0.8425	No
3	–0.0072	–0.2146	–0.6534	–0.9607	0.8316	No
4	–0.0471	–0.3085	–0.5669	–1.0932	0.8628	No
5	–0.1109	–0.4504	–0.8507	–1.2331	0.8912	No
6	–0.0131	–0.2219	–0.5109	–1.0417	0.8512	No
7	–0.0567	–0.3070	–0.5983	–1.0785	0.8596	No
8	–0.0539	–0.3143	–0.9438	–1.2920	0.9018	No
9	–0.0727	–0.3614	–0.8870	–1.2993	0.9031	No
10	–0.0404	–0.2931	–0.9059	–1.3067	0.9043	No
11	–0.0644	–0.3603	–1.1837	–1.2920	0.9018	No
12	–0.0920	–0.4294	–1.5864	–1.3214	0.9068	No
13	–0.1285	–0.5392	–1.6576	–1.4171	0.9218	No
14	–0.1240	–0.5261	–1.6027	–1.4024	0.9196	No
Without transaction costs						
1	0.0818	0.1888	–0.1621	–0.5116	0.6955	No
2	0.1795	0.3770	–0.8148	–0.2613	0.6031	No
3	0.1608	0.2777	–0.1528	–0.2981	0.6172	No
4	0.1572	0.1838	–0.0663	–0.3570	0.6395	No
5	0.1048	0.0418	–0.3500	–0.5779	0.7183	No
6	0.2146	0.2703	–0.0102	–0.3644	0.6422	No
7	0.1924	0.1852	–0.0976	–0.5264	0.7007	No
8	0.1689	0.1780	–0.4431	–0.6662	0.7474	No
9	0.1435	0.1309	–0.3864	–0.7472	0.7725	No
10	0.1618	0.1992	–0.4052	–0.6883	0.7544	No
11	0.1284	0.1319	–0.6831	–0.7325	0.7681	No
12	0.1029	0.0629	–1.0858	–0.8429	0.8004	No
13	0.0574	–0.0469	–1.1570	–0.8797	0.8105	No
14	0.0624	–0.0338	–1.1021	–0.8871	0.8125	No

NOTES: see notes below Table 38.

Figure 29 shows estimates of alpha and monthly mean excess returns to peak for N=2,6,10 with setbacks in between. The largest estimate of alpha (monthly mean excess return) of 0.52% (0.38%) is observed for N=2. The multiple peaks are in contrast to the Seekingalpha simulation results, which show only one distinct peak and generally higher levels of estimated alpha and monthly mean excess returns. Due to the more than one peak in the Blogspot simulation, results depend much more on the parameter N. That is, the existence of positive abnormal returns on the monthly mean for the Blogspot dataset seems less robust. An interpretation of the plot in Figure 29 is that the possible predictive relationship of the investor sentiment index from Blogspot with respect to next period abnormal returns might be valid for specific stocks only and might be in general more random. Between N=2 and N=5, between N=6 and N=9, and between N=10 and N=13 estimates of alpha and the monthly mean excess returns decrease. However, there is no consistent pattern that would support H4.1 or H4.2.

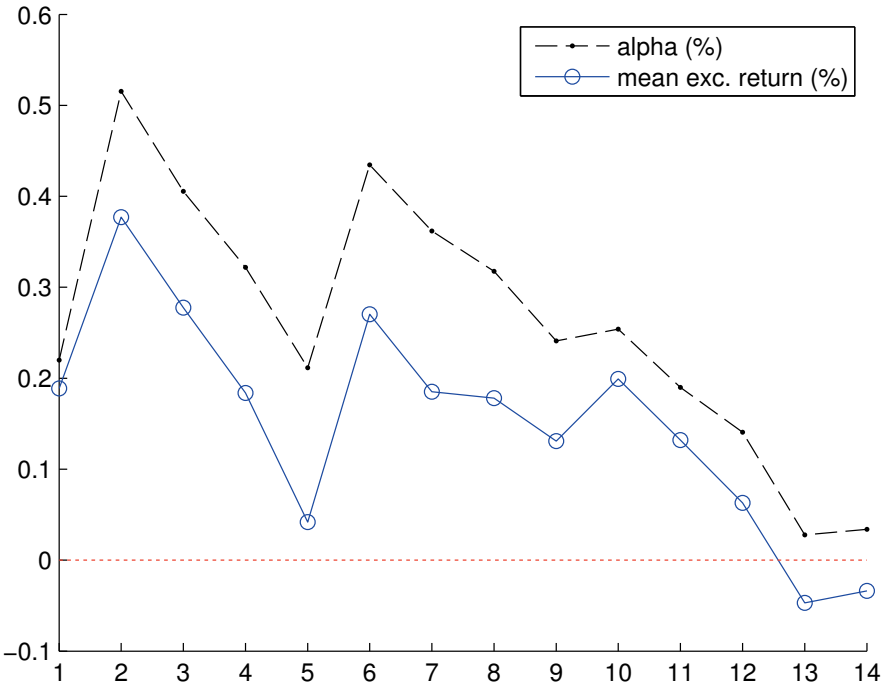


Figure 29: Plot of alpha and mean excess return for the Blogspot long-short portfolio, ignoring transaction costs. The outcome depends on the number of stocks selected (on the horizontal axis) in each period in the long and short portfolios. The monthly alpha (in %) was estimated by Carhart's (1997) model and returns are monthly log returns (in %).

When accounting for transaction costs, estimates of alpha and monthly mean excess returns turn negative for almost all values of N (see Figure 30). Table 43 presents results regarding tests of the null hypotheses $\alpha_N - \alpha_{N+1} \leq 0$ for N=1,2,…,13 of hypothesis H4.1 to check for statistical evidence (accounting for or ignoring transaction costs).

Table 43: H4.1 t-test results for the Blogspot long-short portfolio with N vs. N+1 stocks selected in the long and short sub-portfolios. The tests are regarding the null hypothesis of H4.1 with respect to the difference of Carhart alphas.

N	$\alpha_N - \alpha_{N+1}$ estimate	90% CI bound	*t*-statistic	p	Reject $\alpha_N \leq \alpha_{N+1}$?	RF-test
1	–0.0030	–0.0097	–0.5667	0.7140	No	No
2	0.0011	–0.0045	0.2532	0.4003	No	No
3	0.0008	–0.0041	0.2163	0.4146	No	No
4	0.0011	–0.0037	0.2987	0.3829	No	No
5	–0.0022	–0.0073	–0.5675	0.7142	No	No
6	0.0007	–0.0046	0.1763	0.4302	No	No
7	0.0004	–0.0048	0.1098	0.4564	No	No
8	0.0008	–0.0043	0.1963	0.4224	No	Yes
9	–0.0001	–0.0050	–0.0343	0.5136	No	Yes
10	0.0006	–0.0042	0.1722	0.4318	No	Yes
11	0.0005	–0.0041	0.1376	0.4454	No	Yes
12	0.0011	–0.0031	0.3414	0.3667	No	Yes
13	–0.0001	–0.0042	–0.0189	0.5075	No	Yes

NOTES: see notes below Table 39.

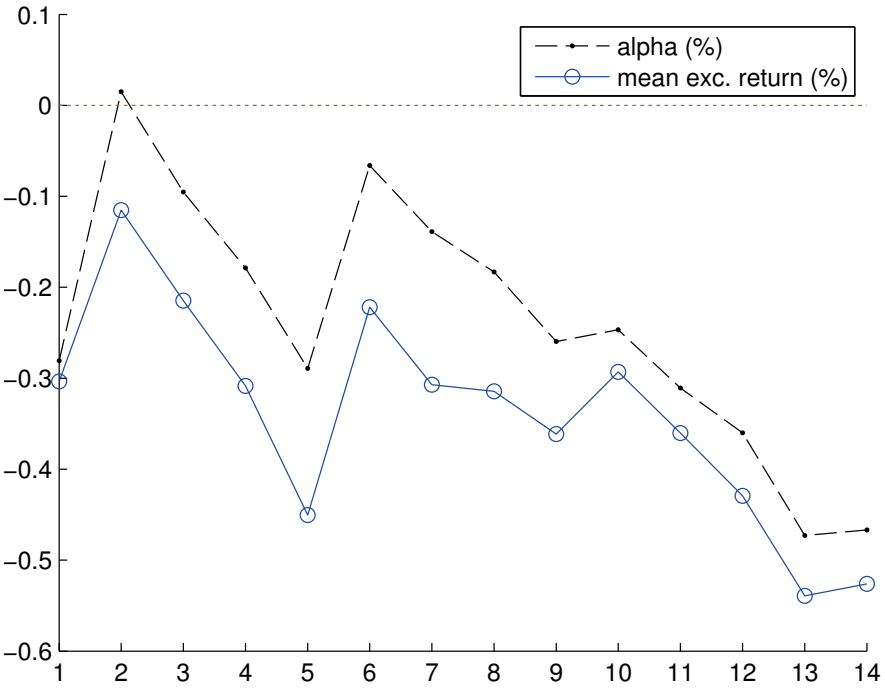


Figure 30: Plot of alpha and mean excess return for the Blogspot long-short portfolio, adjusting for transaction costs. The outcome depends on the number of stocks selected (on the horizontal axis) in each period in the long and short portfolios. The monthly alpha (in %) was estimated by Carhart's (1997) model and returns are monthly log returns (in %).

Table 43 shows the sign of the estimate of the difference of alphas of portfolio simulations with subsequent N to be positive in 9 out of 13 cases, indicating some support for H4.1. However, there is no statistical evidence rejecting the null hypothesis of H4.1.

Related to testing H4.2, 9 of 13 means of the pairwise differences of excess returns are positive (for N=2,3,4,6,7,8,10,11,12). Details of hypothesis test results regarding the null hypothesis of H4.2, relating to the mean of the pairwise differences of excess returns, are not reported because the normality assumption required for the paired samples t-test of this null hypothesis was violated. Rejection of normality was observed for 11 of 13 values of N as indicated by the AD-test or JB-test at the 10% significance level. Among the remaining values of N (i.e., N=4,6), the null hypothesis of a zero or negative mean difference was rejected only for N=4 at the 10% significance level.

Finally, Table 44 shows 10 of 13 medians of paired differences of excess returns of two portfolios with N and N+1 stocks selected to be greater than zero. The high number of positive median differences indicates support for H4.2 (related to the median difference). Furthermore, Wilcoxon's signed rank test (SRT) rejects the null hypothesis of a zero or negative median difference in one case (at the 1% significance level) – and in three cases at the 10% significance level. However, there is no broad rejection of the null hypotheses of H4.2. Note that the table's results are valid when accounting for or when ignoring transaction costs.

Table 44: H4.2 test results for the Blogspot long-short portfolio with *N* vs. *N*+1 stocks selected in the long and short sub-portfolios. The test refers to Wilcoxon's signed rank test of the null hypothesis of H4.2 regarding the median of paired differences of excess log returns of the combined long-short portfolios.

N	Median difference	Wilcoxon's SRT z-score	p	Reject null?
1	0.0014	0.1067	0.4575	No
2	0.0002	0.5558	0.2892	No
3	0.0026	0.8208	0.2059	No
4	0.0017	1.4024	0.0804*	Yes
5	–0.0024	–1.9545	0.9747	No
6	–0.0002	0.7030	0.2410	No
7	0.0007	1.1006	0.1355	No
8	0.0005	0.9312	0.1759	No
9	0.0004	0.4969	0.3096	No
10	0.0003	1.0122	0.1557	No
11	0.0006	1.5717	0.0580*	Yes
12	0.0013	2.6097	0.0045***	Yes
13	0.0000	–0.9021	0.8165	No

NOTES: see notes below Table 40.

Summarizing, H4.1 is not accepted with respect to the Blogspot dataset due to a lack of evidence. Related to H4.2, evidence regarding one (three) values of *N* was found, rejecting the null hypothesis of H4.2 of a negative or zero mean (median) of differences of excess returns. Generally, the simulation results in terms of estimates of alpha and monthly mean excess returns are less robust and depend more on the number of stocks selected in comparison to the Seekingalpha dataset. At least, when ignoring transaction costs, positive estimates of alpha and monthly mean excess returns (as well as Sharpe ratios above the one of the buy-and-hold portfolio) for all $N \leq 12$ indicate some theoretical investment value of the Blogspot investor sentiment indexes and consistency with hypotheses H1.1 and H1.2 (relating to monthly mean excess returns). However, statistical evidence for H1.1 was found only for four values of *N*, when ignoring transaction costs.

4.3.5 Hypothesis H5: Effects in Relation to the Datasets

This section compares results of the long-short portfolio simulation on the Seekingalpha dataset vs. the Blogspot dataset to test hypotheses H5.1 and H5.2. Each simulation was performed over different numbers of stocks selected in the long and short sub-portfolios. Results are separated in (1) results with accounting for transaction costs, and (2) results without transaction costs. The numerical simulation results required for the comparisons in this section have been mostly presented in the previous section. Thus, some new plots that allow for graphical comparisons are presented.

4.3.5.1 With Transaction Costs

Most plots and tabulated results presented in this section are in favor of the portfolio simulation using the monthly investor sentiment index of the Seekingalpha dataset and indicate support for hypothesis H5.1 of higher alpha and H5.2 of a positive mean or median

of differences in excess returns (vs. the buy-and-hold portfolio's returns) of the simulation on the Seekingalpha vs. the Blogspot dataset for a given number N of stocks selected.

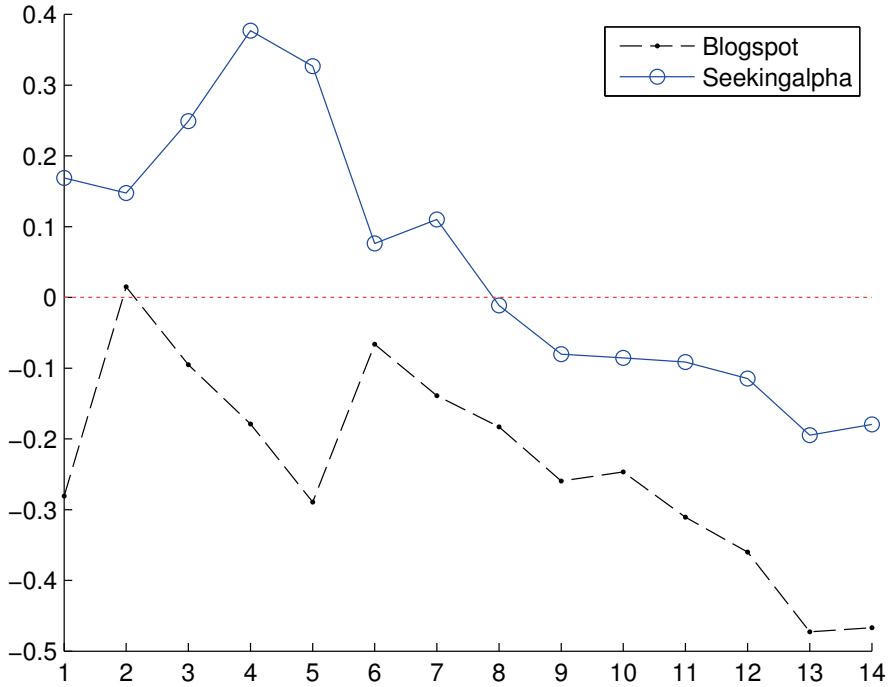


Figure 31: Plot of alpha for the long-short Blogspot & Seekingalpha portfolios, adjusting for transaction costs. The outcome depends on the number of stocks selected in each period in the long and short portfolios. The monthly alpha (in %) was estimated by Carhart's (1997) model.

First, Figure 31 shows the OLS alpha estimates of Carhart's (1997) model. The alpha estimates for the Seekingalpha dataset are higher for each N numbers of stocks selected in the long and short (sub-)portfolios. Especially for N=4,5 the superiority is obvious. The reason for the setback of the Blogspot-based simulation's result for these values of N is unclear (see the H4 section). For $N \geq 6$, the alpha estimates for Seekingalpha are about 0.2% higher.

Despite the graphical superiority of the Seekingalpha-based simulation results with respect to the alpha estimates, only partial statistical evidence rejecting any null hypothesis $\alpha_{A,N} \leq \alpha_{B,N}$ or the equivalent $\alpha_{A,N} - \alpha_{B,N} \leq 0$ of (the alternative hypothesis) H5.1 was found. $N \in \{1,2,\ldots,14\}$ refers to the number of stocks selected in each period in the long/short (sub-)portfolios and A (B) refers to the Seekingalpha (Blogspot) dataset-based portfolio simulation. See Table 45 for detailed results. Whereas all 14 alpha difference estimates are positive and indicate support for H5.1, the null hypothesis of a zero or negative alpha difference was only rejected for N=4,5 at the 10% significance level and for N=5 at the 5% significance level.

Table 45: H5.1 t-test results regarding Seekingalpha vs. Blogspot long-short portfolios. The tests are regarding the null hypothesis of H5.1 with respect to the difference of Carhart alphas of the long-short portfolios with *N* stocks selected in the long and short sub-portfolios.

N	$\alpha_{A,N} - \alpha_{B,N}$ estimate	90% CI bound	*t*-statistic	p	Reject $\alpha_{A,N} \leq \alpha_{B,N}$?	RF-test
1	0.0045	–0.0048	0.6246	0.2668	No	Yes
2	0.0013	–0.0059	0.2371	0.4065	No	No
3	0.0034	–0.0028	0.7094	0.2398	No	Yes
4	0.0056	0.0003	1.3544	0.0892*	Yes	Yes
5	0.0062	0.0014	1.6674	0.0491**	Yes	Yes
6	0.0014	–0.0034	0.3814	0.3518	No	Yes
7	0.0025	–0.0022	0.6861	0.2470	No	Yes
8	0.0017	–0.0028	0.4934	0.3113	No	Yes
9	0.0018	–0.0024	0.5506	0.2915	No	Yes
10	0.0016	–0.0025	0.5014	0.3085	No	Yes
11	0.0022	–0.0018	0.7133	0.2386	No	Yes
12	0.0025	–0.0013	0.8358	0.2025	No	Yes
13	0.0028	–0.0007	1.0353	0.1514	No	Yes
14	0.0029	–0.0006	1.0652	0.1446	No	Yes

NOTES: "$\alpha_{A,N} - \alpha_{B,N}$ estimate" is the OLS-estimate of the coefficient α_{A-B} (with *A* being a proxy for Seekingalpha and *B* being a proxy for Blogspot) in the model in Definition (4.15). See notes below Table 39 for further notes.

Figure 32 visualizes monthly mean excess returns (vs. the buy-and-hold portfolio's returns) for different *N* of both datasets.

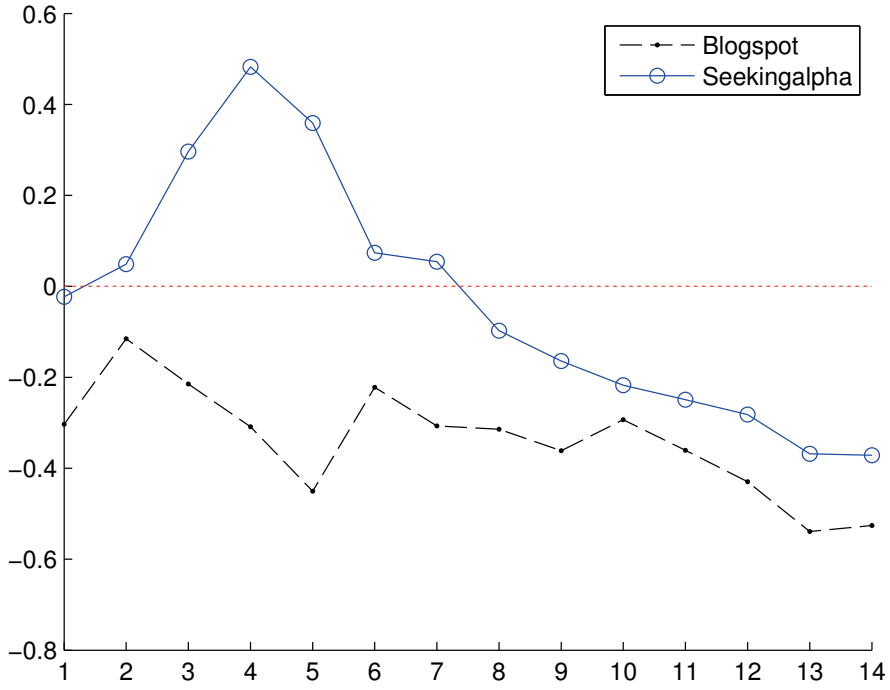


Figure 32: Plot of the mean excess returns of the long-short Blogspot & Seekingalpha portfolios, adjusting for transaction costs. The outcome depends on the number of stocks selected in each period in the long and short portfolios. Returns are monthly log returns (%).

The plot in Figure 32 is similar to the one in Figure 31 and indicates support for H5.2 (relating to monthly mean excess returns). For more evidence, a paired samples *t*-test of the

null hypothesis of the mean of the pairwise differences of excess return time series values of the Seekingalpha vs. the Blogspot portfolio simulations being zero or negative would test, whether higher mean excess returns for the Blogspot simulation can be rejected. For the paired samples *t*-test, the normality assumption regarding the difference time series has to hold. However, the normality assumption was violated for 10 of 14 values of *N* as indicated by the AD-test and the JB-test at the 10% significance level. For the remaining values of *N*=1,2,8,9 the null hypothesis of a zero or negative mean difference was not rejected at the 10% significance level.

Medians of excess returns (vs. the buy-and-hold portfolio) are not plotted because the difference of medians of excess returns does not equal the median of paired differences of excess returns, regarding which hypothesis H5.2 was formulated. Thus, Table 46 presents medians of paired differences of excess returns of the portfolio simulation using the Seekingalpha dataset vs. the Blogspot dataset and shows them to be greater than zero in 13 of 14 cases. The high number of positive median differences indicates support for H5.2 (related to the median difference). However, the corresponding Wilcoxon signed rank test (SRT) rejected the null hypothesis of a zero or negative median difference only for *N*=4,5,14 at the 10% significance level and for *N*=5 at the 5% significance level. On this basis hypothesis H5.2 (regarding the median difference) is accepted only for *N*=4,5,14.

Table 46: H5.2 test results regarding Seekingalpha vs. Blogspot long-short portfolios.

N	Median difference	Wilcoxon's SRT z-score	p	Reject null?
1	–0.0022	0.0994	0.4604	No
2	0.0003	–0.0626	0.5249	No
3	0.0041	0.8576	0.1955	No
4	0.0034	1.3656	0.0860*	Yes
5	0.0057	1.9103	0.0280**	Yes
6	0.0009	0.6515	0.2574	No
7	0.0052	1.0122	0.1557	No
8	0.0038	0.7251	0.2342	No
9	0.0018	0.6589	0.2550	No
10	0.0003	0.5264	0.2993	No
11	0.0002	0.6736	0.2503	No
12	0.0004	0.8871	0.1875	No
13	0.0017	1.1595	0.1231	No
14	0.0026	1.2846	0.0995*	Yes

NOTES: The performance *difference* of the simulated overall portfolio with *N* stocks selected in the long and short sub-portfolios using the Seekingalpha vs. the Blogspot dataset is measured in terms of the median of paired *differences* of excess log returns of the overall portfolios. Excess log returns were calculated as paired difference time series of a simulated overall portfolio's returns using investor sentiment indexes vs. the buy-and-hold portfolio's returns. Wilcoxon's (1945) signed rank test ("SRT") was used to non-parametrically test the null hypothesis of a zero or negative median of paired differences of excess log returns of the simulated overall portfolios. Regarding the SRT, the *z*-score and the *p*-value are reported. The significance level is denoted as follows: *** = 1%, ** = 5%, * = 10%. The null hypothesis of the SRT was rejected (see column "Reject null?") at a *p*-value indicating a 10% (or better) significance level. Results are irrespective of transaction costs.

Figure 33 presents the Sharpe ratio plots for Seekingalpha vs. Blogpost.

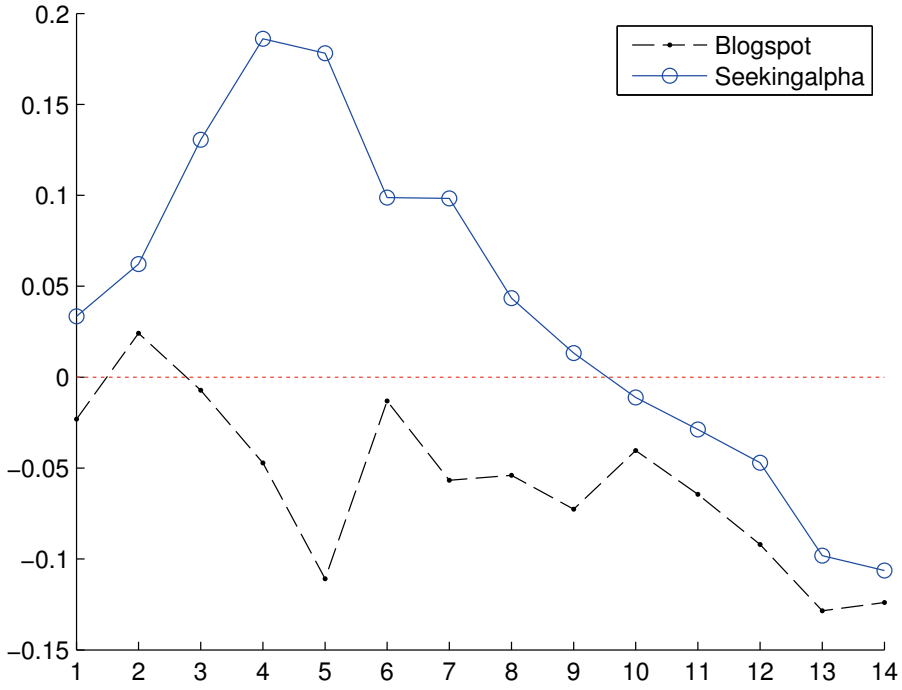


Figure 33: Plot of the monthly Sharpe ratio for the long-short Blogspot & Seekingalpha portfolios, adjusting for transaction costs. The Sharpe ratio depends on the number of stocks selected in each period in the long and short portfolios.

Figure 33 does not relate directly to testing hypotheses H5.1 and H5.2. Still, it helps to compare the investment performance of the long-short portfolio simulation using the two datasets based on the Sharpe ratio. Irrespective of N, the performance of the Seekingalpha-based simulation is better. However, for $N \geq 10$, the Sharpe ratios of the simulations using the two different datasets tend to converge. Convergence of Sharpe ratios seems reasonable because with more and more stocks selected from the overall set of stocks, the possibilities for differentiation become less.

Summarizing, broad indications (based on the sign of alpha difference estimates, mean/median differences of excess returns) were found that support H5.1 and H5.2, favoring the Seekingalpha-based simulation. Statistical evidence (rejecting the respective null hypotheses) was found only for two (three) values of N with respect to H5.1. (H5.2, relating to the median of paired differences of excess returns).

4.3.5.2 Without Transaction Costs

When ignoring transaction costs, the relative performance of the portfolio simulation using the Blogspot vs. the Seekingalpha dataset, which is studied by hypotheses H5.1 and H5.2, is identical. Thus the corresponding tables of results are not repeated here. Regarding the absolute levels of estimated alpha, mean excess returns, and the Sharpe ratio, the simulation outcomes using the Blogspot or the Seekingalpha dataset observe increases. First, Figure 34 shows the OLS alpha estimates of Carhart's (1997) model.

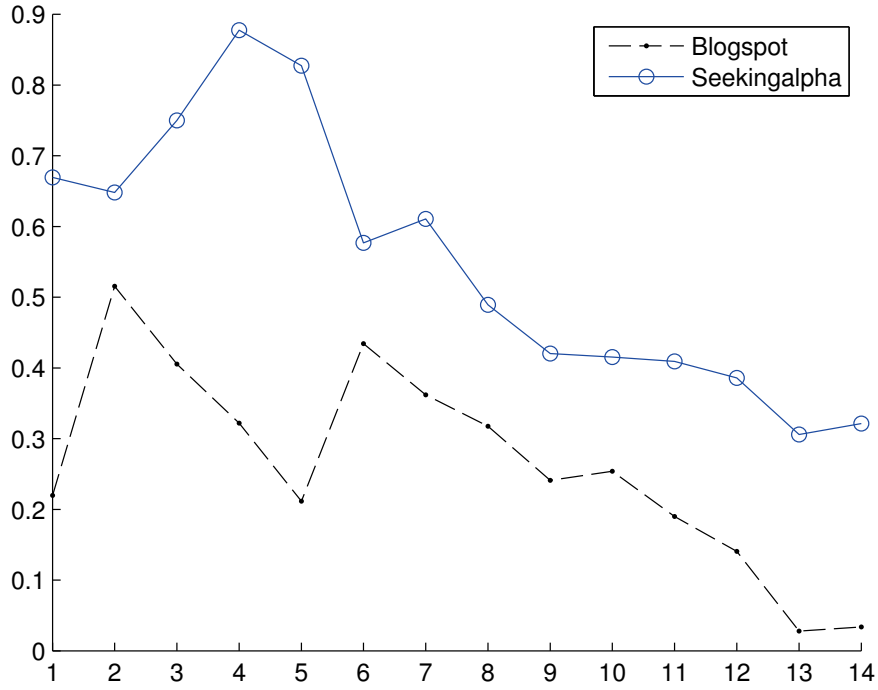


Figure 34: Plot of alpha for the long-short Blogspot & Seekingalpha portfolios, ignoring transaction costs. The outcome depends on the number of stocks selected in each period in the long and short portfolios. The monthly alpha (in %) was estimated by Carhart's (1997) model.

The benchmark-based investment performance is plotted in Figure 35.

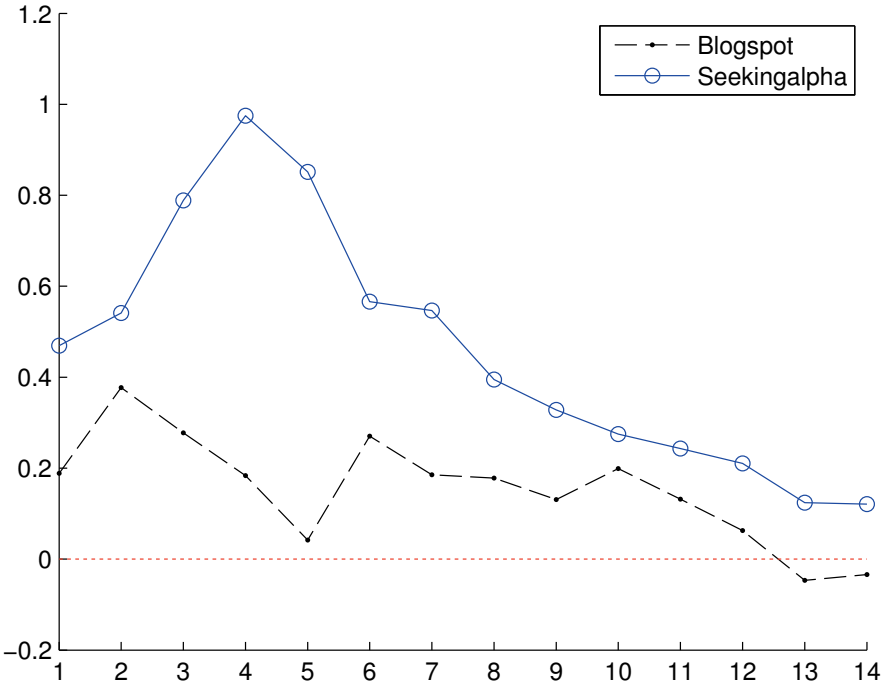


Figure 35: Plot of the mean excess returns of the long-short Blogspot & Seekingalpha portfolios, ignoring transaction costs. The outcome depends on the number of stocks selected in each period in the long and short portfolios. Returns are monthly log returns (%).

Finally, Figure 36 provides the (monthly) Sharpe ratio plots.

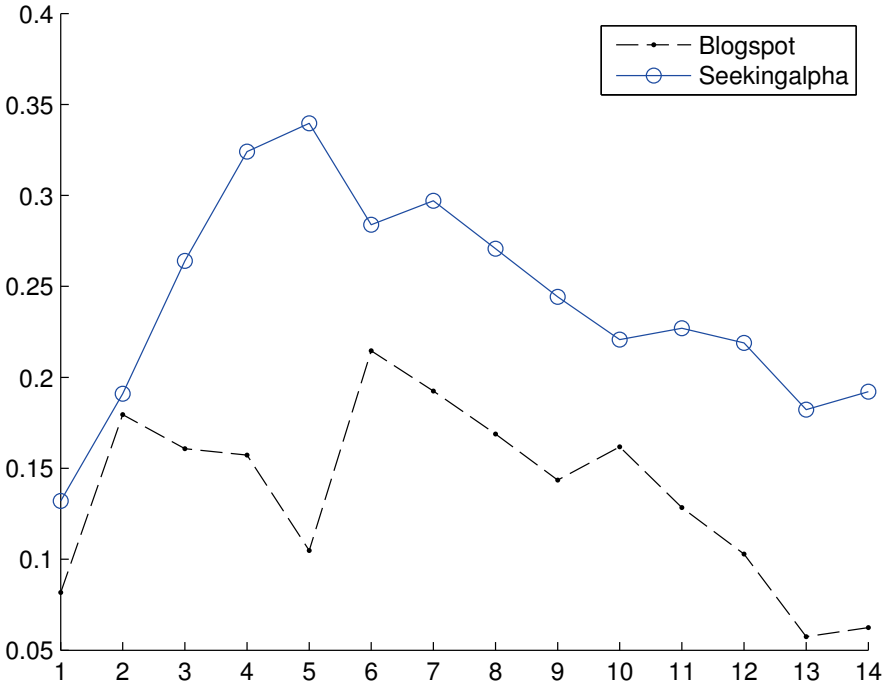


Figure 36: Plot of the monthly Sharpe ratio for the long-short Blogspot & Seekingalpha portfolios, ignoring transaction costs. The Sharpe ratio depends on the number of stocks selected in each period in the long and short portfolios.

The relative shape of the (monthly) Sharpe ratio plots in Figure 36 differ from the ones that account for transaction costs. For N=2, the Sharpe ratio for the Blogspot-based simulation is almost as high as for the Seekingalpha-based simulation. For $N \geq 3$, the Sharpe ratio of the Seekingalpha-based simulation is about 0.1 higher compared to the Blogspot-based simulation's Sharpe ratio. Thus, the Sharpe ratio investment performance measure is also in favor of the Seekingalpha-based simulation when ignoring transaction costs.

4.3.6 Hypothesis H6: Effects in Relation to the Momentum Effect

This section compares results of the long-short portfolio simulation using the Seekingalpha dataset or the Blogspot dataset to results of the long-short portfolio simulation using the momentum strategy for stock selection (see Section 4.2.1) to test hypotheses H6.1 and H6.2. These hypotheses postulate the investor sentiment indexes from blog documents to have a higher effect on abnormal returns than cross-sectional price momentum information. Each simulation was performed over different numbers of stocks selected in the long and short (sub-)portfolios. Before testing H6.1 and H6.2, a test for the existence of abnormal returns due to the price momentum effect was conducted within this study's portfolio simulation setup.

4.3.6.1 Existence of Abnormal Momentum Returns

Table 47 presents alpha estimates of the long-short portfolio simulation using price momentum information. The only positive alpha estimate was found for N=1, when accounting for transaction costs. This alpha estimate appears to be a large positive outlier because all other alpha estimates are much lower. When ignoring transaction costs, all alpha estimates are positive. However, there is no statistical evidence at the 10% significance level rejecting the null hypothesis of a zero or negative alpha for any N=1,2,…,14. That is, there is no statistical evidence for a price momentum effect on alpha in this study's portfolio simulation setup.

Table 47: Test results for positive alpha of the momentum long-short portfolio with N stocks selected in the long and short sub-portfolios. The tests are regarding the null hypothesis of zero or negative alpha.

N	α estimate	90% CI bound	*t*-statistic	p	Reject α≤0?	RF-test
With transaction costs						
1	0.0064	–0.0069	0.6252	0.2672	No	Yes
2	–0.0014	–0.0097	–0.2179	0.5858	No	Yes
3	–0.0008	–0.0086	–0.1261	0.5500	No	Yes
4	–0.0012	–0.0080	–0.2385	0.5938	No	Yes
5	–0.0020	–0.0082	–0.4247	0.6637	No	Yes
6	–0.0018	–0.0073	–0.4133	0.6595	No	Yes
7	–0.0014	–0.0062	–0.3813	0.6478	No	Yes
8	–0.0021	–0.0062	–0.6838	0.7515	No	Yes
9	–0.0030	–0.0065	–1.1092	0.8639	No	Yes
10	–0.0022	–0.0055	–0.8930	0.8121	No	Yes
11	–0.0027	–0.0056	–1.2293	0.8879	No	Yes
12	–0.0033	–0.0059	–1.6896	0.9516	No	Yes
13	–0.0033	–0.0057	–1.7422	0.9565	No	Yes
14	–0.0034	–0.0058	–1.8235	0.9632	No	Yes
Without transaction costs						
1	0.0114	–0.0019	1.1136	0.1352	No	Yes
2	0.0036	–0.0047	0.5666	0.2867	No	Yes
3	0.0042	–0.0036	0.7058	0.2416	No	Yes
4	0.0038	–0.0030	0.7206	0.2371	No	Yes
5	0.0030	–0.0032	0.6308	0.2654	No	Yes
6	0.0032	–0.0023	0.7530	0.2273	No	Yes
7	0.0036	–0.0012	0.9650	0.1694	No	Yes
8	0.0029	–0.0012	0.9182	0.1813	No	Yes
9	0.0020	–0.0015	0.7542	0.2270	No	Yes
10	0.0028	–0.0004	1.1181	0.1342	No	Yes
11	0.0023	–0.0006	1.0186	0.1564	No	Yes
12	0.0017	–0.0009	0.8499	0.1995	No	Yes
13	0.0017	–0.0007	0.9246	0.1796	No	Yes
14	0.0016	–0.0008	0.8768	0.1922	No	Yes

NOTES: see notes below Table 37.

Along this line, Table 48 shows no statistical evidence rejecting the null hypothesis of monthly median excess returns to be zero or negative. The null hypothesis of zero or negative monthly mean excess returns cannot be tested using a *t*-test due to the normality assumption being violated for all values of N. That is, the null hypothesis of a normal distribution of the

excess return time series was rejected at the 10% significance level for all N=1,2,…,14 by both, the JB-test and the AD-test.

Table 48: Test results for positive median excess returns of the momentum portfolio.

N	Sharpe ratio	Monthly mean excess return	Monthly median excess return	Wilcoxon's SRT z-score	p	Reject null?
With transaction costs						
1	0.0852	0.5181	–0.5552	–0.7398	0.7703	No
2	–0.0302	–0.3502	–1.7145	–1.4981	0.9329	No
3	0.0103	–0.1458	–0.8653	–1.2846	0.9005	No
4	–0.0053	–0.2142	–1.5545	–1.7557	0.9604	No
5	–0.0352	–0.3235	–1.9557	–1.8367	0.9669	No
6	–0.0346	–0.3069	–1.8262	–1.6159	0.9469	No
7	–0.0324	–0.2883	–1.3547	–1.5717	0.9420	No
8	–0.0499	–0.3327	–1.6649	–1.5570	0.9403	No
9	–0.0989	–0.4418	–1.5011	–1.6527	0.9508	No
10	–0.0957	–0.4078	–1.5470	–1.5570	0.9403	No
11	–0.1365	–0.4710	–1.4488	–1.6232	0.9477	No
12	–0.1647	–0.5057	–1.6166	–1.6306	0.9485	No
13	–0.1610	–0.4880	–1.4097	–1.6453	0.9500	No
14	–0.1836	–0.5038	–1.2910	–1.6453	0.9500	No
Without transaction costs						
1	0.1452	1.0104	–0.0546	–0.3276	0.6284	No
2	0.0660	0.1421	–1.2139	–0.9754	0.8353	No
3	0.1192	0.3465	–0.3647	–0.7693	0.7791	No
4	0.1201	0.2781	–1.0539	–1.1668	0.8784	No
5	0.0999	0.1688	–1.4551	–1.2478	0.8939	No
6	0.1177	0.1853	–1.3256	–1.0785	0.8596	No
7	0.1379	0.2040	–0.8541	–0.9386	0.8260	No
8	0.1289	0.1596	–1.1643	–0.9975	0.8407	No
9	0.1002	0.0505	–1.0004	–0.9828	0.8371	No
10	0.1274	0.0845	–1.0463	–0.8797	0.8105	No
11	0.1094	0.0213	–0.9482	–0.9312	0.8241	No
12	0.0991	–0.0134	–1.1160	–1.0343	0.8495	No
13	0.1123	0.0043	–0.9091	–1.0196	0.8460	No
14	0.1122	–0.0116	–0.7904	–1.0564	0.8546	No

NOTES: see notes below Table 38, where the simulated portfolio refers to the portfolio simulation using momentum information.

4.3.6.2 Comparing Sentiment Portfolios to Momentum Portfolios

On comparing portfolio simulation results based on the investor sentiment datasets and price momentum information, results are separated in (1) results with accounting for transaction costs, and (2) results without transaction costs.

4.3.6.2.1 With Transaction Costs

The plots in this section compare the estimates of alpha, mean excess return, and Sharpe ratio over different numbers of stocks N selected in the long or short (sub-)portfolios of the simulations when accounting for transaction costs. Except for N=1, the simulation results displayed by the plots show the results based on the Seekingalpha dataset to be superior compared to the momentum portfolio simulation (see Figure 37, Figure 38, and Figure 39).

The Blogspot-based simulation results are roughly on the same level like the ones of the momentum simulation. These results indicate support for H6.1 and H6.2 (regarding mean excess returns) for the Seekingalpha dataset.

Figure 37 shows the Carhart (1997) alpha estimates.

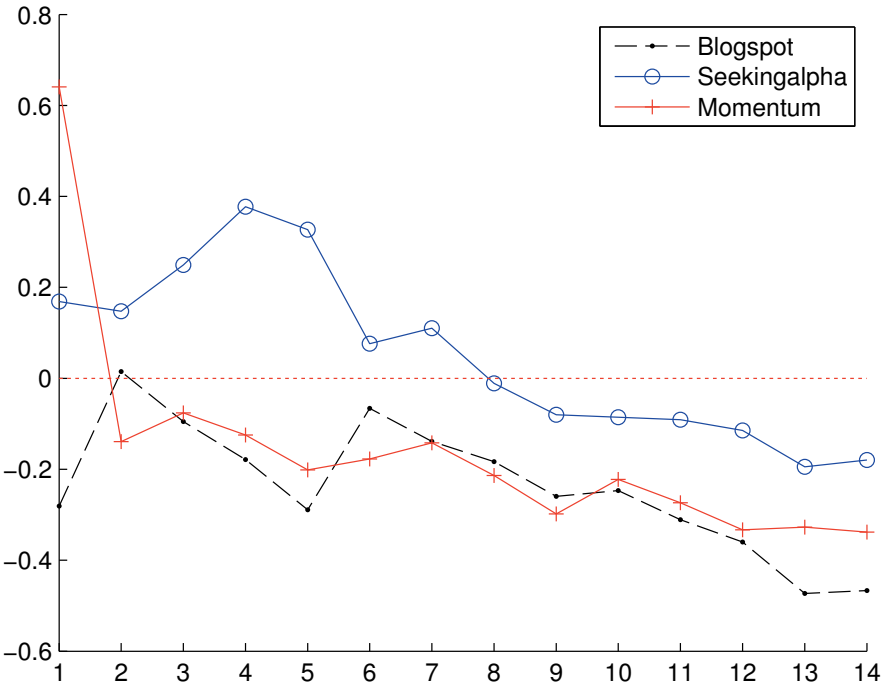


Figure 37: Plot of alpha for the long-short Blogspot & Seekingalpha portfolios and the momentum simulation's portfolio, adjusting for transaction costs. The outcome depends on the number of stocks selected in each period in the long and short portfolios (displayed horizontally). The monthly alpha (in %) was estimated by Carhart's (1997) model.

Table 49 presents the estimates of the differences in alpha between the portfolio simulation using the Seekingalpha dataset vs. the momentum-based simulation. Except for N=1, all estimates are positive, indicating broad support for hypothesis H6.1. However, there is no statistical evidence rejecting any null hypothesis $\alpha_{A,N} \leq \alpha_{B,N}$ or the equivalent $\alpha_{A,N} - \alpha_{B,N} \leq 0$ to (the alternative hypothesis) H6.1 with 10% statistical significance, where A refers to the Seekingalpha-based simulation and B refers to the momentum-based simulation.

Table 49: H6.1 t-test results for Seekingalpha vs. momentum long-short portfolios. The tests are regarding the null hypothesis of H6.1 with respect to the difference of Carhart alphas of the long-short portfolios with *N* stocks selected in the long and short sub-portfolios.

N	$\alpha_{A,N} - \alpha_{B,N}$ estimate	90% CI bound	*t*-statistic	p	Reject $\alpha_{A,N} \leq \alpha_{B,N}$?	RF-test
1	–0.0047	–0.0200	–0.3977	0.6542	No	Yes
2	0.0029	–0.0072	0.3670	0.3572	No	Yes
3	0.0033	–0.0060	0.4520	0.3261	No	Yes
4	0.0050	–0.0028	0.8237	0.2059	No	Yes
5	0.0053	–0.0017	0.9717	0.1667	No	Yes
6	0.0025	–0.0037	0.5217	0.3015	No	Yes
7	0.0025	–0.0031	0.5829	0.2806	No	Yes
8	0.0020	–0.0028	0.5436	0.2939	No	Yes
9	0.0022	–0.0020	0.6656	0.2535	No	Yes
10	0.0014	–0.0026	0.4433	0.3292	No	Yes
11	0.0018	–0.0017	0.6627	0.2544	No	Yes
12	0.0022	–0.0011	0.8504	0.1985	No	Yes
13	0.0013	–0.0018	0.5512	0.2913	No	Yes
14	0.0016	–0.0015	0.6733	0.2511	No	Yes

NOTES: "$\alpha_{A,N} - \alpha_{B,N}$ estimate" is the OLS-estimate of the coefficient α_{A-B} (with *A* being a proxy for Seekingalpha and *B* being a proxy for momentum) in the model in Definition (4.15). See notes below Table 39 for further notes.

When comparing the Blogspot-based portfolio simulation's alphas (*A*) with the momentum-based portfolio simulation's alphas (*B*) for different values of *N*, the sign of the estimate of the alpha difference is positive only for *N*=2,6,7,8,9 (see Table 50). There is no evidence rejecting the null hypothesis in favor of the alternative H6.1 regarding the Blogspot dataset (at the 10% significance level).

Table 50: H6.1 t-test results for Blogspot vs. momentum long-short portfolios. The tests are regarding the null hypothesis of H6.1 with respect to the difference of Carhart alphas of the long-short portfolios with *N* stocks selected in the long and short sub-portfolios.

N	$\alpha_{A,N} - \alpha_{B,N}$ estimate	90% CI bound	*t*-statistic	p	Reject $\alpha_{A,N} \leq \alpha_{B,N}$?	RF-test
1	–0.0092	–0.0234	–0.8374	0.7979	No	Yes
2	0.0015	–0.0077	0.2140	0.4155	No	Yes
3	–0.0002	–0.0088	–0.0290	0.5115	No	Yes
4	–0.0005	–0.0081	–0.0929	0.5369	No	Yes
5	–0.0009	–0.0078	–0.1627	0.5645	No	Yes
6	0.0011	–0.0056	0.2136	0.4156	No	Yes
7	0.0000	–0.0060	0.0064	0.4974	No	Yes
8	0.0003	–0.0051	0.0726	0.4711	No	Yes
9	0.0004	–0.0045	0.1013	0.4597	No	Yes
10	–0.0002	–0.0049	–0.0671	0.5267	No	Yes
11	–0.0004	–0.0048	–0.1075	0.5427	No	Yes
12	–0.0003	–0.0043	–0.0855	0.5340	No	Yes
13	–0.0015	–0.0052	–0.4985	0.6904	No	Yes
14	–0.0013	–0.0051	–0.4367	0.6684	No	Yes

NOTES: see notes below Table 49, with exchanging *A* for Blogspot.

Evaluating H6.2 with respect to monthly mean excess returns (relative to the buy-and-hold portfolio's returns) for different *N*, Figure 38 shows the mean excess returns for all investor sentiment datasets and the momentum-based portfolio simulation. The results of

these simulations are similar to the plot of estimates of alpha in Figure 37. The higher results (except for N=1) compared to the momentum simulation indicate support for H6.2 (relating to monthly mean excess returns) for the Seekingalpha dataset.

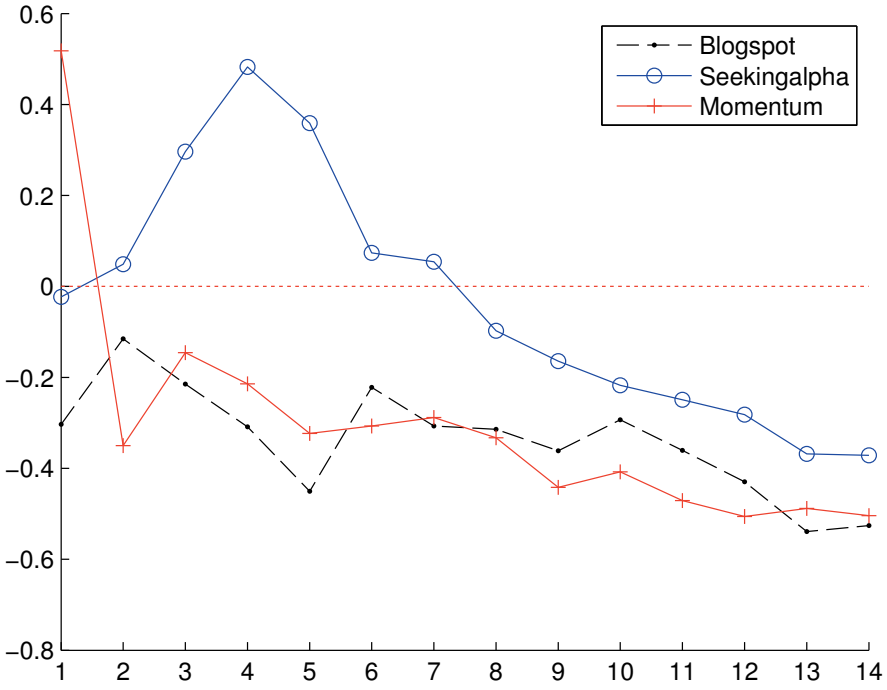


Figure 38: Plot of the mean excess returns for the long-short Blogspot & Seekingalpha portfolios and the momentum simulation's portfolio, adjusting for transaction costs. The outcome depends on the number of stocks selected in each period in the long and short portfolios (displayed horizontally). The returns are monthly log returns (in %).

The null hypothesis of the mean of the pairwise differences of excess return time series values of the Blogspot-based vs. momentum-based portfolio simulations being zero or negative (postulating higher mean excess returns for the momentum portfolio) was tested for N=7,8,…,11. The null hypothesis was not rejected for these values of N (see Table 51). For other values of N, normality was rejected by the AD-test and the JB-test, thus invalidating a possible t-test. Regarding Seekingalpha-based vs. momentum-based simulations' pairwise differences of excess return time series, the mean of these pairwise differences is positive for $N \geq 2$. Hypothesis test results are not provided because of rejection of normality of the pairwise difference time series values for all values of N except for N=3, for which the null hypothesis of the t-test with respect to the mean of the pairwise differences of excess return time series was not rejected at the 10% significance level.

Table 51: Test results for higher mean excess returns of Blogspot vs. momentum long-short portfolios.

N	Mean difference	RN (AD)?	RN (JB)?	*t*-statistic	p	Reject null?
1	–0.0082	Yes	Yes			
2	0.0023	Yes	Yes			
3	–0.0007	Yes	Yes			
4	–0.0009	Yes	Yes			
5	–0.0013	Yes	Yes			
6	0.0009	Yes	Yes			
7	–0.0002	No	No	–0.0534	0.5212	No
8	0.0002	No	No	0.0535	0.4787	No
9	0.0008	No	No	0.2423	0.4047	No
10	0.0011	No	No	0.3839	0.3512	No
11	0.0011	No	No	0.3770	0.3538	No
12	0.0008	Yes	Yes			
13	–0.0005	Yes	Yes			
14	–0.0002	Yes	Yes			

NOTES: The performance *difference* of the simulated overall portfolio with *N* stocks selected in the long and short sub-portfolios using the Blogspot dataset vs. momentum information is measured in terms of the mean of paired differences of excess log returns of the overall portfolios. Excess log returns were calculated as paired difference time series of a simulated overall portfolio's returns (using either the Blogspot dataset or momentum information) vs. the buy-and-hold portfolio's returns. A *t*-test was used to test the null hypothesis of zero or negative mean of paired differences of excess log returns of the simulated overall portfolio using the Blogspot dataset vs. momentum information. The normality assumption of the *t*-test was verified by the AD-test (i.e., Anderson & Darling, 1952, 1954; D'Agostino & Stephens, 1986) and the JB-test (i.e., Jarque & Bera, 1987). The null hypothesis of normally distributed paired differences of excess log returns of the two simulated overall portfolios was rejected at the 10% significance level as indicated by the two columns "RN (AD)?" and "RN (JB)?". Regarding the *t*-test, the "*t*-statistic" and the *p*-value are reported. The null hypothesis of the *t*-test was rejected (see column "Reject null?") at a *p*-value indicating a 10% significance level. Results are irrespective of transaction costs.

Medians of excess returns are not plotted because the difference of medians of excess returns does not equal the median of paired differences of excess returns, regarding which hypothesis H6.2 was formulated. Thus, Table 52 presents medians of paired differences of excess returns (vs. the buy-and-hold portfolio) of the portfolio simulation using the Seekingalpha dataset vs. the momentum-based portfolio simulation. All median differences are greater than zero for N=1,2,…,14, indicating broad support for H6.2 (related to the median difference). The null hypothesis of the alternative H6.2 was rejected for 8 values of N: for N=5,12 at the 5% significance level and for N=4,5,6,7,8,9,11,12 at the 10% significance level.

Table 52: H6.2 test results for Seekingalpha vs. momentum long-short portfolios.

N	Median difference	Wilcoxon's SRT z-score	p	Reject null?
1	0.0012	–0.3276	0.6284	No
2	0.0063	1.1079	0.1339	No
3	0.0032	0.8503	0.1976	No
4	0.0046	1.4097	0.0793*	Yes
5	0.0045	1.6527	0.0492**	Yes
6	0.0083	1.2920	0.0982*	Yes
7	0.0090	1.5717	0.0580*	Yes
8	0.0054	1.4392	0.0750*	Yes
9	0.0058	1.6232	0.0523*	Yes
10	0.0030	1.1153	0.1324	No
11	0.0028	1.3729	0.0849*	Yes
12	0.0023	1.6748	0.0470**	Yes
13	0.0009	1.0858	0.1388	No
14	0.0025	1.1668	0.1216	No

NOTES: The performance *difference* of the simulated overall portfolio with *N* stocks selected in the long and short sub-portfolios using the Seekingalpha dataset vs. momentum information is measured in terms of the median of paired differences of excess log returns of the overall portfolios. Excess log returns were calculated as paired difference time series of a simulated overall portfolio's returns (using either the Seekingalpha dataset or momentum information) vs. the buy-and-hold portfolio's returns. Wilcoxon's (1945) signed rank test ("SRT") was used to non-parametrically test the null hypothesis of zero or negative median of paired differences in excess log returns of the simulated overall portfolios. Regarding the SRT, the *z*-score and the *p*-value are reported. The significance level is denoted as follows: *** = 1%, ** = 5%, * = 10%. The null hypothesis of the SRT was rejected (see column "Reject null?") at a *p*-value indicating a 10% (or better) significance level. Results are irrespective of transaction costs.

Regarding the Blogspot dataset, the medians of paired differences of excess returns vs. the momentum-based simulation's excess returns are positive only in 9 cases (see Table 53). The corresponding Wilcoxon signed rank test results do not reject the null hypothesis of zero or negative median differences. Thus, hypothesis H6.2 (regarding the median difference for the Blogspot dataset) was rejected.

Table 53: H6.2 test results for Blogspot vs. momentum long-short portfolios.

N	Median difference	Wilcoxon's SRT z-score	p	Reject null?
1	0.0018	0.0184	0.4927	No
2	0.0072	0.5926	0.2767	No
3	0.0043	0.0699	0.4721	No
4	–0.0008	0.3938	0.3468	No
5	–0.0024	0.1141	0.4546	No
6	0.0009	0.3791	0.3523	No
7	0.0014	–0.0184	0.5073	No
8	–0.0003	0.1141	0.4546	No
9	0.0029	0.0699	0.4721	No
10	0.0015	0.2466	0.4026	No
11	–0.0011	0.2981	0.3828	No
12	0.0007	0.2172	0.4140	No
13	0.0016	0.0626	0.4751	No
14	–0.0009	–0.0405	0.5161	No

NOTES: see notes below Table 52, with exchanging Seekingalpha for Blogspot.

Finally, Figure 39 compares the investment performance of the long-short portfolio simulation using the two investor sentiment datasets and the momentum-based simulation

based on the (monthly) Sharpe ratio for different *N*. For *N*=1, the momentum-based portfolio simulation observes the highest Sharpe ratio of all simulations. However, it is only slightly higher compared to the Sharpe ratio of the Seekingalpha-based simulation. For *N*>1, the Seekingalpha-based simulation observes the highest Sharpe ratios in comparison to the other simulations. For *N*=3,4,5,6,7 the Seekingalpha-based simulation's Sharpe ratio is higher than the momentum-based simulation's Sharpe ratio for any *N*. The higher Sharpe ratio indicates the monthly investor sentiment index of the Seekingalpha dataset to generate a better investment performance than a strategy informed only by price-based information.

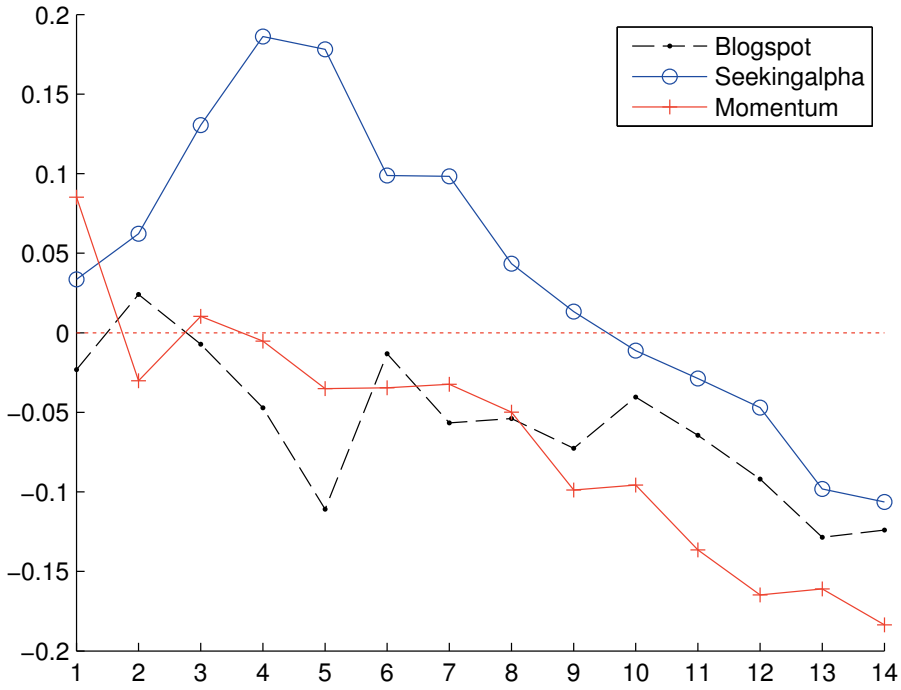


Figure 39: Plot of the monthly Sharpe ratio for the long-short Blogspot & Seekingalpha portfolios and the momentum simulation's portfolio, adjusting for transaction costs. The ratio depends on the number of stocks selected in each period in the long and short portfolios.

Summarizing this section, simulation results indicate support for H6.1 and H6.2 for *N*>1 and the Seekingalpha dataset. Although there is only statistical evidence rejecting the null hypothesis of H6.2 (with respect to the median difference of excess returns) at the 5% and 10% significance level for a broad range of eight values of *N*, the results indicate the Seekingalpha-based investor sentiment indexes to have higher predictive power concerning abnormal returns compared to price momentum information. Although positive estimates of alpha differences and positive means/medians of paired differences in excess returns were observed for Blogspot-based investor sentiment indexes, the level of this outperformance is rather small and there is no statistical evidence.

4.3.6.2.2 Without Transaction Costs

When ignoring transaction costs, the relative performance of the portfolio simulations using the Seekingalpha or the Blogspot dataset in comparison to the momentum portfolio simulation measured by the estimate of the alpha difference, the mean or median of paired

differences of excess returns is identical to the above case when adjusting for transaction costs. Therefore, the plots of absolute levels of estimated alpha and mean excess returns are provided for different numbers of stocks *N* selected in each period. Providing the plot of median excess returns would not make sense because the difference of medians does not equal the median of paired differences, regarding which hypothesis H6.2 was formulated.

Figure 40 shows the alpha estimates with respect to Carhart's (1997) model.

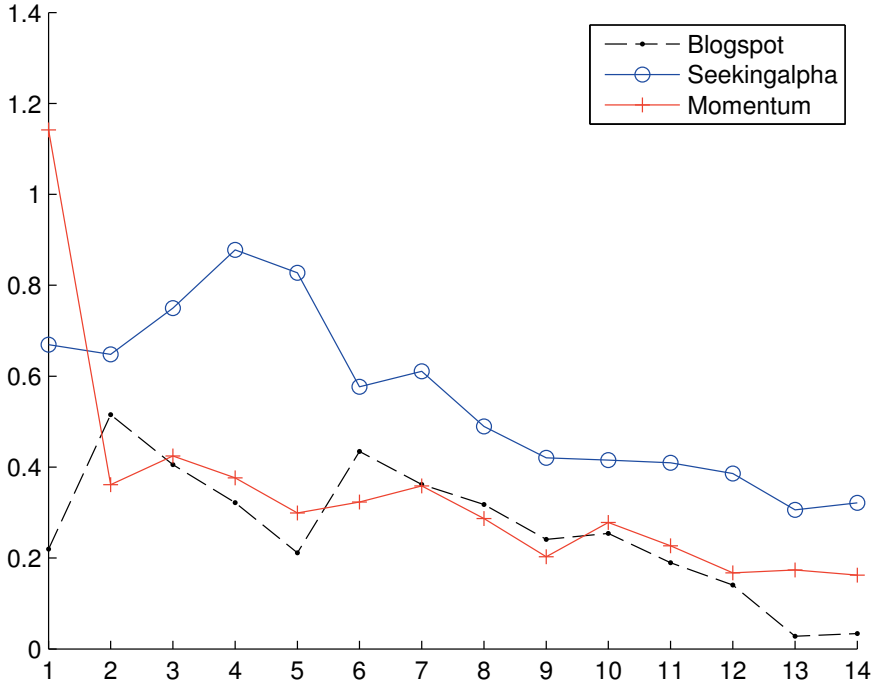


Figure 40: Plot of alpha for the long-short Blogspot & Seekingalpha portfolios and the momentum simulation's portfolio, ignoring transaction costs. The outcome depends on the number of stocks selected in each period in the long and short portfolios (displayed horizontally). The monthly alpha (in %) was estimated by Carhart's (1997) model.

The monthly mean excess returns for the simulations using the Blogspot/Seekingalpha investor sentiment indexes and the momentum-based simulation are plotted in Figure 41.

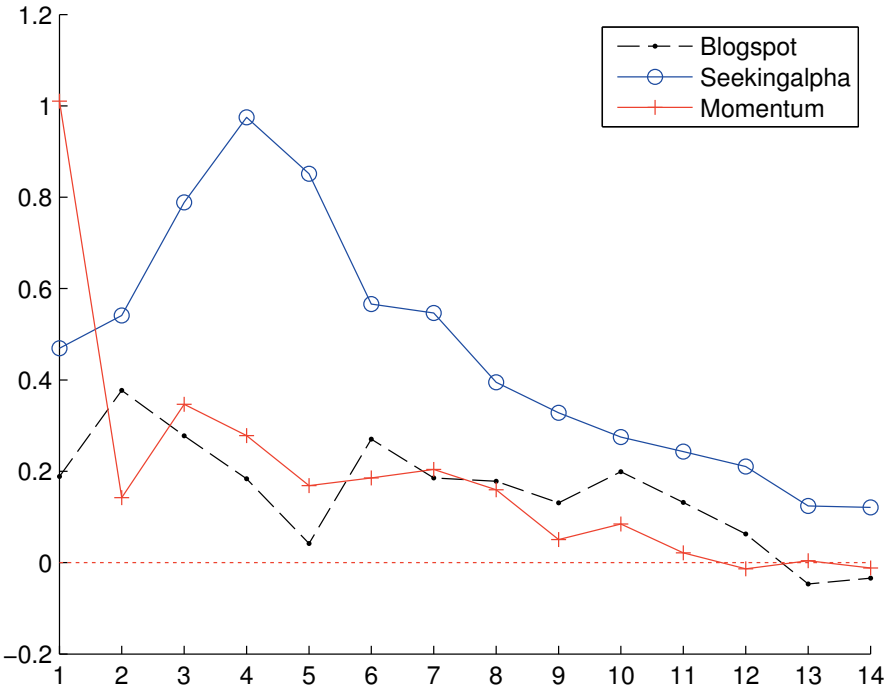


Figure 41: Plot of the mean excess returns for the long-short Blogspot & Seekingalpha portfolios and the momentum simulation's portfolio, ignoring transaction costs. The outcome depends on the number of stocks selected in each period in the long and short portfolios (displayed horizontally). The returns are monthly log returns (in %).

Figure 42 presents the Sharpe ratio plots.

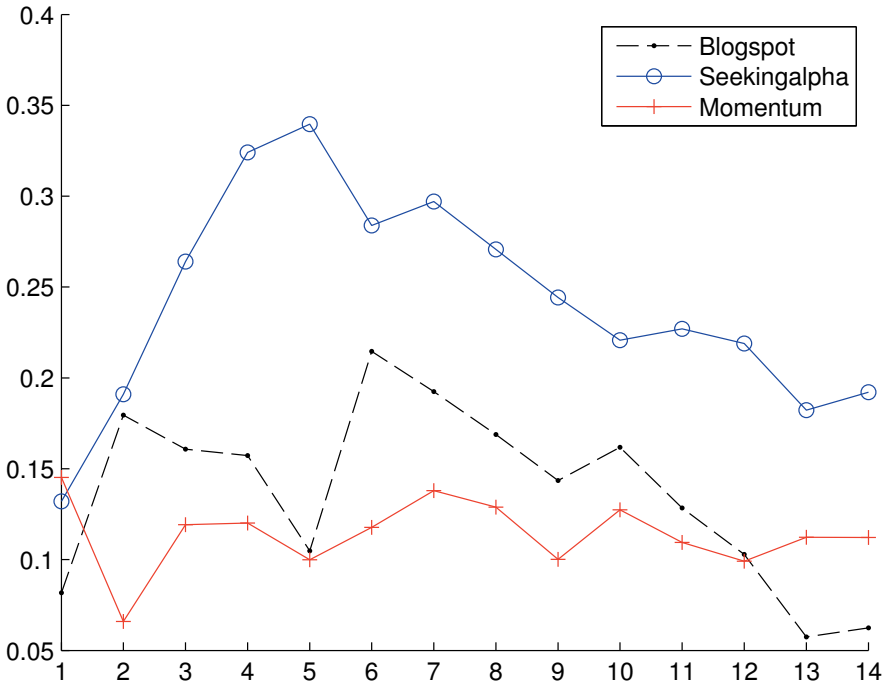


Figure 42: Plot of the monthly Sharpe ratio for the long-short Blogspot & Seekingalpha portfolios and the momentum simulation's portfolio, ignoring transaction costs. The ratio depends on the number of stocks selected in each period in the long and short portfolios.

The (monthly) Sharpe ratio plots in Figure 42 differ in relative shape from the ones in Figure 39, which account for transaction costs: Even the Sharpe ratios of the portfolio simulation using the Blogspot dataset are higher than the momentum-based simulation's Sharpe ratios for most *N*. The higher Sharpe ratios indicate some theoretical investment value for the investor sentiment indexes of the Blogspot dataset. For the Seekingalpha-based simulation, the differences in Sharpe ratios to the momentum-based simulation's Sharpe ratios are larger in comparison to the previous section's plot, which accounts for transaction costs. Thus, the theoretical investment value seems to be even better for Seekingalpha when ignoring transaction costs.

4.4 Summary of Hypotheses Testing

This section provides a summary of the tests of the hypotheses on effects of using aggregated investor sentiment document scores from blog documents for stock selection. Portfolio simulations were conducted to this respect using the datasets of (1) Seekingalpha, and (2) Blogspot, with either ignoring or accounting for transaction costs, as defined in Section 4.2.1. The effects were measured on the portfolio level by monthly mean abnormal returns in terms of estimates of Carhart's (1997) alpha and by monthly mean or median excess returns (versus the buy-and-hold portfolio's returns).

Rejection of a null hypothesis (in favor of a respective alternative hypothesis H1.1, H1.2, …, or H6.2) at a certain significance level by a hypothesis test is marked in bold in the overview tables Table 54, Table 55, and Table 56. If a hypothesis test was not possible due to violation of an underlying assumption of the test, the overview tables provide the value of (the estimate of) the variable of interest in the null hypothesis (e.g., mean excess returns). The sign and magnitude of the (estimated) variable provide at least an indication for judging the hypothesis.

Table 54 shows statistical evidence for posited effects in H1 and H2 to clearly exist for the Seekingalpha dataset when ignoring transaction costs. Strongest evidence regarding positive alpha exists for the long portfolio. This observation suggests primarily the highest investor sentiment indexes to relate to abnormal returns. Regarding the Blogspot-based investor sentiment index, there is no statistical evidence to this respect. However, the estimated alpha of the long-short and the short-only portfolio simulation is signed positively.

Table 54: Overview of test results for H1, H2, and H3 without transaction costs. The tests are with respect to the formulated null hypotheses to study portfolio simulation results.

H	Hypothesis		Result		
	"Using the investor sentiment index (in each period) for selecting the …		*… results (over the whole simulation, on the portfolio level) in …*		
		Dataset	***α≤0?***	***mean(excrt.)≤0?***	***median(excrt.)≤0?***
H1	*… highest (lowest) ranked stocks in a long (short) portfolio …*	Seeking-alpha	***p*=0.0013*** **	mean=0.85% but *t*-test was n.a.	*p*=0.4169
		Blogspot	*p*=0.2094	mean=0.04% but *t*-test was n.a.	*p*=0.7183
H2	*…highest ranked stocks in a long portfolio...*	Seeking-alpha	***p*=0.0004*** **	mean=0.66 % but *t*-test was n.a.	***p*=0.0036*** **
		Blogspot	*p*=0.6704	mean=–0.28 % *t*-test was n.a.	*p*=0.8798
H3	*…lowest ranked stocks in a short portfolio…*	Seeking-alpha	*p*=0.1657	mean=0.59 % but *t*-test was n.a.	*p*=0.6284
		Blogspot	*p*=0.2116	mean=0.05 % but *t*-test was n.a.	*p*=0.6773

NOTES: α = the alpha in Carhart's (1997) model; *** = 1% significance level; "excrt" = excess log returns of the simulated portfolio's log returns using investor sentiment indexes of a certain dataset vs. the buy-and-hold portfolio's log returns; "mean" = monthly mean; "median" = monthly median; "n.a." = not applicable due to violation of (an) assumption(s) of the test.

Table 55 shows the effects to diminish when accounting for transaction costs. Diminishment of effects is expressed in much higher *p*-values and lower estimates of mean or median abnormal returns. However, there is still evidence for an effect on abnormal returns in terms of alpha for the long portfolio simulation using Seekingalpha as formulated in H2.1. Furthermore, mean excess returns are still positive for all three portfolio simulations using the Seekingalpha dataset, thus indicating some positive value for investors.

Table 55: Overview of test results for H1, H2, and H3 with transaction costs. The tests are with respect to the formulated null hypotheses to study portfolio simulation results.

H	Hypothesis		Result		
	"Using the investor sentiment index (in each period) for selecting the …		*… results (over the whole simulation, on the portfolio level) in …*		
		Dataset	***α≤0?***	***mean(excrt.)≤0?***	***median(excrt.)≤0?***
H1	*… highest (lowest) ranked stocks in a long (short) portfolio …*	Seeking-alpha	*p*=0.1091	mean=0.36% but *t*-test was n.a.	*p*=0.6450
		Blogspot	*p*=0.8650	mean=–0.45% *t*-test was n.a.	*p*=0.8912
H2	*…highest ranked stocks in a long portfolio...*	Seeking-alpha	***p*=0.0981***	mean=0.16% but *t*-test was n.a.	*p*=0.2574
		Blogspot	*p*=0.9741	mean=–0.78 % *t*-test was n.a.	*p*=0.9954
H3	*…lowest ranked stocks in a short portfolio…*	Seeking-alpha	*p*=0.5417	mean=0.09% but *t*-test was n.a.	*p*=0.7257
		Blogspot	*p*=0.7200	mean=–0.45 % *t*-test was n.a.	*p*=0.7791

NOTES: see notes below Table 54.

Table 56 presents test results of the null hypotheses of the alternatives H4, H5, and H6.

Table 56: Overview of test results for H4, H5, and H6, irrespective of transaction costs. The tests are with respect to the formulated null hypotheses, comparing different portfolio simulation configurations.

H	Hypothesis		Result		
			… results (over the whole simulation, on the portfolio level) in …		
		Dataset	$\alpha_N - \alpha_{N+1} \leq 0$**?**	***mean(pd(excrt.$_N$, excrt.$_{N+1}$)*** ≤ 0**?**	***median(pd(excrt.$_N$, excrt.$_{N+1}$)*** ≤ 0**?**
H4	*"**Selecting N stocks** in a long portfolio and a short portfolio using the investor sentiment index…**than selecting N+1 stocks**."* [where N=1,2,…,13]	Seeking-alpha	The null hyp. is false for N=1,4, 5,7,8,9,10,11,12 – but not rejected by the hypothesis test at the 10% level	The null hyp. is false for N=4,5, …,13 and **rejected at the 5% level for N=5,7**; for N=2,3, 4,6,8,9,11,12,13 the t-test was n.a.	The null hyp. is false for N=3,5, 7,8,9,10,11,12 and **rejected at the 5% level for N=5,7 and at the 10% level for N=5,7,8,12**
		Blogspot	The null hyp. is false for N=2,3,4, 6,7,8,10,11,12 – but not rejected by the hypothesis test at the 10% level	The null hyp. is false for N=2,3,4,6, 7,8,10,11,12 and **rejected at the 10% level for N=4**; the t-test was applicable for N=4,6	The null hyp. is false for N=1,2,3,4, 7,8,9,10,11,12 and **rejected at the 1% level for N=12 and at the 10% level for N=4,11,12**
		Dataset	$\alpha_{A,N} - \alpha_{B,N} \leq 0$**?**	***mean(pd(excrt.$_{A,N}$, excrt.$_{B,N}$)*** ≤ 0**?**	***median(pd(excrt.$_{A,N}$, excrt.$_{B,N}$)*** ≤ 0**?**
H5	*"**Using the Seekingalpha investor sentiment index (A)** for selecting N stocks in a long portfolio and a short portfolio … **than using the Blogspot investor sentiment index (B)**."* [where N=1,2,…,14]		The null hyp. is false for N=1,2,…,14 – and **rejected** by the hypothesis test **for N=4,5 at the 10% level and for N=5 at the 5% level**	The null hyp. is false for N=1,2, …,14 but not rejected; for N=3,4,5,6,7,10,11, 12,13,14 the t-test was n.a.	The null hyp. is false for N=2,3,…,14 and **rejected at the 5% level for N=5 and at the 10% level for N=4,5,14**
H6	*"**Using the investor sentiment index (A)** for selecting N stocks in a long portfolio and a short portfolio … **than using the stocks' returns (B)**."* [where N=1,2,…,14]	Seeking-alpha	The null hyp. is false for N=2,3, …,14 – but not rejected by the hypothesis test at the 10% level	The null hyp. is false for N=2,3,…,14 but not rejected; the test was applicable only for N=3	The null hyp. is false for N=1,2,…, 14 and **rejected at the 5% level for N=5,12 and at the 10% level for N=4,5,6,7,8,9,11, 12**
		Blogspot	The null hypothesis is false for N=2,6, 7,8,9 – but not rejected by the hypothesis test at the 10% level	The null hyp. is false for N=2,6,8,9, 10,11,12 but not rejected; the test was applicable for N=7,…,11	The null hyp. is false for N=1,2,3,6,7,9,10, 12,13 but not rejected

NOTES: α = the alpha in Carhart's (1997) model; "The null hyp. is false" means that the condition of the null hypothesis was false given the sign of the estimated alpha difference values or mean/median of paired differences of excess returns but the null hypothesis was not necessarily rejected based on statistical evidence; "excrt" = excess returns; "mean" = monthly mean; "median" = monthly median; "pd" = paired differences of two time series as defined in (4.18); "n.a." = not applicable due to violation of (an) assumption(s) of the test.

The test results presented in Table 56 relate to the null hypotheses corresponding to the alternatives H4, H5, and H6, which compare different portfolio simulation results over a range of different numbers of stocks *N* selected in the long and short sub-portfolios. There is only partial statistical evidence for some values of *N* for rejecting either null hypothesis. Still, there is considerable face validity for H4 because the null hypothesis is false for most values of *N* for the Seekingalpha and Blogspot datasets. The partial face validity of H4 suggests that highest abnormal returns are observed in a portfolio simulation with a small number of stocks. For a larger number of stocks, abnormal returns would decline substantially.

Regarding H5, there is also considerable face validity for all *N* (and some statistical evidence for N=4,5,14). Thus, the portfolio simulation using investor sentiment indexes of Seekingalpha should observe higher abnormal returns in general compared to when using investor sentiment indexes of Blogspot. The observed superiority of Seekingalpha might be due to more predictive information, a larger number of reading investors, the classifier of the sentiment orientation not tuned specifically to Blogspot, or partial sparseness of the Blogspot dataset.

Regarding H6.2, there is considerable statistical evidence rejecting the null hypothesis for the Seekingalpha dataset for N=4,5,6,7,8,9,11,12 at the 10% significance level. This evidence suggests that for $N \geq 4$ the information of the investor sentiment index from Seekingalpha results in higher abnormal returns than using the price momentum information for stock selection in the portfolio simulation. Thus, the Seekingalpha-based investor sentiment index can be considered to convey more predictive information towards abnormal returns. That is, it pays for investors to consider the Seekingalpha-based investor sentiment index for stock selection.

5 Conclusions

This section summarizes the research contributions and discusses their practical implications and limitations.

5.1 Research Contributions

This thesis makes three specific contributions to research:

First, this thesis contributes to the behavioral finance literature by advancing the understanding of the effects of investor sentiment from *blog documents* on abnormal returns of a portfolio of stocks. This thesis' study found investor sentiment from blog documents to have a positive effect on abnormal returns of stock portfolios by means of a simulation of a stock selection strategy. The effect is statistically significant when ignoring transaction costs and is largest in magnitude when selecting a small number of stocks in a long and a short sub-portfolio. The effect observed for a long-short portfolio well exceeds the effect of cross-sectional price momentum (in most cases). Statistical evidence for an effect of investor sentiment from blog documents on abnormal returns is largest when selecting stocks only in a long portfolio and prevails also when accounting for transaction costs.

The statistically significant findings specifically relate to (1) blog documents of Seekingalpha, which is host for many (semi-) professional authors, (2) U.S. large capitalization stocks, and (3) monthly aggregates of investor sentiment document scores. This study's monthly aggregates extend on prior studies that mainly used daily measures of investor sentiment (see Section 2.5), and thus provides new findings for the longer term effects of investor sentiment. This study found that highest levels of monthly aggregates of investor sentiment document scores from blog documents have a statistically significant effect on monthly mean/median (portfolio) abnormal returns. This finding corroborates predictions of behavioral finance theory that the effect of investor sentiment can last for long time periods (as discussed in Section 2.2.2). The effects of investor sentiment on large capitalization stocks are not ruled out by behavioral finance theory; however, the effect should be largest for difficult to arbitrage stocks (Baker & Wurgler, 2007, p.149) – such as small capitalization stocks (Baker & Wurgler, 2006). In the light of this thesis' findings, large capitalization stocks are affected substantially as well.

With regard to Seekingalpha, this thesis corroborates results reported by Chen et al. (2014) in the sense that information from Seekingalpha was found to predict stock returns. However, this thesis diverges by having created an explicit investor sentiment classifier with an evaluated accuracy instead of analyzing the number of negative words in documents in relation to abnormal stock returns. Furthermore, this work aggregated investor sentiment document scores into monthly investor sentiment indexes and found them to result in

substantial total portfolio returns in every year of 2007 (13.35%), 2008 (10.99%), 2009 (17.13%), 2010 (9.27%), and 2011 (18.92%) in a long-short stock selection portfolio simulation (ignoring transaction costs). These returns are in contrast to Chen et al.'s (2014) portfolio simulation, which resulted in about <60% portfolio returns in the 2006-2012 period but much smaller portfolio returns of <20% for the three year period 2009–2011. This thesis' result was achieved with a simpler portfolio simulation design of using the current monthly period's investor sentiment indexes for portfolio formation and holding the portfolio over the next monthly period. That is, in contrast to Chen et al. (2014), there were no overlapping holding periods in this thesis' design.

Second, this thesis contributes to better understanding differences between blog platforms as sources of investor sentiment bearing documents. The first source, Seekingalpha, is specialized to investment topics, it targets decision makers, and has (semi-)professional authors, editorial rules, and a large number of readers. The second source, Blogspot, is a large general topic blog platform, which hosts also investment-specific blogs. However, there are no editorial rules and also the number of readers and the impact on decision making is presumably smaller. Whereas Seekingalpha was already subject to research regarding effects on abnormal returns (e.g., Chen et al. (2014), and Fotak (2007)), this thesis contributes to understanding the differences between the two platforms. As discussed above, investor sentiment from blog documents of Seekingalpha relates to abnormal stock returns. In contrast, for the Blogspot blog platform, indications for the same effect seem to also exist but to a smaller extent and with much less statistical evidence.

Third, this thesis contributes to understanding the design of classifiers for the sentiment orientation of investor sentiment from blog documents. A novel classifier with comparably high accuracy of 79.2% was designed, which exceeds the 75.1% reported by O'Hare et al. (2009) on a similar corpus of financial blog documents. The classifier design was informed by the sentiment classification literature and transferred findings from other domains to the financial domain. Furthermore, experiments were conducted to determine the optimal settings for the parameters of the text representation and the SVM *C*-parameter. The classifier allows analyzing textual investor sentiment by an evaluated classifier with estimated accuracy. This is in contrast to prior studies that use a dictionary-based approach with unknown accuracy (e.g., (Chen et al., 2014; Tetlock et al., 2008)). For creating the classifier based on a machine learning approach, a novel corpus of manually annotated blog documents was created. The corpus comprises 638 blog documents from Seekingalpha with a deliberately sampled 50-to-50 percent distribution of positive vs. negative sentiment orientations. This distribution is to help discriminate positive vs. negative vocabulary by providing the same amount of examples. To capture a large variety of vocabulary, different market phases over several years and more than 100 different stocks were covered. The annotations were created by knowledgeable undergraduate students, who achieved a good level of inter-annotator agreement, suggesting a high quality of the corpus.

5.2 Practical Implications

The practical implications of the findings and contributions of this thesis are as follows:

The advantages of investor sentiment from blog documents in comparison to other measures are: Compared to survey-based investor sentiment measures, investor sentiment from blog documents can be collected automatically regarding individual stocks from a large number of blog authors at high frequency, at low cost, and without problems related to questionnaire-based interviews. Compared to market data proxy types of measures of investor sentiment, the benefit of investor sentiment from blog documents is that it is not a proxy but rather a form of a direct measure of investor sentiment, providing potentially information prior to the time when investors act on the market.

The magnitude and statistical significance of the positive abnormal returns found in the Seekingalpha-based simulation (when ignoring transaction costs and partially when accounting for transaction costs) suggests investors to consider investor sentiment indexes based on blog documents from Seekingalpha for their decision making. When accounting for transaction costs, the mean monthly abnormal return (i.e., the estimate of alpha) is still 0.29% (when selecting five stocks in each period in a long portfolio). This result is statistically significant at the 10% level. Due to using only large capitalization stocks in the portfolio simulation, real-world slippage and transaction costs should be comparably low. The design of the monthly period simulation is simple, produces monthly transactions only, and the monthly period decision making may correspond well to fund-type portfolio selection – at least funds are often evaluated on the monthly level (e.g., (Kothari & Warner, 2001)). Whereas the long-short simulation consistently results in positive portfolio returns in every year of the simulation – except the last one, which comprises only one trading month – it might be suitable for hedge funds only. Based on the study's findings, the following practical recommendations can be made when using investor sentiment from blog documents for investment decision making:

A high quality source of investor sentiment such as Seekingalpha should be preferred to other sources and price momentum information.

A long-short stock selection strategy should be used instead of using long or short individually based on the observed magnitude of mean abnormal returns.

A long only stock selection strategy might be used based on the amount of statistical evidence (for non-hedge funds).

A small number of stocks should be selected in each period (based on investor sentiment indexes).

Implementing investment decision making using investor sentiment from blog documents requires time-efficient and cost-efficient classification of the sentiment orientation of a large amount of blog documents being published all the time. To this respect, the machine learning-based classifier offers a solution with comparably high accuracy.

5.3 Limitations

The contributions and practical implications of this thesis must be seen in light of their limitations, which are discussed next.

First, the classifier for the sentiment orientation is constrained to the kind of vocabulary used in Seekingalpha because the training corpus of the classifier contains only blog documents from Seekingalpha. The accuracy for other sources might be lower than for Seekingalpha. Thus, this study's application of the classifier to Blogspot blog documents might be also affected by a lower accuracy.

Second, the size of the corpus, which was used for training the classifier, limits its accuracy. That is, expanding the corpus should lead to (small) increases in the accuracy of the classifier. This expansion might also benefit the results of the portfolio simulation.

Third, the study is constrained by the datasets of investment blog documents, which span a five year period from 2007–2011. In the early years of this time period and especially regarding the Blogspot dataset, there is partial data sparseness. The partial data sparseness might constitute an underlying reason for comparably bad results regarding Blogspot. In further research, the datasets might be expanded to cover also the most recent time periods, other stocks (than DJIA stocks, e.g., non U.S. stocks), and other sources of investment blog documents.

Fourth, regarding the portfolio simulation, there are several limitations: (1) only DJIA stocks were used (i.e., only U.S. large capitalization stocks), (2) only monthly periods were considered (i.e., no daily, or weekly periods), (3) portfolios were held for one period only (i.e., no overlapping periods), and (4) only the level of investor sentiment indexes was used for stock selection (i.e., not the change of the indexes).

Fifth, the abnormal returns detected in terms of alpha are contingent to the used asset pricing model and partially to asymptotic validity of parametric statistical inference methods. Assuming the model (and inference) to be valid, this thesis' results would indicate market inefficiencies. However, due to the joint-hypothesis testing problem (model correctness and market inefficiency), this work does not make this claim. As an alternative to asset pricing model estimation, a benchmark portfolio-based approach was also used. Further non-parametric alternatives to parametric model estimation have remained yet unexplored in this work (e.g., (Chen & Knez, 1996) cited in (Brown & Cliff, 2004, p.21)). As an alternative to parametric inference, non-parametric inference methods were also used in this thesis.

Bibliography

Abarbanell, J. S.; Bushee, B. J. (1997) "Fundamental Analysis, Future Earnings, and Stock Prices". In: *Journal of Accounting Research*. 35 (1), pp. 1–24.

D'Agostino, R. B.; Stephens, M. A. (eds.) (1986) *Goodness-of-Fit Techniques*. New York, NY, USA: Marcel Dekker.

Fellbaum, C. (ed.) (1998) *WordNet: An Electronic Lexical Database*. Cambridge, MA, USA: MIT Press.

Agarwal, N.; Liu, H. (2008) "Blogosphere: Research Issues, Tools, and Applications". In: *ACM SIGKDD Explorations Newsletter*. 10 (1), pp. 18–31.

Akbani, R.; Kwek, S.; Japkowicz, N. (2004) "Applying Support Vector Machines to Imbalanced Datasets". In: *Proceedings of the 15th European Conference on Machine Learning (ECML'04)*. Pisa, Italy, pp. 39–50.

Alcoa (2010) *2009 Annual Report and Form 10-K*. Retrieved 2014-11-12 from URI: http://www.alcoa.com/global/en/investment/pdfs/2009_Annual_Report.pdf

Allen, F.; Carletti, E. (2010) "An Overview of the Crisis: Causes, Consequences, and Solutions". In: *International Review of Finance*. 10 (1), pp. 1–26.

American Association of Individual Investors (2014) "Sentiment Survey". Retrieved 2014-11-30 from URI: http://www.aaii.com/sentimentsurvey

Anderson, T. W.; Darling, D. A. (1952) "Asymptotic Theory of Certain "Goodness of Fit" Criteria Based on Stochastic Processes". In: *The Annals of Mathematical Statistics*. 23 (2), pp. 193–212.

Anderson, T. W.; Darling, D. A. (1954) "A Test of Goodness of Fit". In: *Journal of the American Statistical Association*. 49 (268), pp. 765–769.

Andreevskaia, A.; Bergler, S. (2006) "Mining WordNet for Fuzzy Sentiment: Sentiment Tag Extraction from WordNet Glosses". In: *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL'06)*. Trento, Italy, pp. 209–216.

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

Antweiler, W.; Frank, M. Z. (2006) *Do U.S. Stock Markets Typically Overreact to Corporate News Stories?* Retrieved 2014-11-12 from URI: http://ssrn.com/abstract=878091

Baccianella, S.; Esuli, A.; Sebastiani, F. (2010) "SENTIWORDNET 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining". In: *Proceedings of the 7th Conference on International Language Resources and Evaluation (LREC'10)*. Valletta, Malta, pp. 2200–2204.

Baeza-Yates, R.; Ribeiro-Neto, B. (1999) *Modern Information Retrieval*. Harlow, Essex, UK: Addison Wesley Longman.

Baker, M.; Wurgler, J. (2006) "Investor Sentiment and the Cross-Section of Stock Returns". In: *The Journal of Finance*. 61 (4), pp. 1645–1680.

Baker, M.; Wurgler, J. (2007) "Investor Sentiment in the Stock Market". In: *Journal of Economic Perspectives*. 21 (2), pp. 129–151.

Barber, B. M.; Lyon, J. D. (1997) "Detecting long-run abnormal stock returns: The empirical power and specification of test statistics". In: *Journal of Financial Economics*. 43 (3), pp. 341–372.

Barber, B. M.; Odean, T. (2000) "Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors". In: *The Journal of Finance*. 55 (2), pp. 773–806.

Barber, B. M.; Odean, T.; Zhu, N. (2009) "Systematic noise". In: *Journal of Financial Markets*. 12 (4), pp. 547–569.

Barberis, N.; Shleifer, A.; Vishny, R. W. (1998) "A Model of Investor Sentiment". In: *Journal of Financial Economics*. 49 (3), pp. 307–343.

Bar-Haim, R.; Dinur, E.; Feldman, R.; Fresko, M.; Goldstein, G. (2011) "Identifying and Following Expert Investors in Stock Microblogs". In: *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP'11)*. Edinburgh, UK, pp. 1310–1319.

Barro, R.; Sala-i-Martin, X. (1995) *Economic Growth*. New York, NY, USA: McGraw-Hill.

Bellman, R. E. (1961) *Adaptive control processes: a guided tour*. Princeton, NJ, USA: Princeton University Press.

Berkowitz, S. A.; Logue, D. E.; Noser Jr., E. A. (1988) "The Total Cost of Transactions on the NYSE". In: *The Journal of Finance*. 43 (1), pp. 97–112.

Bernard, V. L.; Thomas, J. K. (1989) "Post-Earnings-Announcement Drift: Delayed Price Response or Risk Premium?". In: *Journal of Accounting Research*. 27, pp. 1–36.

Biber, D. (1993) "Representativeness in Corpus Design". In: *Literary and Linguistic Computing*. 8 (4), pp. 243–257.

Black, F. (1986) "Noise". In: *The Journal of Finance*. 41 (3), pp. 529–543.

Bodie, Z.; Kane, A.; Marcus, A. J. (2009) *Investments*. 8th in. ed., New York, NY, USA: McGraw-Hill Irwin.

Bollen, J.; Mao, H.; Zeng, X. (2011) "Twitter mood predicts the stock market". In: *Journal of Computational Science*. 2 (1), pp. 1–8.

De Bondt, W. F. M. (1993) "Betting on trends: Intuitive forecasts of financial risk and return". In: *International Journal of Forecasting*. 9 (3), pp. 355–371.

De Bondt, W. F. M. (1998) "A portrait of the individual investor". In: *European Economic Review*. 42 (3), pp. 831–844.

Boser, B. E.; Guyon, I. M.; Vapnik, V. N. (1992) "A Training Algorithm for Optimal Margin Classifiers". In: *Proceedings of the 5th Annual Workshop on Computational Learning Theory (COLT'92)*. Pittsburgh, PA, USA, pp. 144–152.

Brealey, R. A.; Myers, S. C.; Allen, F. (2011) *Principles of Corporate Finance*. 10th gl. ed., New York, NY, USA: McGraw-Hill Irwin.

Brown, G. W.; Cliff, M. T. (2004) "Investor sentiment and the near-term stock market". In: *Journal of Empirical Finance*. 11 (1), pp. 1–27.

Brown, G. W.; Cliff, M. T. (2005) "Investor Sentiment and Asset Valuation". In: *The Journal of Business*. 78 (2), pp. 405–440.

Brown, S. J.; Warner, J. B. (1980) "Measuring Security Price Performance". In: *Journal of Financial Economics*. 8 (3), pp. 205–258.

Brown, S. J.; Warner, J. B. (1985) "Using Daily Stock Returns: The Case of Event Studies". In: *Journal of Financial Economics*. 14 (1), pp. 3–31.

Burghardt, M. (2010) "Retail Investor Sentiment and Behavior – an Empirical Analysis". PhD Thesis. Karlsruhe Institute of Technology.

Campbell, J. Y.; Lo, A. W.; MacKinlay, A. C. (1997) "Event Study Analysis". In: *The Econometrics of Financial Markets*. Princeton, NJ, USA: Princeton University Press, pp. 149–180.

Carhart, M. (1997) "On Persistence in Mutual Fund Performance". In: *The Journal of Finance*. 52 (1), pp. 57–82.

Chang, K.-W.; Hsieh, C.-J.; Lin, C.-J. (2008) "Coordinate Descent Method for Large-scale L2-loss Linear Support Vector Machines". In: *Journal of Machine Learning Research*. 9, pp. 1369–1398.

Chen, H.; De, P.; Hu, Y. J.; Hwang, B. H. (2014) "Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media". In: *Review of Financial Studies*. 27 (5), pp. 1–37.

Chen, Z.; Knez, P. J. (1996) "Portfolio performance measurement: theory and applications". In: *Review of Financial Studies*. 9, pp. 511–555.

Chordia, T.; Shivakumar, L. (2002) "Momentum, Business Cycle, and Time-Varying Expected Return". In: *The Journal of Finance*. 57 (2), pp. 985–1019.

Clark, D. A. (2013) "5 Stocks Under $10 To Buy With Solid Fundamentals And Upside". Retrieved 2014-04-23 from URI: http://seekingalpha.com/article/1411551-5-stocks-under-10-to-buy-with-solid-fundamentals-and-upside

Coca-Cola Company (2010) *Fiscal Year 2009 Form 10-K*. Retrieved 2014-11-12 from URI: http://www.sec.gov/Archives/edgar/data/21344/000104746910001476/a2195739z10-k.htm

Cochrane, J. H. (2005) *Asset Pricing*. Revised ed., Princeton, NJ, USA: Princeton University Press.

Cohen, J. (1960) "A Coefficient of Agreement for Nominal Scales". In: *Educational and Psychological Measurement*. 20 (1), pp. 37–46.

Cohen, J. (1968) "Weighted Kappa: Nominal Scale Agreement with Provision for Scales Disagreement or Partial Credit". In: *Psychological Bulletin*. 70 (4), pp. 213–220.

Cortes, C.; Vapnik, V. (1995) "Support-Vector Networks". In: *Machine Learning*. 20 (3), pp. 273–297.

Covel, M. W. (2004) *Trend Following*. London, UK: Financial Times Prentice Hall.

Cover, T. M. (1965) "Geometrical and Statistical Properties of Systems of Linear Inequalities with Applications in Pattern Recognition". In: *IEEE Transactions on Electronic Computers*. 14 (3), pp. 326–334.

Cunningham, H.; Maynard, D.; Bontcheva, K.; Tablan, V.; Aswani, N.; Roberts, I.; Gorrell, G.; Funk, A.; Roberts, A.; Damljanovic, D.; Heitz, T.; Greenwood, M. A.; Saggion, H.; Petrak, J.; Li, Y.; Peters, W. (2011) *Developing Language Processing Components with GATE Version 6 (a User Guide)*. University of Sheffield. Retrieved 2014-11-12 from URI: http://gate.ac.uk/releases/gate-6.1-build3913-ALL/tao.pdf

Cunningham, P. (2007) *Dimension Reduction*. Report, Nr. UCD-CSI-2007-7. University College Dublin. Retrieved 2014-11-13 from URI: https://www.csi.ucd.ie/files/UCD-CSI-2007-7.pdf

Daniel, K.; Grinblatt, M.; Titman, S.; Wermers, R. (1997) "Measuring Mutual Fund Performance with Characteristics-Based Benchmarks". In: *The Journal of Finance*. 52 (3), pp. 1035–1058.

Daniel, K.; Hirshleifer, D.; Subrahmanyam, A. (1998) "Investor Psychology and Security Market Under- and Overreactions". In: *The Journal of Finance*. 53 (6), pp. 1839–1885.

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

Downing, D.; Clark, J. (2010) *Business Statistics*. 5th ed., Hauppauge, NY, USA: Barron's Educational Series.

Dubin, D. (2004) "The Most Influential Paper Gerard Salton Never Wrote". In: *Library Trends*. 52 (4), pp. 748–764.

Dumais, S.; Platt, J.; Heckerman, D.; Sahami, M. (1998) "Inductive learning algorithms and representations for text categorization". In: *Proceedings of the 7th international Conference on Information and Knowledge Management (CIKM '98)*. Bethesda, MD, USA, pp. 148–155.

Edgecliffe, A. (1999) "eToys Surges after Listing". *Financial Times*. *1999-05-21*, p. 29.

Edwards, W. (1968) "Conservatism in human information processing". In: Kleinmuntz, B. (ed.) *Formal Representation of Human Judgment*. New York, NY, USA: John Wiley & Sons, pp. 17–52.

Eling, M.; Schuhmacher, F. (2007) "Does the choice of performance measure influence the evaluation of hedge funds?". In: *Journal of Banking and Finance*. 31 (9), pp. 2632–2647.

Elton, E. J.; Gruber, M. J. (1997) "Modern portfolio theory, 1950 to date". In: *Journal of Banking & Finance*. 21 (11–12), pp. 1743–1759.

Esuli, A.; Sebastiani, F. (2006) "SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining". In: *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC'06)*. Genoa, Italy, pp. 417–422.

Fama, E. F. (1965a) "The Behavior of Stock Market Prices". In: *The Journal of Business*. 38 (1), pp. 34–105.

Fama, E. F. (1965b) "Random Walks in Stock Market Prices". In: *Financial Analysts Journal*. 21 (5), pp. 55–59.

Fama, E. F. (1970) "Efficient Capital Markets: A Review of Theory and Empirical Work". In: *The Journal of Finance*. 25 (2), pp. 383–417.

Fama, E. F. (1991) "Efficient Capital Markets: II". In: *The Journal of Finance*. 46 (5), pp. 1575–1617.

Fama, E. F. (1998) "Market Efficiency, Long-Term Returns, and Behavioral Finance". In: *Journal of Financial Economics*. 49 (3), pp. 283–306.

Fama, E. F.; French, K. R. (1988) "Permanent and Temporary Components of Stock Prices". In: *Journal of Political Economy*. 96 (2), pp. 246–273.

Fama, E. F.; French, K. R. (1992) "The Cross-Section of Expected Stock Returns". In: *The Journal of Finance*. 47 (2), pp. 427–465.

Fama, E. F.; French, K. R. (1993) "Common Risk Factors in the Returns on Stocks and Bonds". In: *Journal of Financial Economics*. 33 (1), pp. 3–56.

Fama, E. F.; French, K. R. (1996) "Multifactor Explanations of Asset Pricing Anomalies". In: *The Journal of Finance*. 51 (1), pp. 55–84.

Fama, E. F.; French, K. R. (2004) "The Capital Asset Pricing Model: Theory and Evidence". In: *Journal of Economic Perspectives*. 18 (3), pp. 25–46.

Feibel, B. J. (2003) *Investment Performance Measurement*. Hoboken, NJ, USA: John Wiley & Sons.

Fisher, I. (1974) *The Theory of Interest: As Determined by Impatience to Spend Income and Opportunity to Invest It*. 1934 ed., Clifton, Bristol, UK: Augustus M. Kelley (Reprint).

Fisher, K. L.; Statman, M. (2000) "Investor Sentiment and Stock Returns". In: *Financial Analysts Journal*. 56 (2), pp. 16–23.

Fleiss, J. L. (1971) "Measuring Nominal Scale Agreement Among Many Raters". In: *Psychological Bulletin*. 76 (5), pp. 378–382.

Fleiss, J. L.; Levin, B.; Paik, M. C. (2003) "The Measurement of Interrater Agreement". In: Balding, D. J.; Cressie, N. A. C.; Fisher, N. I.; Johnstone, I. M.; Kadane, J. B.; Ryan, L. M.; Scott, D. W.; Smith, A. F. M.; Teugels, J. L.; Barnett, V.; Hunter, J. S.; Kendall, D. G. (eds.) *Statistical Methods for Rates and Proportions*. 3rd ed., Hoboken, NJ, USA: John Wiley & Sons, pp. 598–626.

Forman, G. (2003) "An Extensive Empirical Study of Feature Selection Metrics for Text Classification". In: *Journal of Machine Learning Research*. 3, pp. 1289–1305.

Fotak, V. (2007) "The Impact of Blog Recommendations on Security Prices and Trading Volumes". Retrieved 2014-11-12 from URI: http://ssrn.com/abstract=1089868

Fung, W.; Hsieh, D. A. (2013) "Hedge Funds". In: Constantinides, G. M.; Harris, M.; Stulz, R. M. (eds.) *Handbook of the Economics of Finance, Volume 2, Part B*. Amsterdam, The Netherlands: Elsevier, pp. 1063–1125.

Fürnkranz, J. (1998) *A Study Using n-gram Features for Text Categorization*. Report, Nr. OEFAI-TR-98-30. Austrian Research Institute for Artificial Intelligence. Retrieved 2014-11-12 from URI: http://www.ofai.at/cgi-bin/get-tr?download=1&paper=oefai-tr-98-30.pdf

General Inquirer (2014) "Descriptions of Inquirer Categories and Use of Inquirer Dictionaries". Retrieved 2014-12-15 from URI: http://www.wjh.harvard.edu/~inquirer/homecat.htm

Gibbons, J. D.; Chakraborti, S. (2011) *Nonparametric Statistical Inference*. 5th ed., Boca Raton, FL, USA: Chapman & Hall/CRC.

Gilbert, E.; Karahalios, K. (2010) "Widespread Worry and the Stock Market". In: *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media (ICWSM'10)*. Washington, DC, USA, pp. 58–65.

Godbole, N.; Srinivasaiah, M.; Skiena, S. (2007) "Large-Scale Sentiment Analysis for News and Blogs". In: *Proceedings of the International Conference on Weblogs and Social Media (ICWSM'07)*. Boulder, CO, USA, pp. 219–222.

Granger, C. W. J. (1969) "Investigating Causal Relations by Econometric Models and Cross-spectral Methods". In: *Econometrica*. 37 (3), pp. 424–438.

Grinold, R. C.; Kahn, R. N. (2000) *Active Portfolio Management: a Quantitative Approach for Providing Superior Returns and Controlling Risk*. 2nd ed., New York, NY, USA: McGraw-Hill.

Grossman, S. J.; Stiglitz, J. E. (1980) "On the Impossibility of Informationally Efficient Markets". In: *The American Economic Review*. 70 (3), pp. 393–408.

Groth, S. S.; Muntermann, J. (2011) "An intraday market risk management approach based on textual analysis". In: *Decision Support Systems*. Elsevier B.V. 50 (4), pp. 680–691.

Groves, R. M. (2004) *Survey Errors and Survey Costs*. Hoboken, NJ, USA: John Wiley & Sons.

Gujarati, D. (1970a) "Use of Dummy Variables in Testing for Equality Between Sets of Coefficients in Two Linear Regressions: A Note". In: *The American Statistican*. 24 (1), pp. 50–52.

Gujarati, D. (1970b) "Use of Dummy Variables in Testing for Equality Between Sets of Coefficients in Linear Regressions: A Generalization". In: *The American Statistician*. 24 (5), pp. 18–22.

Gujarati, D. N.; Porter, D. C. (2009) *Basic Econometrics*. 5th in. ed., New York, NY, USA: McGraw-Hill.

Hastie, T.; Tibshirani, R.; Friedman, J. (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed., New York, NY, USA: Springer.

Hatzivassiloglou, V.; McKeown, K. R. (1997) "Predicting the Semantic Orientation of Adjectives". In: *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics (ACL/EACL'97)*. Madrid, Spain, pp. 174–181.

Henry, G. T. (1990) *Practical Sampling*. Newbury Park, CA, USA: Sage.

Hepple, M. (2000) "Independence and Commitment: Assumptions for Rapid Training and Execution of Rule-based POS Taggers". In: *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics (ACL'00)*. Hong Kong, China, pp. 278–285.

Herring, S. C.; Scheidt, L. A.; Bonus, S.; Wright, E. (2004) "Bridging the Gap: A Genre Analysis of Weblogs". In: *Proceedings of the 37th Hawaii International Conference on System Sciences (HICSS'04)*. Big Island, HI, USA, pp. 1–11.

Hirshleifer, D. (2001) "Investor Psychology and Asset Pricing". In: *The Journal of Finance*. 56 (4), pp. 1533–1597.

Hohlfeld, R.; Dörsam, S. (2008) "Börse im Blog. Eine Analyse medienintegrierter und unabhängiger Finanz-Weblogs". In: Quandt, T.; Schweiger, W. (eds.) *Journalismus online - Partizipation oder Profession?*. Wiesbaden, Germany: VS Verlag für Sozialwissenschaften, pp. 95–110.

Home Depot (2010) *Fiscal Year 2009 Form 10-K*. Retrieved 2014-11-13 from URI: http://www.sec.gov/Archives/edgar/data/354950/000119312510067178/d10k.htm

Hong, H.; Lim, T.; Stein, J. C. (2000) "Bad News Travels Slowly: Size, Analyst Coverage, and the Profitability of Momentum Strategies". In: *The Journal of Finance*. 55 (1), pp. 265–295.

Hossain, M. Z.; Bhatti, M. I. (2003) "Recent Development in Econometric Analysis of Model Selection". In: *Managerial Science*. 29 (7), pp. 90–108.

Hsieh, C.-J.; Chang, K.-W.; Lin, C.-J.; Keerthi, S. S.; Sundararajan, S. (2008) "A Dual Coordinate Descent Method for Large-scale Linear SVM". In: *Proceedings of the 25th International Conference on Machine Learning (ICML'08)*. Helsinki, Finland, pp. 408–415.

Hsu, C.; Chang, C.; Lin, C. (2010) *A practical guide to support vector classification*. National Taiwan University. Retrieved 2013-09-09 from URI: http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf

Hu, M.; Liu, B. (2004) "Mining and Summarizing Customer Reviews". In: *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'04)*. Seattle, WA, USA, pp. 168–177.

Investors Intelligence (2014a) "Investors Intelligence Sentiment Poll". Retrieved 2014-11-30 from URI: http://www.investorsintelligence.com/content/20784/sentiment_data__quick_explanation.pdf

Investors Intelligence (2014b) "US Advisors' Sentiment Report - Signals when you need them - near important market tops and bottoms". Retrieved 2014-11-30 from URI: http://www.investorsintelligence.com/x/us_advisors_sentiment.html

Jaffe, J. F. (1974) "The Effect of Regulation on Insider Trading". In: *The Bell Journal of Economics and Management Science*. 5 (1), pp. 93–121.

Jarque, C. M.; Bera, A. K. (1987) "A Test for Normality of Observations and Regression Residuals". In: *International Statistical Review*. 55 (2), pp. 163–172.

Jegadeesh, N.; Titman, S. (1993) "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency". In: *The Journal of Finance*. 48 (1), pp. 65–91.

Jegadeesh, N.; Titman, S. (2001) "Profitability of Momentum Strategies: An Evaluation of Alternative Explanations". In: *The Journal of Finance*. 56 (2), pp. 699–720.

Jegadeesh, N.; Titman, S. (2011) "Momentum". In: *Annual Review of Financial Economics*. 3 (1), pp. 493–509.

Jensen, M. C. (1968) "The performance of mutual funds in the period 1945-1964". In: *The Journal of Finance*. 23 (2), pp. 389–416.

Joachims, T. (1998) "Text Categorization with Support Vector Machines: Learning with Many Relevant Features". In: *Proceedings of the 10th European Conference on Machine Learning (ECML'98)*. Chemnitz, Germany, pp. 137–142.

Joachims, T. (2002) *Learning to Classify Text Using Support Vector Machines*. Norwell, MA, USA: Kluwer Academic Publishers.

Joachims, T. (2006) "Training Linear SVMs in Linear Time". In: *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'06)*. Philadelphia, PA, USA, pp. 217–226.

Kaplan, A. M.; Haenlein, M. (2010) "Users of the world, unite! The challenges and opportunities of Social Media". In: *Business Horizons*. 53 (1), pp. 59–68.

Keim, D. B.; Madhavan, A. (1998) "The Cost of Institutional Equity Trades". In: *Financial Analysts Journal*. 54 (4), pp. 50–69.

Khalid (2011) "How Big is Blogosphere [Infographic]". Retrieved 2012-04-12 from URI: http://www.invesp.com/blog/business/how-big-is-blogsphere.html

Klein, A.; Altuntas, O.; Haeusser, T.; Kessler, W. (2011) "Extracting Investor Sentiment from Weblog Texts: A Knowledge-based Approach". In: *Proceedings of the 13th IEEE International Conference on Commerce and Enterprise Computing (CEC'11)*. Luxembourg, Luxembourg, pp. 1–9.

Klein, A.; Altuntas, O.; Riekert, M.; Dinev, V. (2013) "A Combined Approach for Extracting Financial Instrument-specific Investor Sentiment from Weblogs". In: *Proceedings of the 11th International Conference on Wirtschaftsinformatik (WI'13)*. Leipzig, Germany, pp. 691–705.

Kohavi, R. (1995) "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection". In: *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI'95)*. Montreal, QC, Canada, pp. 1137–1143.

Kothari, S. P.; Warner, J. B. (2001) "Evaluating Mutual Fund Performance". In: *The Journal of Finance*. 56 (5), pp. 1985–2010.

Kothari, S. P.; Warner, J. B. (2007) "Econometrics of Event Studies". In: Eckbo, B. E. (ed.) *Handbook of Corporate Finance: Empirical Corporate Finance (Volume 1)*. Amsterdam, The Netherlands: Elsevier, pp. 4–36.

Krishnamurthy, S. (2002) "The Multidimensionality of Blog Conversations: The Virtual Enactment of September 11". Maastricht, The Netherlands: Internet Research 3.0.

Kumar, A.; Lee, C. M. C. (2006) "Retail Investor Sentiment and Return Comovements". In: *The Journal of Finance*. 61 (5), pp. 2451–2486.

Kyle, A. S. (1985) "Continuous Auctions and Insider Trading". In: *Econometrica*. 53 (6), pp. 1315–1336.

Langer, E. J.; Roth, J. (1975) "Heads I win, tails it's chance: The illusion of control as a function of the sequence of outcomes in a purely chance task". In: *Journal of Personality and Social Psychology*. 32 (6), pp. 951–955.

Lee, C. M. C.; Shleifer, A.; Thaler, R. H. (1991) "Investor Sentiment and the Closed-End Fund Puzzle". In: *The Journal of Finance*. 46 (1), pp. 75–109.

Leinweber, D.; Sisk, J. (2011) "Relating news analytics to stock returns". In: Mitra, G.; Mitra, L. (eds.) *The Handbook of News Analytics in Finance*. 1st ed., Chichester, West Sussex, UK: John Wiley & Sons, pp. 149–172.

Leopold, E.; Kindermann, J. (2002) "Text Categorization with Support Vector Machines. How to Represent Texts in Input Space?". In: *Machine Learning*. 46 (1–3), pp. 423–444.

Lerman, K.; Blair-Goldensohn, S.; McDonald, R. (2009) "Sentiment summarization: Evaluating and learning user preferences". In: *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL'09)*. Athens, Greece, pp. 514–522.

Lintner, J. (1965) "Security Prices, Risk, and Maximal Gains From Diversification". In: *The Journal of Finance*. 20 (4), pp. 587–615.

Liu, B. (2010) "Sentiment Analysis and Subjectivity". In: Indurkhya, N.; Damerau, F. J. (eds.) *Handbook of Natural Language Processing*. 2nd ed., Boca Raton, FL, USA: Chapman and Hall/CRC Press, pp. 627–666.

Liu, H.; Singh, P. (2004) "ConceptNet - A Practical Commonsense Reasoning Tool-Kit". In: *BT Technology Journal*. 22 (4), pp. 211–226.

Livejournal (2015) "Livejournal". Retrieved 2015-02-18 from URI: http://www.livejournal.com/

De Long, J. B.; Shleifer, A.; Summers, L. H.; Waldmann, R. J. (1990a) "Positive Feedback Investment Strategies and Destabilizing Rational Speculation". In: *The Journal of Finance*. 45 (2), pp. 379–395.

De Long, J. B.; Shleifer, A.; Summers, L. H.; Waldmann, R. J. (1990b) "Noise Trader Risk in Financial Markets". In: *Journal of Political Economy*. 98 (4), pp. 703–738.

De Long, J. B.; Shleifer, A.; Summers, L. H.; Waldmann, R. J. (1991) "The Survival of Noise Traders in Financial Markets". In: *The Journal of Business*. 64 (1), pp. 1–19.

Loughran, T.; McDonald, B. (2011) "When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-Ks". In: *The Journal of Finance*. 66 (1), pp. 67–97.

Loughran, T.; McDonald, B. (2014) "Loughran and McDonald Financial Sentiment Dictionaries". Retrieved 2014-12-15 from URI: http://www3.nd.edu/~mcdonald/Word_Lists.html

Loughran, T.; Ritter, J. R. (1995) "The New Issues Puzzle". In: *The Journal of Finance*. 50 (1), pp. 23–51.

Luhn, H. P. (1957) "A Statistical Approach to Mechanized Encoding and Searching of Literary Information". In: *IBM Journal of Research and Development*. 1 (4), pp. 309–317.

Lyon, J. D.; Barber, B. M.; Tsai, C.-L. (1999) "Improved Methods for Tests of Long-Run Abnormal Stock Returns". In: *The Journal of Finance*. 54 (1), pp. 165–201.

Manning, C. D.; Raghavan, P.; Schütze, H. (2009) *Introduction to Information Retrieval*. Online ed., New York, NY, USA: Cambridge University Press. Retrieved 2014-11-13 from URI: http://nlp.stanford.edu/IR-book/

Markowitz, H. M. (1952) "Portfolio Selection". In: *The Journal of Finance*. 7 (1), pp. 77–91.

Markowitz, H. M. (1959) *Portfolio Selection - Efficient Diversification of Investments*. New York, NY, USA: John Wiley & Sons.

McCallum, A.; Nigam, K. (1998) "A Comparison of Event Models for Naive Bayes Text Classification". In: *Proceedings of the 15th National Conference on Artificial Intelligence Workshop on Learning for Text Categorization (AAAI'98)*. Madison, WI, USA

McIntyre, D. A.; Allen, A. C. (2009) "Best 25 Financial Blogs". Retrieved 2013-07-24 from URI: http://www.time.com/time/business/article/0,8599,1873144-1,00.html

Mejova, Y.; Srinivasan, P. (2011) "Exploring Feature Definition and Selection for Sentiment Classifiers". In: *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media (ICWSM'11)*. Barcelona, Spain, pp. 546–549.

Melville, P.; Gryc, W.; Lawrence, R. D. (2009) "Sentiment analysis of blogs by combining lexical knowledge with text classification". In: *Proceedings of the 15th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'09)*. Paris, France, pp. 1275–1283.

Michaud, R. O. (1993) "Are Long-Short Equity Strategies Superior?". In: *Financial Analysts Journal*. 49 (6), pp. 44–49.

Miller, D. T.; Ross, M. (1975) "Self-serving bias in attribution of causality: Fact or fiction?". In: *Psychological Bulletin*. 82 (2), pp. 213–225.

Mishne, G. (2007) "Applied Text Analytics for Blogs". PhD Thesis. University of Amsterdam.

Missen, M. M. S.; Boughanem, M.; Cabanac, G. (2013) "Opinion mining: reviewed from word to document level". In: *Social Network Analysis and Mining*. 3 (1), pp. 107–125.

Mitchell, M. L.; Stafford, E. (2000) "Managerial Decisions and Long-Term Stock Price Performance". In: *The Journal of Business*. 73 (3), pp. 287–329.

Mitchell, T. M. (1997) *Machine Learning*. Boston, MA, USA: WCB/McGraw-Hill.

Mitra, L.; Mitra, G. (2011) "Applications of news analytics in finance: A review". In: Mitra, G.; Mitra, L. (eds.) *The Handbook of News Analytics in Finance*. Hoboken, NJ, USA: John Wiley & Sons, pp. 1–40.

Mladenic, D.; Grobelnik, M. (1998) "Word sequences as features in text-learning". In: *Proceedings of the 17th Electrotechnical and Computer Science Conference (ERK'98)*. Ljubljana, Slovenia, pp. 145–148.

Modigliani, F.; Modigliani, L. (1997) "Risk-adjusted performance – how to measure it and why". In: *Journal of Portfolio Management*. 23 (2), pp. 45–54.

Moskowitz, T. J.; Ooi, Y. H.; Pedersen, L. H. (2012) "Time series momentum". In: *Journal of Financial Economics*. 104 (2), pp. 228–250.

Mossin, J. (1966) "Equilibrium in a Capital Asset Market". In: *Econometrica*. 34 (4), pp. 768–783.

Mullen, T.; Collier, N. (2004) "Sentiment analysis using support vector machines with diverse information sources". In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP'04)*. Barcelona, Spain, pp. 412–418.

Murphy, J. J. (1999) *Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications*. Paramus, NJ, USA: New York Institute of Finance.

Murugesan, S. (2007) "Understanding Web 2.0". In: *IT Professional*. 9 (4), pp. 34–41.

Nardi, B. A.; Schiano, D. J.; Gumbrecht, M.; Swartz, L. (2004) "Why We Blog". In: *Communications of the ACM*. 47 (12), pp. 41–46.

Nasukawa, T.; Yi, J. (2003) "Sentiment Analysis: Capturing Favorability Using Natural Language Processing Definition of Sentiment Expressions". In: *Proceedings of the 2nd International Conference on Knowledge Capture (K-CAP'03)*. Sanibel Island, Florida, USA, pp. 70–77.

Newey, W. K.; West, K. D. (1987) "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix". In: *Econometrica*. 55 (3), pp. 703–708.

Ng, V.; Dasgupta, S.; Arifin, S. M. N. (2006) "Examining the Role of Linguistic Knowledge Sources in the Automatic Identification and Classification of Reviews". In: *Proceedings of the Joint Conference of the International Committee on Computational Linguistics and the Association for Computational Linguistics (COLING/ACL'06)*. Sydney, Australia, pp. 611–618.

Nielsen Company (2007) "Word of Mouth, the Most Powerful Selling Tool: Nielsen Global Survey". Retrieved 2013-07-13 from URI: http://www.nielsen.com/us/en/press-room/2007/Word-of-Mouth_the_Most_Powerful_Selling_Tool__Nielsen_Global_Survey.html

NM Incite (2012) "Buzz in the Blogosphere: Millions more bloggers and blog readers". Retrieved 2015-04-28 from URI: http://www.nielsen.com/us/en/insights/news/2012/buzz-in-the-blogosphere-millions-more-bloggers-and-blog-readers.html

O'Hare, N.; Davy, M.; Bermingham, A.; Ferguson, P.; Sheridan, P.; Gurrin, C.; Smeaton, A. F. (2009) "Topic-Dependent Sentiment Analysis Of Financial Blogs". In: *Proceedings of the 1st International CIKM Workshop on Topic-sentiment Analysis for Mass Opinion (TSA'09)*. Hong Kong, China, pp. 9–16.

Odean, T. (1998) "Volume, Volatility, Price, and Profit When All Traders Are Above Average". In: *The Journal of Finance*. 53 (6), pp. 1887–1934.

Oh, C.; Sheng, O. R. L. (2011) "Investigating Precitive Power of Stock Micro Blog Sentiment in Forecasting Future Stock Price Directional Movement". In: *Proceedings of 32nd International Conference on Information Systems (ICIS'11)*. Shanghai, China

Ou, J. A.; Penman, S. H. (1989) "Financial Statement Analysis and the Prediction of Stock Returns". In: *Journal of Accounting and Economics*. 11 (4), pp. 295–329.

Pang, B.; Lee, L. (2004) "A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts". In: *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL'04)*. Barcelona, Spain, pp. 271–278.

Pang, B.; Lee, L. (2008) "Opinion Mining and Sentiment Analysis". In: *Foundations and Trends in Information Retrieval*. 2 (1–2), pp. 1–135.

Pang, B.; Lee, L.; Vaithyanathan, S. (2002) "Thumbs up? Sentiment Classification using Machine Learning Techniques". In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP'02)*. Philadelphia, PA, USA, pp. 79–86.

Penman, S. H. (2013) *Financial Statement Analysis and Security Valuation*. 5th in. ed., New York, NY, USA: McGraw-Hill.

Poitras, G. (2010) "Arbitrage: Historical Perspectives". In: Cont, R. (ed.) *Encyclopedia of Quantitative Finance*. Hoboken, NJ, USA: John Wiley & Sons

Pontiff, J. (1996) "Costly Arbitrage: Evidence from Closed-End Funds". In: *The Quarterly Journal of Economics*. 111 (4), pp. 1135–1151.

Poterba, J. M.; Summers, L. H. (1988) "Mean Reversion in Stock Prices". In: *Journal of Financial Economics*. 22 (1), pp. 27–59.

Princeton University (2014) "WordNet A lexical database for English". Retrieved 2014-12-15 from URI: http://wordnet.princeton.edu/

Quantcast (2013) "Number of Visitors of seekingalpha.com". Retrieved 2013-07-24 from URI: https://www.quantcast.com/seekingalpha.com

Raghavan, V. V.; Wong, S. K. M. (1986) "A Critical Analysis of Vector Space Model for Information Retrieval". In: *Journal of the American Society for Information Science*. 37 (5), pp. 279–287.

Rajan, R.; Servaes, H. (1997) "Analyst Following of Initital Public Offerings". In: *The Journal of Finance*. 52 (2), pp. 507–529.

S&P Dow Jones Indices LLC (2013) "Dow Jones Industrial Average - Historical Components". Retrieved 2014-04-29 from URI: http://www.djindexes.com/mdsidx/downloads/brochure_info/Dow_Jones_Industrial_Average_Historical_Components.pdf

S&P Dow Jones Indices LLC (2014) "Dow Jones Industrial Average - Overview". Retrieved 2014-05-16 from URI: https://www.djaverages.com/?go=industrial-overview

Salton, G. (1979) "Mathematics and Information Retrieval". In: *Journal of Documentation*. 35 (1), pp. 1–29.

Salton, G. (1989) *Automatic text processing: The transformation, analysis, and retrieval of information by computer*. Boston, MA, USA: Addison-Wesley Longman.

Salton, G.; Buckley, C. (1988) "Term-weighting approaches in automatic text retrieval". In: *Information Processing & Management*. 24 (5), pp. 513–523.

Salton, G.; McGill, M. J. (1983) *Introduction to Modern Information Retrieval*. New York, NY, USA: McGraw-Hill.

Salton, G.; Wong, A.; Yang, C. S. (1975) "A Vector Space Model for Automatic Indexing". In: *Communications of the ACM*. 18 (11), pp. 613–620.

Salton, G.; Yang, C. S. (1973) "On the Specification of Term Values in Automatic Indexing". In: *Journal of Documentation*. 29 (4), pp. 351–372.

Schmidt, J. (2007) "Blogging practices: An analytical framework". In: *Journal of Computer-Mediated Communication*. 12 (4), pp. 1409–1427.

Schölkopf, B.; Smola, A. J. (2002) *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press.

Schumaker, R. P.; Zhang, Y.; Huang, C.-N.; Chen, H. (2012) "Evaluating Sentiment in Financial News Articles". In: *Decision Support Systems*. 53 (3), pp. 1–25.

Sebastiani, F. (2002) "Machine Learning in Automated Text Categorization". In: *ACM Computing Surveys*. 34 (1), pp. 1–47.

Seekingalpha (2014a) "Seeking Alpha's History and Founder". Retrieved 2014-12-03 from URI: http://seekingalpha.com/page/history_and_founder

Seekingalpha (2014b) "Article Stats". Retrieved 2014-12-03 from URI: http://seekingalpha.com/listing/articles_stats

Seekingalpha (2014c) "Contributor Stats". Retrieved 2014-12-03 from URI: http://seekingalpha.com/listing/contributors_stats

Seekingalpha (2014d) "Article Submission Guidelines". Retrieved 2014-12-03 from URI: http://seekingalpha.com/page/article-submission-guidelines

Seekingalpha (2014e) "Who Reads Seeking Alpha?". Retrieved 2014-12-03 from URI: http://seekingalpha.com/page/who_reads_sa

Seekingalpha (2014f) "Become a Seeking Alpha Contributor". Retrieved 2014-12-03 from URI: http://seekingalpha.com/page/become-a-seeking-alpha-contributor

SentiWordNet (2014) "SentiWordNet". Retrieved 2014-12-15 from URI: http://sentiwordnet.isti.cnr.it/

Shaikh, M. A. M.; Prendinger, H.; Ishizuka, M. (2008) "Sentiment Assessment of Text By Analyzing Linguistic Features and Contextual Valence Assignment". In: *Applied Artificial Intelligence*. 22 (6), pp. 558–601.

Sharpe, W. F. (1963) "A Simplified Model for Portfolio Analysis". In: *Management Science*. 9 (2), pp. 277–293.

Sharpe, W. F. (1964) "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk". In: *The Journal of Finance*. 19 (3), pp. 425–442.

Sharpe, W. F. (1966) "Mutual Fund Performance". In: *The Journal of Business*. 39 (1), pp. 119–138.

Sharpe, W. F.; Alexander, G. J.; Bailey, J. V. (1998) *Investments*. 6th ed., Upper Saddle River, NJ, USA: Prentice Hall.

Shefrin, H. (2008) "Introduction". In: *A Behavioral Approach to Asset Pricing*. 2nd ed., Burlington, MA, USA: Academic Press, pp. 1–13.

Shiller, R. J. (2000) *Irrational Exuberance*. Princeton, NJ, USA: Princeton University Press.

Shiller, R. J. (2003) "From Efficient Markets Theory to Behavioral Finance". In: *Journal of Economic Perspectives*. 17 (1), pp. 83–104.

Shleifer, A. (2000) *Inefficient Markets - An Introduction to Behavioral Finance*. Oxford, Oxfordshire, UK: Oxford University Press.

Shleifer, A.; Summers, L. H. (1990) "The Noise Trader Approach to Finance". In: *Journal of Economic Perspectives*. 4 (2), pp. 19–33.

Shleifer, A.; Vishny, R. W. (1997) "The Limits of Arbitrage". In: *The Journal of Finance*. 52 (1), pp. 35–55.

Siegel, J. J. (1992) "Equity Risk Premia, Corporate Profit Forecasts, and Investor Sentiment around the Stock Crash of October 1987". In: *The Journal of Business*. 65 (4), pp. 557–570.

Sifry, D. (2007) "The State of the Live Web, April 2007". Retrieved 2011-12-16 from URI: http://www.sifry.com/alerts/archives/000493.html

Smailović, J.; Grčar, M.; Lavrač, N.; Žnidaršič, M. (2014) "Stream-based active learning for sentiment analysis in the financial domain". In: *Information Sciences*. 285, pp. 181–203.

Smola, A. J.; Schölkopf, B. (2004) "A tutorial on support vector regression". In: *Statistics and Computing*. 14 (3), pp. 199–222.

Sokolova, M.; Lapalme, G. (2009) "A systematic analysis of performance measures for classification tasks". In: *Information Processing & Management*. 45 (4), pp. 427–437.

Sparck Jones, K. (1972) "A statistical interpretation of term specificity and its application in retrieval". In: *Journal of Documentation*. 28 (1), pp. 11–21.

Spinn3r (2015) "Features". Retrieved 2015-02-18 from URI: http://spinn3r.com/features

Stone, P. J.; Dunphy, D. C.; Smith, M. S.; Ogilvie, D. M. (1966) *The General Inquirer: A Computer Approach to Content Analysis*. Cambridge, MA, USA: MIT Press.

Subrahmanyam, A. (2007) "Behavioural Finance: A Review and Synthesis". In: *European Financial Management*. 14 (1), pp. 12–29.

Sudman, S. (1976) *Applied Sampling*. New York, NY, USA: Academic Press.

Tang, H.; Tan, S.; Cheng, X. (2009) "A survey on sentiment detection of reviews". In: *Expert Systems with Applications*. 36 (7), pp. 10760–10773.Technorati (2011a) "State of the Blogsphere 2011: Part 1". Retrieved 2012-04-13 from URI: http://technorati.com/social-media/article/state-of-the-blogosphere-2011-part1/

Technorati (2011b) "State of the Blogosphere 2011: Introduction and Methodology". Retrieved 2012-04-12 from URI: http://technorati.com/social-media/article/state-of-the-blogosphere-2011-introduction/

Technorati (2011c) "State of the Blogsphere 2011: Part 2". Retrieved 2012-04-13 from URI: http://technorati.com/blogging/article/state-of-the-blogosphere-2011-part2/

Technorati (2011d) "State of the Blogsphere 2011: Part 3". Retrieved 2012-04-13 from URI: http://technorati.com/blogging/article/state-of-the-blogosphere-2011-part3/

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

Tetlock, P. C.; Saar-Tsechansky, M.; Macskassy, S. (2008) "More than Words: Quantifying Language to Measure Firms' Fundamentals". In: *The Journal of Finance*. 63 (3), pp. 1437–1467.

Theil, H. (1961) *Economic Forecasts and Policy*. 2nd ed., Amsterdam, The Netherlands: North Holland.

Tsytsarau, M.; Palpanas, T. (2012) "Survey on mining subjective data on the web". In: *Data Mining and Knowledge Discovery*. 24 (3), pp. 478–514.

Turney, P. D.; Littman, M. L. (2003) "Measuring Praise and Criticism: Inference of Semantic Orientation from Association". In: *ACM Transactions on Information Systems*. 21 (4), pp. 315–346.

Tversky, A.; Kahneman, D. (1974) "Judgment under Uncertainty: Heuristics and Biases". In: *Science*. 185 (4157), pp. 1124–1131.

Vapnik, V. (1982) *Estimation of Dependences Based on Empirical Data*. New York, NY, USA: Springer.

Vapnik, V.; Chervonenkis, A. (1974) *Theory of Pattern Recognition [in Russian]*. Moscow, Russia: Nauka.

Vapnik, V.; Lerner, A. (1963) "Pattern recognition using generalized portrait method". In: *Automation and Remote Control*. 24, pp. 774–780.

Vapnik, V. N. (1999) "An Overview of Statistical Learning Theory". In: *IEEE Transactions on Neural Networks*. 10 (5), pp. 988–999.

Vapnik, V. N. (2000) *The Nature of Statistical Learning Theory*. 2nd ed., New York, NY, USA: Springer.

Wang, S.; Manning, C. D. (2012) "Baselines and bigrams: Simple, good sentiment and topic classification". In: *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics - Volume 2 (ACL'12)*. Jeju, South Korea, pp. 90–94.

Wilcoxon, F. (1945) "Individual Comparisons by Ranking Methods". In: *Biometrics Bulletin*. 1 (6), pp. 80–83.

Williams, B. (1978) *A Sampler on Sampling*. New York, NY, USA: John Wiley & Sons.

Wilson, T. A. (2008) "Fine-Grained Subjectivity and Sentiment Analysis: Recognizing the Intensity, Polarity, and Attitudes of Private States". PhD Thesis. University of Pittsburgh.

Wooldridge, J. M. (2013) *Introductory Econometrics: A Modern Approach*. 5th in. ed., Andover, UK: Cengage Learning.

Yang, Y. (1999) "An Evaluation of Statistical Approaches to Text Categorization". In: *Information Retrieval*. 90 (1), pp. 69–90.

Yang, Y.; Liu, X. (1999) "A re-examination of text categorization methods". In: *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '99)*. Berkeley, CA, USA, pp. 42–49.

Yang, Y.; Pedersen, J. O. (1997) "A Comparative Study on Feature Selection in Text Categorization". In: *Proceedings of the 14th International Conference on Machine Learning (ICML'97)*. Nashville, TN, USA, pp. 412–420.

Yazici, B.; Yolacan, S. (2007) "A comparison of various tests of normality". In: *Journal of Statistical Computation and Simulation*. 77 (2), pp. 175–183.

Yi, J.; Nasukawa, T.; Bunescu, R.; Niblack, W. (2003) "Sentiment Analyzer: Extracting Sentiments About a Given Topic Using Natural Language Processing Techniques". In: *Proceedings of the 3rd IEEE International Conference on Data Mining (ICDM'03)*. Melbourne, FL USA, pp. 427–434.

Zhang, C. (2008) "Defining, Modeling, and Measuring Investor Sentiment". Thesis. University of California, Berkley.

Zhang, W.; Skiena, S. (2010) "Trading Strategies to Exploit Blog and News Sentiment". Retrieved 2014-11-13 from URI: http://www.cs.sunysb.edu/~skiena/lydia/blogtrading.pdf

A Appendix

A.1 Sample Blog Documents

The following figures provide details for the blog documents referred to in Figure 1 in the introduction Section 1.1.

Thursday, June 29, 2006

iPod Cell Phones coming, but why and when?

Some insightful commentary over at roughly drafted.com regarding Apple iPod and adding a cell phone capability to it - or adding iPod - ness to a cell phone. Tough to tell what would happen in this Chicken/Egg debate.

Figure 43: Excerpt of a Blogspot document about an iPod with cell phone capabilities. Publication date: 2006-06-29. URI: http://ipod-info.blogspot.de/2006/06/ipod-cell-phones-coming-but-why-and.html, retrieved 2014-06-20.

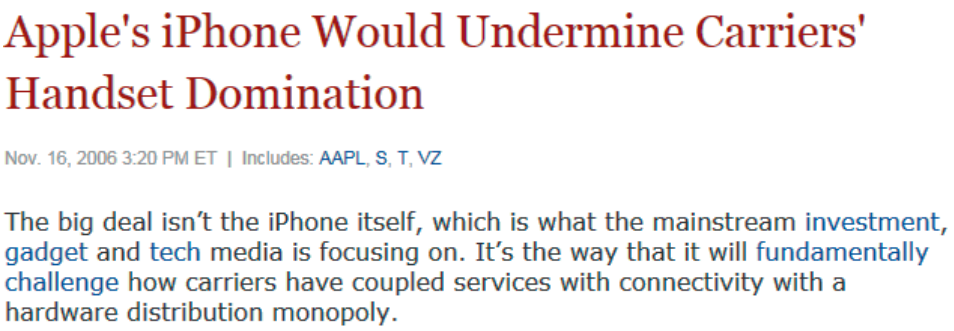
Thursday, September 07, 2006

An Apple iPod Cell Phone Rumor A Day.....

I have said a few times that Apple could be a mobile marketing and mobile commerce powerhouse if they introduced an iPod cell phone.

iTunes could provide the conduit for the PC and mobile worlds.

Figure 44: Excerpt of a Blogspot document about Apple as a cell phone provider. Publication date: 2006-09-07. URI: http://theponderingprimate.blogspot.de/2006/09/apple-ipod-cell-phone-rumor-day.html, retrieved 2014-06-20.



Apple's iPhone Would Undermine Carriers' Handset Domination

Nov. 16, 2006 3:20 PM ET | Includes: AAPL, S, T, VZ

The big deal isn't the iPhone itself, which is what the mainstream investment, gadget and tech media is focusing on. It's the way that it will fundamentally challenge how carriers have coupled services with connectivity with a hardware distribution monopoly.

Figure 45: Excerpt of a Seekingalpha document about Apple's upcoming iPhone release. Publication date: 2006-11-16. URI: http://seekingalpha.com/article/20780-apples-iphone-would-undermine-carriers-handset-domination, retrieved 2014-11-10.

Gizmodo: Apple to Announce iPhone on Monday

Dec. 15, 2006 1:52 PM ET | 5 comments | About: Apple Inc. (AAPL)

By Carl Howe

Gizmodo has a three sentence story claiming the iPhone will be announced on Monday, December 18. I can't claim to have any information on the topic, but I can perhaps shed some light on why Apple (AAPL) might do this (if they do).

Figure 46: Excerpt of a Seekingalpha document about an anticipated iPhone announcement. Publication date: 2006-12-15. URI: http://seekingalpha.com/article/22522-gizmodo-apple-to-announce-iphone-on-monday, retrieved 2014-06-20.

Apple Reinvents the Phone with iPhone

MACWORLD SAN FRANCISCO—January 9, 2007—Apple® today introduced iPhone, combining three products—a revolutionary mobile phone, a widescreen iPod® with touch controls, and a breakthrough Internet communications device with desktop-class email, web browsing, searching and maps—into one small and lightweight handheld device. iPhone introduces an entirely new user interface based on a large multi-touch display and pioneering new software, letting users control iPhone with just their fingers. iPhone also ushers in an era of software power and sophistication never before seen in a mobile device, which completely redefines what users can do on their mobile phones.

"iPhone is a revolutionary and magical product that is literally five years ahead of any other mobile phone," said Steve Jobs, Apple's CEO. "We are all born with the ultimate pointing device—our fingers—and iPhone uses them to create the most revolutionary user interface since the mouse."

Figure 47: Excerpt of the official press release of Apple's iPhone announcement. Publication date: 2007-01-09. URI: http://www.apple.com/pr/library/2007/01/09Apple-Reinvents-the-Phone-with-iPhone.html, retrieved 2014-06-23.

A.2 Corpus of Blog Documents

The corpus proposed in this thesis consists of blog documents retrieved from the Seekingalpha blog platform. Table 57 lists the URIs and publication dates of the 591 blog documents that were annotated with one or more stocks but with one sentiment orientation. Table 58 lists the URIs and publication dates of the 47 blog documents that were annotated with multiple stocks and with multiple sentiment orientations. The overall Corpus A is thus made up of all 638 blog documents. Finally, Table 59 lists the stocks to which the annotations of investor sentiment in the overall corpus refer to.

Table 57: Seekingalpha blog documents annotated with one sentiment orientation. All blog documents except the three ones in italic and grey background make up Corpus B described in Section 3.1.2.

URI	Publication date
http://seekingalpha.com/article/8857-cemex-an-excellent-proxy-for-infrastructure-growth-cx	2006-02-14
http://seekingalpha.com/article/59134-mcdonald-s-global-exposure-perfect-for-an-economic-slowdown	2008-01-06
http://seekingalpha.com/article/60480-boeing-s-dreamliner-delays-will-cost-them	2008-01-17
http://seekingalpha.com/article/60527-united-technologies-molten-salt-solar-power-generation	2008-01-17
http://seekingalpha.com/article/60766-ibm-also-looks-safe	2008-01-20
http://seekingalpha.com/article/61160-bank-of-america-is-columbia-a-problem	2008-01-23
http://seekingalpha.com/article/63065-the-past-due-loan-problem-at-american-express	2008-02-05
http://seekingalpha.com/article/64294-cisco-layoffs-in-the-offing	2008-02-12
http://seekingalpha.com/article/64373-why-is-disney-failing-in-hong-kong	2008-02-13
http://seekingalpha.com/article/65318-wal-mart-excelling-in-a-dismal-environment	2008-02-20
http://seekingalpha.com/article/67640-wal-mart-remains-undervalued	2008-03-07
http://seekingalpha.com/article/71929-wal-mart-strong-international-growth-especially-latin-america	2008-04-11
http://seekingalpha.com/article/72123-ge-s-earnings-miss-what-ever-happened-to-warning-investors-first	2008-04-13
http://seekingalpha.com/article/72441-intel-s-quarter-on-target-tech-sector-exhales	2008-04-15
http://seekingalpha.com/article/72450-johnson-johnson-a-good-long-term-buy	2008-04-16
http://seekingalpha.com/article/72709-buy-microsoft-ahead-of-earnings-citi	2008-04-17
http://seekingalpha.com/article/72626-will-pfizer-hit-a-home-run-with-avant-s-cancer-treatment	2008-04-17
http://seekingalpha.com/article/73054-has-pfizer-s-crown-jewel-lipitor-lost-its-value	2008-04-21
http://seekingalpha.com/article/73530-unitedhealth-group-an-unsustainable-model	2008-04-23
http://seekingalpha.com/article/73505-unitedhealth-looks-like-dead-money	2008-04-23
http://seekingalpha.com/article/74188-american-express-false-sense-of-security	2008-04-25
http://seekingalpha.com/article/76287-disney-beats-and-iger-speaks	2008-05-08
http://seekingalpha.com/article/77394-chevron-rising-revenues-and-analyst-estimates	2008-05-15
http://seekingalpha.com/article/78565-faa-bans-chantix-will-pfizer-be-smoked-out-of-the-stop-smoking-market	2008-05-23
http://seekingalpha.com/article/78938-home-depot-continues-to-impress-on-dividends	2008-05-27
http://seekingalpha.com/article/79148-hewlett-packard-shows-solid-q2-results-estimates-rise	2008-05-28
http://seekingalpha.com/article/80146-dupont-agriculture-and-solar-in-one-safe-stock	2008-06-08
http://seekingalpha.com/article/80660-the-market-sees-microsoft-losing-its-grip	2008-06-10
http://seekingalpha.com/article/81576-ge-more-bad-news-to-come	2008-06-17
http://seekingalpha.com/article/82484-bofa-ceo-lewis-the-next-to-fall	2008-06-24
http://seekingalpha.com/article/82439-ten-reasons-to-like-american-express	2008-06-24
http://seekingalpha.com/article/83016-the-curious-case-of-hernan-arbizu	2008-06-27
http://seekingalpha.com/article/84386-alcoa-analysts-positive-on-good-quarter-results	2008-07-10
http://seekingalpha.com/article/86395-unitedhealth-investors-lose-faith-in-management	2008-07-23
http://seekingalpha.com/article/87636-the-long-case-for-bank-of-america	2008-07-29
http://seekingalpha.com/article/91708-interested-in-bank-of-america-consider-the-preferred-shares	2008-08-20
http://seekingalpha.com/article/92131-mercks-gardasil-a-risky-and-unnecessary-vaccine	2008-08-22
http://seekingalpha.com/article/93373-exxon-shareholders-suffer-a-windfall-loss-of-13-7	2008-08-31
http://seekingalpha.com/article/95366-jp-morgan-buyer-s-remorse-on-bear-stearns	2008-09-14
http://seekingalpha.com/article/99290-how-not-to-think-about-investing	2008-10-10
http://seekingalpha.com/article/99794-cisco-s-ceo-sees-technology-led-productivity-gains-in-this-downturn	2008-10-14
http://seekingalpha.com/article/101476-american-express-stock-up-profits-down	2008-10-23
http://seekingalpha.com/article/101497-merck-continues-to-thin-out	2008-10-23
http://seekingalpha.com/article/103988-cisco-reports-wednesday-more-estimate-cuts	2008-11-04
http://seekingalpha.com/article/105286-how-to-protect-against-falling-caterpillar	2008-11-11
http://seekingalpha.com/article/105796-no-surprises-as-intel-reduces-sales-guidance	2008-11-13
http://seekingalpha.com/article/106027-ge-s-dividend-assertion-is-dangerous	2008-11-14
http://seekingalpha.com/article/106036-ge-s-immelt-buys-shares-should-you	2008-11-14

URI	Publication date
http://seekingalpha.com/article/106829-ge-not-so-good-things-come-to-light	2008-11-19
http://seekingalpha.com/article/108509-get-paid-to-go-long-bank-of-america	2008-12-01
http://seekingalpha.com/article/110592-caterpillar-does-it-deserve-the-goldman-downgrade	2008-12-14
http://seekingalpha.com/article/110675-florida-s-health-insurance-deception	2008-12-15
http://seekingalpha.com/article/111570-s-p-lowers-ge-outlook-a-downgrade-may-be-possible	2008-12-19
http://seekingalpha.com/article/111999-bank-of-america-optimism-is-unwarranted	2008-12-23
http://seekingalpha.com/article/112699-pfizer-a-case-study-of-the-market-s-extremely-low-valuations	2008-12-31
http://seekingalpha.com/article/113685-intel-q4-worse-than-expected	2009-01-07
http://seekingalpha.com/article/113830-alcoa-makes-cuts-but-analysts-say-not-enough	2009-01-08
http://seekingalpha.com/article/114375-will-disney-s-china-expansion-succeed	2009-01-12
http://seekingalpha.com/article/114465-alcoa-is-hard-to-believe-in	2009-01-13
http://seekingalpha.com/article/114578-bank-of-america-may-have-to-cut-dividend	2009-01-13
http://seekingalpha.com/article/114827-earnings-preview-intel	2009-01-14
http://seekingalpha.com/article/114828-earnings-preview-j-p-morgan	2009-01-14
http://seekingalpha.com/article/114672-recession-hits-alcoa-raw-material-demand-substantially-reduced	2009-01-14
http://seekingalpha.com/article/115035-the-short-case-for-jpmorgan	2009-01-17
http://seekingalpha.com/article/115571-thank-god-for-ibm	2009-01-21
http://seekingalpha.com/article/115940-microsoft-misses-q2-plans-to-lay-off-5-000	2009-01-22
http://seekingalpha.com/article/115809-more-declines-anticipated-for-ge-plus-ratings-risk-and-expected-dividend-cut-ubs	2009-01-22
http://seekingalpha.com/article/116462-buy-sell-or-hold-ibm-runs-away-with-the-market	2009-01-26
http://seekingalpha.com/article/116485-what-caterpillar-s-report-tells-us	2009-01-26
http://seekingalpha.com/article/116669-outlook-is-bleak-for-bank-of-america	2009-01-27
http://seekingalpha.com/article/117045-earnings-show-that-at-t-can-weather-a-downturn	2009-01-28
http://seekingalpha.com/article/145978-is-microsoft-heading-for-30	2009-01-29
http://seekingalpha.com/article/117317-pressure-builds-at-microsoft	2009-01-29
http://seekingalpha.com/article/127511-pfizer-takes-steps-to-provide-healthcare-for-china-s-poor	2009-02-03
http://seekingalpha.com/article/118674-disney-should-go-shopping	2009-02-05
http://seekingalpha.com/article/118649-is-disney-losing-its-magic	2009-02-05
http://seekingalpha.com/article/119299-disney-s-disappointing-q1-earnings	2009-02-09
http://seekingalpha.com/article/119928-boeing-s-bad-balance-sheet-may-doom-it	2009-02-11
http://seekingalpha.com/article/121185-buy-sell-or-hold-coke-still-has-the-fizz	2009-02-18
http://seekingalpha.com/article/121490-hewlett-packard-expect-guidance-below-street-estimate	2009-02-19
http://seekingalpha.com/article/123573-microsoft-eps-estimates-down-as-pc-demand-deteriorates	2009-03-02
http://seekingalpha.com/article/124351-wal-mart-leads-retailers-again	2009-03-05
http://seekingalpha.com/article/125354-intel-might-come-out-of-recession-stronger	2009-03-11
http://seekingalpha.com/article/125280-pali-downgrades-disney-after-three-month-decline	2009-03-11
http://seekingalpha.com/article/125824-is-microsoft-headed-for-back-to-back-down-years	2009-03-13
http://seekingalpha.com/article/126122-mcdonald-s-an-excellent-defensive-pick	2009-03-16
http://seekingalpha.com/article/127619-jpmorgan-will-help-extend-the-rally-cramer	2009-03-24
http://seekingalpha.com/article/128074-chevron-is-the-future-cramer	2009-03-26
http://seekingalpha.com/article/130189-are-strike-worries-causing-at-t-s-stock-to-lag	2009-04-08
http://seekingalpha.com/article/130201-new-reports-of-wal-mart-russia	2009-04-08
http://seekingalpha.com/article/130960-can-intel-drive-the-market-lower	2009-04-15
http://seekingalpha.com/article/131282-jpmorgan-chase-proves-banking-isn-t-dead	2009-04-16
http://seekingalpha.com/article/131806-earnings-preview-caterpillar	2009-04-20
http://seekingalpha.com/article/131932-ibm-s-q1-results-offer-even-more-reason-for-aggressive-m-a	2009-04-21
http://seekingalpha.com/article/132499-at-t-earnings-better-than-expected-shares-headed-up	2009-04-23
http://seekingalpha.com/article/137034-cisco-order-inflows-insufficient-to-arrest-steep-revenue-declines	2009-05-12
http://seekingalpha.com/article/137161-disney-shares-fully-valued-time-to-sell	2009-05-12
http://seekingalpha.com/article/137969-my-microsoft-investment-summary	2009-05-17
http://seekingalpha.com/article/138208-ecuador-a-pr-issue-for-chevron	2009-05-18
http://seekingalpha.com/article/139595-danone-stock-price-down-but-still-at-risk	2009-05-26
http://seekingalpha.com/article/140703-microsoft-s-got-talent	2009-06-01
http://seekingalpha.com/article/141063-can-procter-gamble-cope-with-compensating-consumer-behavior	2009-06-03
http://seekingalpha.com/article/142256-microsoft-should-benefit-from-windows-7-and-bing-in-2010	2009-06-09
http://seekingalpha.com/article/142533-home-depot-s-not-so-awful-news	2009-06-11
http://seekingalpha.com/article/144824-procter-gamble-exposure-to-improving-economy	2009-06-23
http://seekingalpha.com/article/144949-boeing-s-dreamliner-nightmare-may-present-investor-opportunity	2009-06-24
http://seekingalpha.com/article/145081-boeing-s-dreamliner-still-grounded-but-its-stock-could-soar	2009-06-24
http://seekingalpha.com/article/145616-why-investors-should-avoid-american-express	2009-06-26
http://seekingalpha.com/article/146793-american-express-faces-bearish-trade	2009-07-02
http://seekingalpha.com/article/147200-options-bearish-on-merck	2009-07-06
http://seekingalpha.com/article/147675-why-i-bought-cat	2009-07-08
http://seekingalpha.com/article/149893-earnings-preview-unitedhealth-group	2009-07-20
http://seekingalpha.com/article/150608-why-ge-is-no-longer-a-financial-company	2009-07-22
http://seekingalpha.com/article/152028-investors-can-write-off-ge-capital-for-the-next-5-years	2009-07-29
http://seekingalpha.com/article/153033-disney-shows-stability-but-no-improvement	2009-08-02

URI	Publication date
http://seekingalpha.com/article/153664-sec-rewards-ge-accounting-fraud	2009-08-04
http://seekingalpha.com/article/156798-home-depot-weathering-the-downturn	2009-08-18
http://seekingalpha.com/article/159490-chevron-shows-strength-in-tough-times	2009-09-02
http://seekingalpha.com/article/160403-anheuser-busch-inbev-still-room-for-further-upside	2009-09-08
http://seekingalpha.com/article/162360-ibm-s-upside-is-the-market-s-upside	2009-09-20
http://seekingalpha.com/article/162763-exxon-mobil-is-a-buy	2009-09-22
http://seekingalpha.com/article/163295-ibm-downgraded-on-limited-revenue-growth-potential	2009-09-24
http://seekingalpha.com/article/164432-why-is-it-time-to-short-bofa	2009-10-02
http://seekingalpha.com/article/166573-philips-electronics-ready-for-the-recovery	2009-10-14
http://seekingalpha.com/article/166920-ge-ahead-of-earnings-sentiment-trends-downward	2009-10-16
http://seekingalpha.com/article/169013-general-electric-generally-down	2009-10-27
http://seekingalpha.com/article/171458-looking-for-further-upside-in-heineken	2009-11-05
http://seekingalpha.com/article/172512-wal-mart-does-not-save-families-3-100-a-year	2009-11-10
http://seekingalpha.com/article/174338-amex-to-acquire-revolution-money-for-300-million	2009-11-19
http://seekingalpha.com/article/174815-hewlett-packard-sentiment-moves-upward-ahead-of-earnings	2009-11-23
http://seekingalpha.com/article/176312-why-is-bank-of-america-so-bad-at-picking-its-leaders	2009-12-03
http://seekingalpha.com/article/183094-ibm-s-sentiment-heads-down-ahead-of-earnings	2010-01-19
http://seekingalpha.com/article/185117-united-technologies-just-beats-but-revenues-drop	2010-01-28
http://seekingalpha.com/article/188902-l-oreal-seems-expensive-better-returns-elsewhere	2010-02-17
http://seekingalpha.com/article/189227-home-depot-shows-upside-potential-as-housing-sector-recovers	2010-02-18
http://seekingalpha.com/article/189985-will-intel-continue-to-lose-server-market-share	2010-02-22
http://seekingalpha.com/article/193560-exxon-mobil-well-positioned-for-future-growth	2010-03-14
http://seekingalpha.com/article/194637-disney-is-bringing-the-magic-back	2010-03-19
http://seekingalpha.com/article/199345-cemex-just-walk-away	2010-04-19
http://seekingalpha.com/article/200015-do-bank-of-americas-blowout-earnings-actually-blow	2010-04-21
http://seekingalpha.com/article/200884-what-is-driving-the-v-shaped-recovery	2010-04-26
http://seekingalpha.com/article/201071-dupont-tops-estimates-guides-higher	2010-04-27
http://seekingalpha.com/article/201602-long-term-trade-bets-on-jpmorgan	2010-04-29
http://seekingalpha.com/article/202490-alcoa-shows-potential-bearish-pattern	2010-05-03
http://seekingalpha.com/article/202437-united-technologies-dividend-stock-analysis	2010-05-03
http://seekingalpha.com/instablog/604052-element-alpha/66639-whos-betting-on-exxon-buffett-and-soros	2010-05-04
http://seekingalpha.com/article/203659-3m-company-dividend-stock-analysis	2010-05-07
http://seekingalpha.com/article/203654-procter-gamble-long-term-view	2010-05-07
http://seekingalpha.com/article/204159-ibm-a-nugget-in-plain-sight	2010-05-10
http://seekingalpha.com/article/204358-iron-man-2-makes-hollywood-history-disney-viacom-surge-higher	2010-05-11
http://seekingalpha.com/article/204496-is-mcdonald-s-stock-as-much-of-a-bargain-as-its-dollar-menu	2010-05-11
http://seekingalpha.com/article/204797-cisco-another-positive-quarter-for-earnings	2010-05-12
http://seekingalpha.com/article/204781-cisco-earnings-return-to-strong-balanced-growth	2010-05-12
http://seekingalpha.com/article/204974-disconnect-cisco-shares-slump-on-strong-report	2010-05-13
http://seekingalpha.com/article/204763-ibm-s-2015-roadmap-to-20-per-share-earnings	2010-05-13
http://seekingalpha.com/article/204982-why-the-outlook-for-travelers-is-mixed	2010-05-13
http://seekingalpha.com/article/205280-merck-s-busy-road-ahead	2010-05-16
http://seekingalpha.com/article/205401-earnings-preview-hewlett-packard-wal-mart-dell	2010-05-17
http://seekingalpha.com/article/205460-earnings-preview-the-home-depot	2010-05-17
http://seekingalpha.com/article/205565-pulse-check-caterpillar-growth-threatened-by-stronger-dollar	2010-05-17
http://seekingalpha.com/article/205631-beiersdorf-stock-looks-fully-priced	2010-05-18
http://seekingalpha.com/article/205890-earnings-scorecard-chevron	2010-05-19
http://seekingalpha.com/article/205818-hewlett-packard-still-on-top-of-the-pc-heap-after-a-strong-quarter	2010-05-19
http://seekingalpha.com/article/206907-boeing-q1-earnings-make-analysts-nervous	2010-05-25
http://seekingalpha.com/article/207174-hewlett-packard-high-growth-low-valuation	2010-05-26
http://seekingalpha.com/article/207115-intel-from-a-growth-company-to-an-income-stock	2010-05-26
http://seekingalpha.com/article/207450-chevron-corporation-dividend-stock-analysis	2010-05-28
http://seekingalpha.com/article/208024-is-at-t-a-good-investment-part-i	2010-06-02
http://seekingalpha.com/article/208068-unitedhealth-increases-annual-dividend-from-0-03-to-0-50-a-share	2010-06-02
http://seekingalpha.com/article/208263-is-at-t-a-good-investment-part-ii	2010-06-03
http://seekingalpha.com/article/208315-microsoft-the-unfriendliness-discount	2010-06-03
http://seekingalpha.com/article/208453-coca-cola-dividend-stock-analysis	2010-06-04
http://seekingalpha.com/article/209472-dupont-is-on-a-growth-trajectory	2010-06-10
http://seekingalpha.com/article/209667-the-case-for-pfizer	2010-06-11
http://seekingalpha.com/article/210984-exxon-mobil-a-good-buy-and-hold-candidate	2010-06-21
http://seekingalpha.com/article/212593-traders-back-away-from-home-depot	2010-06-30
http://seekingalpha.com/article/213534-how-wal-mart-s-margins-could-be-squeezed-by-rising-chinese-manufacturing-costs	2010-07-07
http://seekingalpha.com/article/214474-jpmorgan-chase-earnings-preview	2010-07-14
http://seekingalpha.com/article/214707-earnings-scorecard-alcoa	2010-07-15
http://seekingalpha.com/article/214651-intel-reports-blowout-quarter	2010-07-15
http://seekingalpha.com/article/214926-ge-lights-on-lights-off	2010-07-16
http://seekingalpha.com/article/215031-alcoa-s-many-warning-flags	2010-07-18

URI	Publication date
http://seekingalpha.com/article/215079-bank-of-america-why-it-s-scary-being-a-bank-in-america	2010-07-19
http://seekingalpha.com/article/215233-earnings-update-ibm-tops-profits-on-thinner-revenue-and-issues-strong-guidance	2010-07-19
http://seekingalpha.com/article/217860-intel-turns-bearish	2010-08-01
http://seekingalpha.com/article/219631-next-hp-ceo-needs-to-deliver-growth-and-vision-not-just-cost-cuts	2010-08-09
http://seekingalpha.com/article/220459-hewlett-packard-eight-day-losing-streaks	2010-08-13
http://seekingalpha.com/article/220642-5-reasons-why-exxon-mobil-is-a-buy	2010-08-16
http://seekingalpha.com/article/220621-why-i-m-buying-tech	2010-08-16
http://seekingalpha.com/article/222110-boeing-s-government-contract-breeds-a-nice-options-play	2010-08-24
http://seekingalpha.com/article/222678-intel-confirms-markets-pessimism	2010-08-27
http://seekingalpha.com/article/222876-at-t-offers-a-stable-dividend-with-room-to-grow	2010-08-30
http://seekingalpha.com/article/226459-boeing-is-overvalued-by-at-least-25	2010-09-22
http://seekingalpha.com/article/226728-american-express-shares-its-a-bummer-being-a-bank	2010-09-23
http://seekingalpha.com/article/227065-hp-is-so-beaten-down-the-ceo-choice-hardly-matters	2010-09-26
http://seekingalpha.com/article/227230-wal-mart-management-goes-shopping-in-south-africa-why-I-are-not-impressed	2010-09-27
http://seekingalpha.com/article/227346-hewlett-packard-on-an-elevator-heading-down	2010-09-28
http://seekingalpha.com/article/237612-merck-a-win-in-the-race-for-safe-profitable-heart-health-drug	2010-10-18
http://seekingalpha.com/article/231412-caterpillar-beats-with-strong-sales-in-all-regions	2010-10-21
http://seekingalpha.com/article/231468-travelers-beats-in-q3-raises-outlook	2010-10-21
http://seekingalpha.com/article/232156-4-reasons-why-intel-is-a-buy-in-this-economy	2010-10-26
http://seekingalpha.com/article/234636-lowe-s-vs-home-depot-which-is-the-best-value-investment	2010-11-04
http://seekingalpha.com/article/236267-boeing-for-long-term-investors-only	2010-11-11
http://seekingalpha.com/article/236354-can-cisco-rebound	2010-11-11
http://seekingalpha.com/article/236287-two-steps-forward-two-steps-back-for-cisco	2010-11-11
http://seekingalpha.com/article/240092-general-electric-a-higher-rated-higher-yielding-bond-to-consider	2010-12-05
http://seekingalpha.com/article/240179-microsoft-market-leader-with-exceptional-cash-flow	2010-12-06
http://seekingalpha.com/article/240996-the-bullish-case-for-3m	2010-12-09
http://seekingalpha.com/article/242847-going-long-ge	2010-12-20
http://seekingalpha.com/article/244663-6-most-vulnerable-stocks-in-the-dow	2011-01-04
http://seekingalpha.com/article/245539-boeing-business-picks-up-but-stock-still-looks-risky	2011-01-09
http://seekingalpha.com/article/246225-time-is-running-out-for-sap	2011-01-12
http://seekingalpha.com/article/247063-johnson-johnson-offers-good-value-with-a-margin-of-safety	2011-01-18
http://seekingalpha.com/article/247664-whirlpool-could-have-fourth-quarter-upside-surprise	2011-01-20
http://seekingalpha.com/article/247636-united-health-group-beats-on-higher-revenue	2011-01-20
http://seekingalpha.com/article/247977-microsoft-the-undervalued-giant-makes-a-comeback	2011-01-23
http://seekingalpha.com/article/248169-mcdonald-s-reports-in-line-better-sales-in-all-regions	2011-01-24
http://seekingalpha.com/article/251538-investing-in-coca-cola	2011-02-08
http://seekingalpha.com/article/251387-investment-guru-eric-sprott-says-silver-will-soar-in-2011	2011-02-08
http://seekingalpha.com/article/251722-3m-medium-term-outperformance-still-unlikely-despite-buybacks	2011-02-09
http://seekingalpha.com/article/252158-mcdonald-s-a-simple-investment-at-a-discount	2011-02-10
http://seekingalpha.com/article/252967-analysts-pfizer-s-research-cuts-should-help-boost-stock-price	2011-02-15
http://seekingalpha.com/article/256401-pfizer-the-market-s-undervalued-ugly-duckling	2011-03-04
http://seekingalpha.com/article/257276-mcdonald-s-potential-upside-from-customer-traffic-growth	2011-03-09
http://seekingalpha.com/article/423581-verizon-provides-generous-dividends-and-stock-appreciation	2011-03-09
http://seekingalpha.com/article/423771-verizon-winning-tech-means-big-dividends-for-investors	2011-03-09
http://seekingalpha.com/article/260940-verizon-breaks-out	2011-03-30
http://seekingalpha.com/article/261113-microsoft-potential-30-to-40-upside	2011-03-31
http://seekingalpha.com/article/256373-why-i-finally-gave-up-on-johnson-johnson	2011-04-04
http://seekingalpha.com/article/262659-siemens-revamps-its-industrial-business	2011-04-08
http://seekingalpha.com/article/263456-coca-cola-attractive-upside-at-no-cost	2011-04-14
http://seekingalpha.com/article/263759-2-catalysts-that-make-microsoft-a-screaming-buy	2011-04-15
http://seekingalpha.com/article/263904-deutsche-bank-scaling-up-its-m-a-business	2011-04-17
http://seekingalpha.com/article/264318-united-technologies-great-business-so-so-price	2011-04-19
http://seekingalpha.com/article/264757-verizon-beats-on-strength-in-wireless-fios	2011-04-21
http://seekingalpha.com/article/265983-philips-electronics-a-bright-light-that-will-shine-again	2011-04-27
http://seekingalpha.com/article/266586-microsoft-after-earnings-still-a-screaming-buy	2011-04-29
http://seekingalpha.com/article/266882-intel-shares-are-still-undervalued	2011-05-01
http://seekingalpha.com/article/266987-basing-index-investing-on-a-single-stock-s-performance-is-not-a-valid-strategy	2011-05-02
http://seekingalpha.com/article/267517-why-cisco-could-be-the-cornerstone-of-your-portfolio	2011-05-04
http://seekingalpha.com/instablog/743797-tomaspray/177943-4-worst-bank-stocks	2011-05-13
http://seekingalpha.com/article/269862-merck-is-the-perfect-rx-for-this-volatile-market	2011-05-13
http://seekingalpha.com/article/270122-why-at-t-needs-t-mobile	2011-05-16
http://seekingalpha.com/article/272640-cisco-systems-unwanted-and-unloved	2011-05-31
http://seekingalpha.com/article/272842-wal-mart-international-a-true-growth-story	2011-06-01
http://seekingalpha.com/article/273732-jp-morgan-undervalued-thanks-to-a-disciplined-dimon	2011-06-07
http://seekingalpha.com/article/273945-upside-and-downside-scenarios-for-dupont	2011-06-08

URI	Publication date
http://seekingalpha.com/article/274174-jp-morgan-a-new-trading-idea-for-this-week	2011-06-09
http://seekingalpha.com/article/274040-procter-gamble-still-a-strong-buy-after-all-this-time	2011-06-09
http://seekingalpha.com/article/276004-why-wal-mart-is-losing-market-share-in-china	2011-06-21
http://seekingalpha.com/article/281823-bank-of-america-a-disappointment-as-expected	2011-06-26
http://seekingalpha.com/article/276892-universal-corporation-dividend-champion-trading-below-book-value	2011-06-28
http://seekingalpha.com/article/277373-at-t-a-dividend-stock-pick-for-the-next-5-years	2011-06-30
http://seekingalpha.com/article/278878-cons-outweigh-the-pros-for-the-travelers-companies	2011-07-11
http://seekingalpha.com/article/279304-jpmorgan-chase-shows-technical-buy-signal	2011-07-13
http://seekingalpha.com/article/280006-unitedhealth-earnings-preview	2011-07-18
http://seekingalpha.com/article/280386-time-to-buy-cisco	2011-07-20
http://seekingalpha.com/article/281344-mcdonald-s-limited-upside-moderate-risk	2011-07-24
http://seekingalpha.com/article/282295-ge-makes-a-critical-mistake	2011-07-27
http://seekingalpha.com/article/283765-microsoft-will-remain-the-king-of-cash	2011-08-02
http://seekingalpha.com/article/283974-the-8-worst-stocks-of-the-dow	2011-08-02
http://seekingalpha.com/article/283826-telefonica-a-defensive-stock-to-play-the-european-debt-crisis	2011-08-02
http://seekingalpha.com/article/284173-mobile-market-share-windows-phone-7-5-is-just-the-beginning	2011-08-03
http://seekingalpha.com/article/284253-insmed-the-definition-of-risk	2011-08-03
http://seekingalpha.com/article/286032-what-windows-8-means-to-microsoft-s-stock	2011-08-09
http://seekingalpha.com/article/286362-sap-looks-to-add-more-china-to-its-growth	2011-08-10
http://seekingalpha.com/article/286793-just-one-stock-overseas-growth-strong-margins-and-dividends-boost-mcdonald-s	2011-08-11
http://seekingalpha.com/article/287220-high-growth-modest-dividend-microsoft-provides-great-value-at-current-price	2011-08-14
http://seekingalpha.com/article/287558-a-bullish-case-for-microsoft	2011-08-15
http://seekingalpha.com/article/287953-wal-mart-more-international-fewer-shares-bigger-dividends	2011-08-17
http://seekingalpha.com/article/289026-3m-is-one-of-the-cornerstones-of-any-portfolio	2011-08-22
http://seekingalpha.com/article/288821-hewlett-packard-s-strategic-shift-is-a-big-mistake	2011-08-22
http://seekingalpha.com/article/288939-why-mcdonald-s-should-continue-to-outperform-the-market	2011-08-22
http://seekingalpha.com/article/289820-can-wal-mart-get-any-better	2011-08-25
http://seekingalpha.com/article/289892-microsoft-looks-very-attractive-before-the-windows-8-0-release	2011-08-25
http://seekingalpha.com/article/289994-why-coca-cola-should-be-in-your-portfolio	2011-08-26
http://seekingalpha.com/article/290364-microsoft-the-next-growth-opportunity	2011-08-29
http://seekingalpha.com/article/290745-unilever-delivers-good-looking-growth	2011-08-30
http://seekingalpha.com/article/290571-american-express-is-a-buy-for-growth-and-value	2011-08-30
http://seekingalpha.com/article/290651-is-walt-disney-a-good-play-for-a-bullish-autumn	2011-08-30
http://seekingalpha.com/article/291065-3m-innovative-company-with-robust-cash-flows	2011-09-01
http://seekingalpha.com/article/291763-home-depot-s-60-upside-potential	2011-09-06
http://seekingalpha.com/article/292180-ge-is-an-attractive-buy-and-hold-for-long-term-investors	2011-09-07
http://seekingalpha.com/article/292347-sanofi-another-encouraging-announcement	2011-09-08
http://seekingalpha.com/article/293400-sandisk-bargain-price-stock-poised-to-rally	2011-09-13
http://seekingalpha.com/article/293425-trouble-in-europe-spells-opportunity-for-these-3-stocks	2011-09-13
http://seekingalpha.com/article/293652-buy-merck-for-blue-chip-safety-with-a-nearly-5-yield	2011-09-14
http://seekingalpha.com/article/293956-a-bullish-stance-on-alcatel-lucent	2011-09-15
http://seekingalpha.com/article/293851-pfizer-offers-investors-tremendous-upside-potential	2011-09-15
http://seekingalpha.com/article/294270-the-market-is-wrong-about-united-technologies-and-a-lot-of-other-companies	2011-09-18
http://seekingalpha.com/article/294360-merck-s-competitive-advantage-buy-and-hold-but-sell-if-necessary	2011-09-19
http://seekingalpha.com/article/295089-microsoft-raises-quarterly-dividends-just-as-bernanke-is-about-to-make-them-even-more-valuable	2011-09-21
http://seekingalpha.com/article/295055-microsoft-the-dividend-increase-to-0-20-per-quarter-portends-a-material-increase-in-the-stock-price	2011-09-21
http://seekingalpha.com/article/295431-iconix-growing-fast-with-a-solid-stock	2011-09-22
http://seekingalpha.com/article/295326-5-stocks-poised-to-double-in-price	2011-09-22
http://seekingalpha.com/article/295606-microsoft-to-unlock-synergistic-value-through-cloud-integration	2011-09-23
http://seekingalpha.com/article/295694-bank-of-america-is-breaking-down-to-new-lows	2011-09-24
http://seekingalpha.com/article/296166-chevron-is-too-cheap-at-less-than-7-times-this-year-s-expected-eps	2011-09-27
http://seekingalpha.com/article/296542-5-bank-stocks-that-look-ready-to-tumble-further	2011-09-28
http://seekingalpha.com/article/296682-general-electric-dividend-and-technical-support-should-put-floor-under-stock-price	2011-09-29
http://seekingalpha.com/article/295988-behind-hp-s-sell-off-more-bad-management-decisions	2011-09-29
http://seekingalpha.com/article/296686-go-to-wal-mart-for-some-comfort	2011-09-29
http://seekingalpha.com/article/296859-drilling-wealth-caterpillar-s-tremendous-potential	2011-09-30
http://seekingalpha.com/article/296824-total-undervalued-but-short-term-risks-remain	2011-09-30
http://seekingalpha.com/article/296936-ariba-a-desired-monopoly-in-the-tech-space	2011-09-30
http://seekingalpha.com/article/296983-do-not-buy-verizon-s-cloud-hype	2011-09-30
http://seekingalpha.com/article/297026-dialing-for-dollars-telefonica-answers-the-call	2011-10-02
http://seekingalpha.com/article/297306-barnes-noble-could-be-the-next-borders	2011-10-03
http://seekingalpha.com/article/297566-novo-nordisk-market-fears-create-a-buying-opportunity	2011-10-04

URI	Publication date
http://seekingalpha.com/article/298022-at-t-trading-at-a-technical-bottom-and-yielding-over-6	2011-10-06
http://seekingalpha.com/article/298122-wall-street-sees-a-big-rally-ahead	2011-10-06
http://seekingalpha.com/article/298091-rio-tinto-undervalued-at-current-earnings-levels	2011-10-06
http://seekingalpha.com/article/298265-ibm-nowhere-close-to-peak-a-safe-bet	2011-10-07
http://seekingalpha.com/article/298409-5-tech-stocks-to-avoid	2011-10-09
http://seekingalpha.com/article/298700-agony-from-lipitor-s-expiration-won-t-be-exclusive-to-pfizer	2011-10-10
http://seekingalpha.com/article/298688-vivendi-milking-a-french-cash-cow	2011-10-10
http://seekingalpha.com/article/298677-mgm-s-prospects-grow-with-huge-opportunity-in-macau	2011-10-10
http://seekingalpha.com/article/298942-clearwire-may-go-bankrupt-by-next-year	2011-10-11
http://seekingalpha.com/article/299229-big-blue-keeps-rolling-on	2011-10-12
http://seekingalpha.com/article/299225-knight-capital-group-loves-volatility	2011-10-12
http://seekingalpha.com/article/299139-molycorp-and-general-moly	2011-10-12
http://seekingalpha.com/article/299029-tootsie-roll-overvalued-by-all-metrics	2011-10-12
http://seekingalpha.com/article/299193-is-vonage-calling-investors-back	2011-10-12
http://seekingalpha.com/news-article/2008352-lawsuit-says-mastercard-visa-fix-atm-fees	2011-10-12
http://seekingalpha.com/article/299013-why-ford-is-a-strong-buy	2011-10-12
http://seekingalpha.com/article/299280-2-stocks-that-will-outperform-in-the-next-sell-off	2011-10-13
http://seekingalpha.com/article/299726-zillow-admit-it-you-can-t-help-looking-at-this-stock	2011-10-14
http://seekingalpha.com/article/299564-stericycle-overdue-for-a-downturn	2011-10-14
http://seekingalpha.com/article/299742-halloween-indicator-trick-or-treat	2011-10-15
http://seekingalpha.com/article/299836-going-long-the-world-s-third-largest-aircraft-manufacturer	2011-10-16
http://seekingalpha.com/article/299810-chipotle-s-expanding-worldwide-concept-leads-to-huge-future-projected-earnings	2011-10-16
http://seekingalpha.com/article/299945-itron-set-to-rise-after-tough-past-6-months	2011-10-17
http://seekingalpha.com/article/299926-why-google-should-be-scared-of-siri	2011-10-17
http://seekingalpha.com/article/299957-xyratex-has-gone-up-too-far-too-fast	2011-10-17
http://seekingalpha.com/article/299886-goldman-sachs-in-a-lose-lose-situation	2011-10-17
http://seekingalpha.com/article/299955-5-stocks-that-just-turned-very-bullish	2011-10-17
http://seekingalpha.com/article/300251-kinder-morgan-s-el-paso-acquisition-is-a-smart-move	2011-10-18
http://seekingalpha.com/article/300222-short-selling-green-mountain-coffee-roasters	2011-10-18
http://seekingalpha.com/article/300445-intel-reports-its-most-profitable-quarter-ever	2011-10-19
http://seekingalpha.com/article/300457-2-stocks-with-potential-for-growth	2011-10-19
http://seekingalpha.com/article/300927-american-express-seems-to-be-on-a-slow-pitch	2011-10-20
http://seekingalpha.com/article/301534-why-you-should-sell-whole-foods	2011-10-24
http://seekingalpha.com/article/301528-4-must-haves-for-value-investors	2011-10-24
http://seekingalpha.com/article/302182-caterpillar-strong-growth-going-forward	2011-10-26
http://seekingalpha.com/article/304839-cramer-s-black-list-3-sell-and-4-avoid-ideas	2011-11-03
http://seekingalpha.com/article/304932-united-technologies-appears-undervalued-compared-to-aerospace-peers	2011-11-03
http://seekingalpha.com/article/305270-activision-unveils-2-new-games	2011-11-04
http://seekingalpha.com/article/305268-is-groupon-another-pets-com	2011-11-04
http://seekingalpha.com/article/305737-balchem-small-company-big-profits	2011-11-06
http://seekingalpha.com/article/305733-3-stocks-to-short-or-sell-right-now	2011-11-06
http://seekingalpha.com/article/305738-the-future-of-neurometrix	2011-11-06
http://seekingalpha.com/article/305809-inhibitex-the-next-hot-pharma-stock	2011-11-07
http://seekingalpha.com/article/305807-5-stocks-offering-dividends-that-will-double-your-money-in-5-years	2011-11-07
http://seekingalpha.com/article/305788-a-pair-of-dividend-giants-unloved-by-investors-but-showing-promise-for-the-next-20-years	2011-11-07
http://seekingalpha.com/article/306103-why-it-s-time-to-buy-intel	2011-11-08
http://seekingalpha.com/article/306206-j-c-penney-a-disaster-waiting-to-happen	2011-11-08
http://seekingalpha.com/article/306490-ge-makes-for-a-good-value-buy	2011-11-09
http://seekingalpha.com/article/306477-should-annaly-capital-be-a-core-holding-of-your-retirement-portfolio	2011-11-09
http://seekingalpha.com/article/306484-5-biotech-stocks-with-much-upside-potential	2011-11-09
http://seekingalpha.com/article/306917-can-groupon-compete-not-likely	2011-11-10
http://seekingalpha.com/article/307222-how-safe-is-the-doctor-s-dividend	2011-11-11
http://seekingalpha.com/article/307831-unitedhealth-group-a-strong-buy-despite-risks	2011-11-14
http://seekingalpha.com/article/307843-no-nba-means-short-msg	2011-11-14
http://seekingalpha.com/article/307848-kindle-fire-reviews-not-good-sell-amazon	2011-11-15
http://seekingalpha.com/article/308489-why-ford-and-gm-are-good-buys-right-now	2011-11-16
http://seekingalpha.com/article/308702-investors-should-short-allianz-insurer-is-too-vulnerable-to-european-sovereign-debt	2011-11-17
http://seekingalpha.com/article/308739-the-sky-is-the-limit-for-boeing	2011-11-17
http://seekingalpha.com/article/308695-exide-technologies-has-huge-upside-potential	2011-11-17
http://seekingalpha.com/article/308641-5-dividend-stocks-that-can-keep-surging-higher	2011-11-17
http://seekingalpha.com/article/308680-new-lumia-800-phone-should-boost-nokia-microsoft	2011-11-17
http://seekingalpha.com/article/308948-exxon-mobil-for-the-long-term	2011-11-18
http://seekingalpha.com/article/309141-3-reasons-amazon-com-looks-primed-for-a-netflix-like-drop	2011-11-20
http://seekingalpha.com/article/309195-at-t-my-stock-pick-of-2012	2011-11-21

URI	Publication date
http://seekingalpha.com/article/309466-7-stocks-showing-strong-resistance-poised-to-pop-in-the-next-few-days	2011-11-21
http://seekingalpha.com/article/309416-difficult-but-not-impossible-future-for-j-c-penny	2011-11-21
http://seekingalpha.com/article/310864-united-technologies-hidden-dividend-star	2011-11-30
http://seekingalpha.com/article/311700-conocophillips-great-dividend-with-solid-prospects-for-growth	2011-12-04
http://seekingalpha.com/article/311980-bank-of-america-is-a-clear-buy-now	2011-12-05
http://seekingalpha.com/article/311907-income-investor-it-s-time-to-buy-at-t	2011-12-05
http://seekingalpha.com/article/311981-for-activision-blizzard-it-s-game-on	2011-12-05
http://seekingalpha.com/article/311980-bank-of-america-is-a-clear-buy-now	2011-12-05
http://seekingalpha.com/article/311800-exxon-mobil-the-long-and-short-view-of-a-dividend-powerhouse	2011-12-05
http://seekingalpha.com/article/311759-inhibitex-22-for-shares-don-t-believe-the-hype-redux	2011-12-05
http://seekingalpha.com/article/311988-why-apple-s-cheap	2011-12-06
http://seekingalpha.com/article/312008-ford-and-sirius-profit-from-strong-automotive-trend	2011-12-06
http://seekingalpha.com/article/312015-new-dividend-kings-5-high-yield-stocks-for-2012	2011-12-06
http://seekingalpha.com/article/312024-2012-the-year-microsoft-finally-breaks-out	2011-12-06
http://seekingalpha.com/article/312520-my-favorite-dividend-growth-stock-in-the-oil-patch-chevron	2011-12-07
http://seekingalpha.com/article/312514-will-the-cloud-save-cisco	2011-12-07
http://seekingalpha.com/article/312513-darden-restaurants-offers-a-opportunity-for-long-term-investors	2011-12-07
http://seekingalpha.com/article/312464-mcclatchy-could-be-headed-to-zero	2011-12-07
http://seekingalpha.com/article/313460-apple-is-ripe-for-another-sell-off	2011-12-13
http://seekingalpha.com/article/313440-bearish-outlooks-for-alcoa-century-aluminum	2011-12-13
http://seekingalpha.com/article/313442-chipotle-mexican-grill-is-looking-too-inflated	2011-12-13
http://seekingalpha.com/article/314131-the-downside-to-pepsico-and-coca-cola	2011-12-15
http://seekingalpha.com/article/314375-cramer-s-black-list-5-stocks-to-avoid	2011-12-16
http://seekingalpha.com/article/314302-johnson-johnson-a-great-dividend-but-not-much-else	2011-12-16
http://seekingalpha.com/article/315161-united-technologies-the-search-for-rising-dividends	2011-12-20
http://seekingalpha.com/article/316113-7-reasons-to-buy-ibm	2011-12-27
http://seekingalpha.com/article/316620-dupont-is-a-great-buy-and-hold-forever-stock	2011-12-29
http://seekingalpha.com/article/316785-3m-headed-for-potential-breakout	2011-12-30
http://seekingalpha.com/article/316767-alcoa-your-best-stock-to-buy-for-2012	2011-12-30
http://seekingalpha.com/article/249650-bofa-valuation-confirms-its-current-bearish-inflection-point	2011-12-31
http://seekingalpha.com/article/1094421-cisco-at-8-month-high-now-what	2012-01-04
http://seekingalpha.com/article/317662-united-technologies-dividend-growth-analysis	2012-01-05
http://seekingalpha.com/article/1097951-chevron-lawsuit-will-weigh-heavily-on-long-term-value	2012-01-07
http://seekingalpha.com/article/318856-alcoa-is-your-best-buy-for-2012-after-earnings	2012-01-11
http://seekingalpha.com/article/319447-mcdonald-s-is-20-overvalued-investors-should-sell	2012-01-13
http://seekingalpha.com/article/513841-dividend-yield-levels-may-boost-procter-gamble-to-all-time-high	2012-01-20
http://seekingalpha.com/article/321137-intel-s-revenue-is-falling	2012-01-22
http://seekingalpha.com/article/353031-although-cheap-cisco-is-a-case-of-bad-capital-management	2012-02-09
http://seekingalpha.com/article/358371-4-high-yield-dividend-stocks-to-buy-1-to-avoid	2012-02-10
http://seekingalpha.com/article/366291-how-unitedhealth-won-the-health-it-race	2012-02-14
http://seekingalpha.com/article/373921-at-t-now-is-the-time-to-buy	2012-02-17
http://seekingalpha.com/article/378861-why-you-should-avoid-chevron-at-these-prices	2012-02-21
http://seekingalpha.com/article/390381-oil-is-surging-but-here-s-why-you-should-avoid-exxon	2012-02-24
http://seekingalpha.com/article/398771-procter-gamble-is-still-growing	2012-02-28
http://seekingalpha.com/article/414101-johnson-johnson-a-high-dividend-low-risk-stock	2012-03-06
http://seekingalpha.com/article/417401-will-verizon-shares-also-benefit-from-its-joint-venture-with-coinstar	2012-03-07
http://seekingalpha.com/article/424031-procter-gamble-buy-now-for-profits-in-2013	2012-03-09
http://seekingalpha.com/article/426971-johnson-johnson-a-solid-buy-candidate-for-2012	2012-03-12
http://seekingalpha.com/article/434741-hewlett-packard-a-blue-chip-in-transition-attractive-value-play	2012-03-14
http://seekingalpha.com/article/519801-capture-a-profit-on-boeing-s-slowdown	2012-03-23
http://seekingalpha.com/article/455481-hewlett-packard-now-a-solid-buy	2012-03-25
http://seekingalpha.com/article/458041-is-pfizer-pfinished	2012-03-26
http://seekingalpha.com/article/485561-pfizer-s-2012-cost-cutting-and-no-growth	2012-04-09
http://seekingalpha.com/article/505391-coca-cola-decent-results-do-not-make-a-screaming-buy	2012-04-18
http://seekingalpha.com/article/508101-it-s-too-late-to-buy-intel	2012-04-18
http://seekingalpha.com/article/521151-why-the-sell-off-in-wal-mart-is-overdone	2012-04-24
http://seekingalpha.com/article/536061-verizon-subscriber-growth-will-push-stock-above-41-by-2013	2012-04-27
http://seekingalpha.com/article/542041-procter-gamble-s-problems-go-beyond-costs	2012-04-30
http://seekingalpha.com/article/549971-watch-3m-as-it-moves-sideways	2012-05-02
http://seekingalpha.com/instablog/514543-timerfrank/581081-bearish-reversal-for-intel-corp-nasdaq-intc	2012-05-04
http://seekingalpha.com/article/560531-pfizer-alzheimer-s-drugs-will-carry-stock-to-new-highs-in-2013	2012-05-04
http://seekingalpha.com/article/560701-time-to-sell-in-may-and-buy-microsoft-at-31	2012-05-04
http://seekingalpha.com/article/560231-why-at-t-should-be-in-your-portfolio-now	2012-05-04
http://seekingalpha.com/article/564461-3m-impressive-product-portfolio-and-diversification-but-look-elsewhere-for-upside-potential	2012-05-07
http://seekingalpha.com/article/563991-4-healthcare-buys-some-undervalued-more-than-30	2012-05-07
http://seekingalpha.com/article/566641-intel-2-major-reasons-to-buy-now-for-gains-in-2013	2012-05-07

URI	Publication date
http://seekingalpha.com/article/566911-johnson-johnson-is-not-a-good-long-term-investment-in-2012	2012-05-07
http://seekingalpha.com/article/571541-time-to-diy-and-take-profits-in-home-depot	2012-05-08
http://seekingalpha.com/article/571321-united-technologies-still-bearish-a-little-longer	2012-05-08
http://seekingalpha.com/article/573341-4-reasons-merck-will-break-40-by-2013	2012-05-09
http://seekingalpha.com/article/574971-4-reasons-to-buy-boeing	2012-05-09
http://seekingalpha.com/article/578271-3-reasons-why-coca-cola-is-a-buy	2012-05-10
http://seekingalpha.com/article/577871-disney-when-you-wish-upon-a-star	2012-05-10
http://seekingalpha.com/article/578201-microsoft-presents-50-upside	2012-05-10
http://seekingalpha.com/article/582131-jpmorgan-the-worst-most-popular-stock	2012-05-11
http://seekingalpha.com/article/590341-4-reasons-hewlett-packard-is-worth-buying-near-52-week-lows	2012-05-15
http://seekingalpha.com/article/591271-microsoft-short-term-bear-long-term-bull	2012-05-15
http://seekingalpha.com/article/597631-home-depot-a-strong-buy-for-value-at-49	2012-05-17
http://seekingalpha.com/article/205720-home-depot-solid-dividend-stock	2012-05-18
http://seekingalpha.com/article/600891-jpmorgan-mess-will-drive-stock-below-28-this-summer	2012-05-18
http://seekingalpha.com/article/599591-new-signs-the-market-will-likely-remain-oversold-at-least-one-more-week	2012-05-18
http://seekingalpha.com/article/604171-chevron-could-tumble-5-on-brazilian-mess	2012-05-21
http://seekingalpha.com/article/606261-lowe-s-and-home-depot-could-fall-20	2012-05-21
http://seekingalpha.com/article/608531-jpmorgan-will-continue-to-tumble-on-2-billion-mess	2012-05-22
http://seekingalpha.com/article/613341-3-new-reasons-chevron-is-an-oil-and-gas-buy	2012-05-23
http://seekingalpha.com/article/615891-cisco-fall-in-stock-price-has-created-great-entry-point-for-long-term-investors	2012-05-24
http://seekingalpha.com/article/614551-verizon-short-term-sell-off-could-sink-stock	2012-05-24
http://seekingalpha.com/article/585251-why-an-opportunity-to-buy-chevron-at-90-per-share-could-be-coming	2012-05-24
http://seekingalpha.com/article/616801-johnson-johnson-could-slip-5-on-lower-margins	2012-05-25
http://seekingalpha.com/article/624081-report-card-exxon-mobil-and-selected-peers	2012-05-29
http://seekingalpha.com/article/622691-these-tech-stocks-could-sink-this-quarter	2012-05-29
http://seekingalpha.com/article/625521-hp-could-tumble-20-30	2012-05-30
http://seekingalpha.com/article/627211-ibm-20-hike-likely-by-2013	2012-05-30
http://seekingalpha.com/article/627741-4-reasons-investors-should-get-ready-to-buy-exxon-for-67-per-share	2012-05-31
http://seekingalpha.com/article/629701-no-news-is-good-news-for-wal-mart	2012-05-31
http://seekingalpha.com/article/631861-is-exxon-mobil-on-sale-below-80	2012-06-01
http://seekingalpha.com/article/631471-mcdonald-s-a-potentially-great-price-or-the-chance-to-double-your-yield	2012-06-01
http://seekingalpha.com/instablog/435182-john-mylant/693331-do-you-know-how-to-profit-on-chevron-in-a-bearish-market	2012-06-03
http://seekingalpha.com/article/633151-time-to-buy-alcoa	2012-06-03
http://seekingalpha.com/article/636421-strong-future-growth-prospects-for-general-electric-skepticism-misplaced	2012-06-04
http://seekingalpha.com/article/647621-market-not-lovin-mcdonald-s-month	2012-06-08
http://seekingalpha.com/article/653541-3m-s-dividend-is-overvalued-and-growth-is-risky	2012-06-12
http://seekingalpha.com/article/652481-coke-is-close-to-its-conservative-intrinsic-value	2012-06-12
http://seekingalpha.com/article/665501-the-magic-of-disney-as-an-investment-for-the-long-term	2012-06-18
http://seekingalpha.com/article/677341-buffett-doesn-t-like-procter-gamble-either	2012-06-22
http://seekingalpha.com/article/686091-avoid-these-4-dividend-stocks-exposed-to-the-strong-dollar	2012-06-27
http://seekingalpha.com/article/706311-ge-buy-on-strong-industry-average-growth	2012-07-06
http://seekingalpha.com/article/720611-how-jpmorgan-just-lost-a-huge-source-of-profits-now-a-terrible-investment	2012-07-14
http://seekingalpha.com/article/726101-the-slow-crawl-caterpillar-is-undervalued-but-not-an-obvious-buy-yet	2012-07-17
http://seekingalpha.com/article/728111-making-money-with-ibm	2012-07-18
http://seekingalpha.com/article/731531-why-windows-8-made-me-sell-microsoft	2012-07-19
http://seekingalpha.com/article/774641-dupont-poised-for-long-term-growth	2012-08-02
http://seekingalpha.com/article/784031-disney-a-good-buy-ahead-of-earnings	2012-08-06
http://seekingalpha.com/article/784091-microsoft-the-boring-returns-killer	2012-08-06
http://seekingalpha.com/article/800561-why-you-should-not-buy-coca-cola	2012-08-11
http://seekingalpha.com/article/825421-4-reasons-to-buy-hewlett-packard-shares-on-excessive-bearishness	2012-08-23
http://seekingalpha.com/article/826801-7-compelling-reasons-to-consider-jpmorgan-chase	2012-08-24
http://seekingalpha.com/article/828851-american-express-number-one-credit-card	2012-08-27
http://seekingalpha.com/article/833221-the-retirement-portfolio-american-express-best-in-class	2012-08-28
http://seekingalpha.com/article/841311-can-procter-gamble-save-10-billion	2012-09-01
http://seekingalpha.com/article/858851-pfizer-s-pipeline-could-push-stock-20-higher-by-2014	2012-09-11
http://seekingalpha.com/article/865671-timing-an-entry-point-for-coca-cola-important	2012-09-13
http://seekingalpha.com/article/868931-a-bullish-case-for-boeing	2012-09-15
http://seekingalpha.com/article/874671-general-electric-is-primed-for-a-leap	2012-09-18
http://seekingalpha.com/article/878321-3-reasons-why-you-should-buy-caterpillar-today	2012-09-20
http://seekingalpha.com/article/881171-expect-more-bearishness-from-boeing-in-the-short-term	2012-09-21
http://seekingalpha.com/article/882341-don-t-expect-dupont-to-rise-soon	2012-09-22
http://seekingalpha.com/instablog/743797-tomaspray/1108661-avoid-these-2-high-yield-tech-giants	2012-09-25

URI	Publication date
http://seekingalpha.com/article/889741-ibm-still-a-buy-with-a-new-chairman	2012-09-26
http://seekingalpha.com/article/908791-caterpillar-is-a-dividend-growth-stock-to-buy-at-this-time	2012-10-06
http://seekingalpha.com/article/910871-alcoa-declares-dividend-it-cannot-afford	2012-10-08
http://seekingalpha.com/article/912061-cisco-rebounds-with-a-strong-performance	2012-10-09
http://seekingalpha.com/article/916501-hp-a-bad-buy-as-profits-continue-declining	2012-10-10
http://seekingalpha.com/article/915861-wal-mart-strike-reminds-us-what-time-it-is	2012-10-10
http://seekingalpha.com/article/922171-avoid-verizon-due-to-overvaluation	2012-10-13
http://seekingalpha.com/article/922071-procter-gamble-is-severely-overvalued	2012-10-13
http://seekingalpha.com/article/922401-intel-is-the-investor-ready-to-dump-the-stock	2012-10-14
http://seekingalpha.com/article/922871-chevron-can-t-seem-to-turn-the-bad-news-faucet-off	2012-10-15
http://seekingalpha.com/article/924671-i-am-still-looking-for-more-bears-for-caterpillar	2012-10-15
http://seekingalpha.com/article/927391-3-reasons-to-buy-alcoa-as-analysts-see-50-upside-potential	2012-10-16
http://seekingalpha.com/article/930851-bank-of-america-s-financial-performance	2012-10-17
http://seekingalpha.com/article/932251-gas-drag-may-cause-exxon-mobil-to-sag	2012-10-18
http://seekingalpha.com/article/935991-4-reasons-to-sell-merck-now-and-buy-it-back-later-for-less	2012-10-19
http://seekingalpha.com/article/934461-why-I-re-still-huge-fans-of-intel	2012-10-19
http://seekingalpha.com/article/936911-general-electric-slowing-order-growth-will-limit-any-significant-uptick	2012-10-20
http://seekingalpha.com/article/936721-sell-bp-too-many-troubles-make-the-stock-a-value-trap	2012-10-20
http://seekingalpha.com/article/937361-hewlett-packard-too-speculative-for-my-taste	2012-10-21
http://seekingalpha.com/article/937301-why-stocks-like-verizon-and-at-t-are-due-for-a-significant-pullback	2012-10-21
http://seekingalpha.com/article/938511-avoid-these-5-dividend-stocks	2012-10-22
http://seekingalpha.com/article/940091-caterpillar-s-earnings-spell-bad-news-for-global-economy	2012-10-22
http://seekingalpha.com/article/938001-it-s-time-to-sell-these-4-surging-stocks	2012-10-22
http://seekingalpha.com/article/940511-microsoft-feeling-the-windows-pane	2012-10-23
http://seekingalpha.com/article/946721-at-t-s-q3-financial-performance	2012-10-24
http://seekingalpha.com/article/947581-forget-general-electric-limited-upside-with-a-15-downside	2012-10-24
http://seekingalpha.com/article/944761-the-dysfunctional-tech-giant-you-must-avoid-now	2012-10-24
http://seekingalpha.com/article/952101-be-aware-of-the-picture-ibm-s-sale-figures-are-painting-for-2013	2012-10-25
http://seekingalpha.com/article/952311-why-cisco-shares-could-become-the-next-tech-value-trap	2012-10-25
http://seekingalpha.com/article/954641-verizon-an-overpriced-telecom-to-avoid	2012-10-26
http://seekingalpha.com/article/960621-this-is-how-i-see-making-money-on-chevron-through-year-end	2012-10-30
http://seekingalpha.com/article/967431-i-expect-3m-is-not-finished-moving-down	2012-11-01
http://seekingalpha.com/article/980101-is-now-the-time-to-dump-this-telecom-giant	2012-11-05
http://seekingalpha.com/article/982251-intel-estimate-cuts-overvalue-stock	2012-11-06
http://seekingalpha.com/article/983801-pfizer-lipitor-vanishing-revenue	2012-11-06
http://seekingalpha.com/article/983621-will-legal-problems-bury-bank-of-america	2012-11-06
http://seekingalpha.com/article/986151-caterpillar-declining-order-book-and-negative-free-cash-flow	2012-11-07
http://seekingalpha.com/article/988751-disney-buy-espn-and-mickey-mouse-for-long-term-gains	2012-11-07
http://seekingalpha.com/article/992651-3-reasons-why-chevron-could-be-heading-back-to-52-week-lows	2012-11-08
http://seekingalpha.com/article/990941-expect-dupont-to-be-bearish-and-range-bound-through-mid-2013	2012-11-08
http://seekingalpha.com/article/996441-3-reasons-why-home-depot-shares-are-poised-for-a-sharp-drop-by-year-end	2012-11-09
http://seekingalpha.com/article/999351-pfizer-overvalued-at-current-price-levels	2012-11-12
http://seekingalpha.com/article/1018791-dr-buffett-has-diagnosed-johnson-johnson-as-too-big-to-succeed	2012-11-19
http://seekingalpha.com/article/1017811-why-is-mcdonald-s-falling-and-where-is-it-heading	2012-11-19
http://seekingalpha.com/article/1021631-home-depot-or-lowe-s-actually-neither	2012-11-20
http://seekingalpha.com/article/1020831-hp-s-whitman-running-out-of-excuses-and-running-out-of-time	2012-11-20
http://seekingalpha.com/article/1019701-the-amd-buyout-rumor-is-it-worth-speculating	2012-11-20
http://seekingalpha.com/article/1025201-caterpillar-is-still-a-buy-despite-being-downgraded-by-j-p-morgan	2012-11-23
http://seekingalpha.com/article/1028381-the-threat-to-disney-s-profit-train-is-sports	2012-11-26
http://seekingalpha.com/article/1035721-long-term-short-term-investing-advice-on-mcdonald-s	2012-11-29
http://seekingalpha.com/article/1040731-intel-hit-by-more-downgrades	2012-12-03
http://seekingalpha.com/article/1044311-3-reasons-why-intel-shares-could-be-headed-for-18-or-less	2012-12-04
http://seekingalpha.com/article/1043151-some-are-bullish-on-caterpillar-but-i-don-t-see-why	2012-12-04
http://seekingalpha.com/article/1044591-unitedhealth-group-the-time-to-buy-this-stock-is-now	2012-12-04
http://seekingalpha.com/article/1043061-walt-disney-is-a-must-buy-dividend-growth-stock	2012-12-04
http://seekingalpha.com/article/1050351-intel-margin-pressures-could-take-the-stock-lower-in-early-2013	2012-12-06
http://seekingalpha.com/article/1056861-where-are-the-intel-bulls	2012-12-11
http://seekingalpha.com/article/1063621-the-bear-case-for-hewlett-packard	2012-12-13
http://seekingalpha.com/article/1064231-bank-of-america-it-is-time-to-take-profits	2012-12-14
http://seekingalpha.com/article/1068381-accounting-for-hp-s-bad-valuations	2012-12-17
http://seekingalpha.com/article/1067981-cisco-dumps-linksys-still-stuck-in-telco-trap	2012-12-17
http://seekingalpha.com/article/1067771-intel-ignoring-the-facts-can-cost-investors-money-evLen-with-a-4-40-dividend-yield	2012-12-17
http://seekingalpha.com/article/1071891-the-financial-future-seems-bright-buy-boeing	2012-12-18
http://seekingalpha.com/article/1075861-2-drug-makers-i-ve-considered-shorting-in-the-near-term	2012-12-20
http://seekingalpha.com/article/1076601-what-s-worrisome-about-apple	2012-12-20
http://seekingalpha.com/article/1079271-occ-is-after-jpm-s-whale-as-trading-loss-swells-over-6-billion	2012-12-23

URI	Publication date
http://seekingalpha.com/article/1081961-wait-for-oil-to-drop-to-75-in-2013-before-buying-chevron	2012-12-26
http://seekingalpha.com/article/1084411-boeing-will-be-fine-going-into-2013	2012-12-27
http://seekingalpha.com/article/1087521-why-chevron-is-going-lower-in-2013	2012-12-31
http://seekingalpha.com/article/1089601-dupont-short-term-uncertainty-long-term-confidence	2013-01-01
http://seekingalpha.com/article/1095491-microsoft-seems-ripe-for-continued-pain	2013-01-04
http://seekingalpha.com/article/1096481-jpmorgan-chase-world-s-best-universal-bank	2013-01-06
http://seekingalpha.com/article/1101741-alcoa-kicks-off-earnings-season	2013-01-09
http://seekingalpha.com/article/1101841-procter-gamble-restructuring-priced-in-time-to-sell	2013-01-09
http://seekingalpha.com/article/1107391-3-reasons-to-sell-boeing-shares-while-it-trades-near-52-week-highs	2013-01-11
http://seekingalpha.com/article/1108881-buy-cisco-now-to-enjoy-healthy-gains-in-2013	2013-01-13
http://seekingalpha.com/article/1113421-why-it-s-too-early-to-buy-at-t-or-verizon-on-the-downgrades	2013-01-15
http://seekingalpha.com/article/1117161-boeing-gone-bad	2013-01-17
http://seekingalpha.com/article/1120111-at-t-uncertainty-and-red-flags-could-mean-a-buying-opportunity	2013-01-18
http://seekingalpha.com/article/1119511-intel-s-earnings-report-not-great-but-not-terrible	2013-01-18
http://seekingalpha.com/article/1123291-why-caterpillar-is-a-good-investment-now	2013-01-21
http://seekingalpha.com/article/1126331-johnson-johnson-a-stable-long-term-buy	2013-01-23
http://seekingalpha.com/article/1151971-why-chevron-is-good-for-the-long-term	2013-02-02
http://seekingalpha.com/article/1152791-disney-the-force-is-strong-with-this-one	2013-02-03
http://seekingalpha.com/article/1217011-alcoa-a-great-long-term-investment	2013-02-22
http://seekingalpha.com/article/1218391-intel-is-a-buy	2013-02-22
http://seekingalpha.com/article/1230771-home-depot-s-3-year-growth-potential	2013-02-27
http://seekingalpha.com/article/1250231-bank-of-america-buy-now-before-dividend-increases	2013-03-06
http://seekingalpha.com/article/204801-big-blue-has-investors-seeing-green	2010-05-12
http://seekingalpha.com/article/293840-mcdonald-s-a-stock-i-m-lovin	2011-09-15

Table 58: Seekingalpha blog documents annotated with multiple sentiment orientations with respect to different stocks.

URI	Publication date
http://seekingalpha.com/article/80393-pfizer-continues-to-fall-further	2008-06-06
http://seekingalpha.com/article/207120-verizons-data-business-benefits-from-android-overtaking-iphone	2010-05-26
http://seekingalpha.com/article/210772-is-microsoft-missing-the-ipad-curve	2010-06-18
http://seekingalpha.com/article/274086-4-big-sells-and-2-big-buys-by-bruce-berkowitz	2011-06-09
http://seekingalpha.com/article/275255-2-dow-stocks-to-avoid-and-2-to-obtain	2011-06-16
http://seekingalpha.com/article/275278-stock-wars-cisco-vs-intel	2011-06-16
http://seekingalpha.com/article/282199-5-stocks-under-20-that-should-trade-for-30	2011-07-27
http://seekingalpha.com/article/293078-home-depot-great-companies-don-t-necessarily-make-great-investments	2011-09-12
http://seekingalpha.com/article/295077-home-healthcare-stocks-one-to-buy-one-to-avoid	2011-09-21
http://seekingalpha.com/article/295079-smooth-sailing-for-carnival-cruise-lines	2011-09-21
http://seekingalpha.com/article/297183-5-must-own-energy-stocks-for-2012	2011-10-03
http://seekingalpha.com/article/298020-sell-apple-buy-sprint	2011-10-06
http://seekingalpha.com/article/312013-4-biotech-stocks-to-buy-1-to-avoid	2011-12-06
http://seekingalpha.com/article/312509-caterpillar-deere-will-both-break-100	2011-12-07
http://seekingalpha.com/article/312537-3-agriculture-stocks-to-buy-now-2-to-avoid	2011-12-07
http://seekingalpha.com/article/318824-3-nasdaq-stocks-to-avoid-1-to-buy	2012-01-11
http://seekingalpha.com/article/358401-big-pharma-3-to-avoid-1-to-buy	2012-02-10
http://seekingalpha.com/article/546551-my-favorite-super-major-oil-company	2012-05-01
http://seekingalpha.com/article/563321-review-of-analyst-downgrades-this-week-part-iv	2012-05-06
http://seekingalpha.com/article/572101-tactical-option-trades-for-the-cautious-investor	2012-05-08
http://seekingalpha.com/article/574571-james-barrow-top-bullish-picks-why-you-should-care	2012-05-09
http://seekingalpha.com/article/579311-2-credit-card-companies-to-buy-2-to-avoid	2012-05-10
http://seekingalpha.com/article/593531-boeing-not-a-value-play-but-lockheed-could-soar	2012-05-16
http://seekingalpha.com/article/598251-thursday-options-brief-cat-rrgb-roc	2012-05-17
http://seekingalpha.com/article/600021-pepsi-vs-coke-the-cola-wars	2012-05-18
http://seekingalpha.com/article/601371-why-cramer-sees-a-big-potential-drop-for-exxon-and-chevron	2012-05-18
http://seekingalpha.com/article/603251-hewlett-packard-75-profit-likely-by-mid-2014	2012-05-20
http://seekingalpha.com/article/605101-hewlett-packard-could-tumble-15-by-2013	2012-05-21
http://seekingalpha.com/article/605471-s-p-500-index-and-big-10-weekly-outlook-week-of-may-21st	2012-05-21
http://seekingalpha.com/article/624231-george-soros-top-holdings-buy-or-sell	2012-05-29
http://seekingalpha.com/article/629191-are-these-9-tech-blue-chip-stocks-worth-your-time	2012-05-31
http://seekingalpha.com/article/631401-bank-of-america-avoid-this-sinking-stock-now	2012-06-01
http://seekingalpha.com/article/635261-the-must-own-stock-for-a-natural-gas-rebound	2012-06-04
http://seekingalpha.com/article/639391-4-reasons-mcdonald-s-shares-might-keep-dropping-to-60-or-less	2012-06-05
http://seekingalpha.com/article/688001-2-dogs-of-the-dow-to-buy-1-to-avoid	2012-06-27

URI	Publication date
http://seekingalpha.com/article/878691-at-t-a-spring-loaded-stock-on-iphone-5-release	2012-09-20
http://seekingalpha.com/article/896281-a-close-look-at-home-depot-cautiously-optimistic-outlook	2012-09-30
http://seekingalpha.com/article/912991-johnson-johnson-the-goldman-downgrade-is-wrong-buy-the-dips	2012-10-09
http://seekingalpha.com/article/919381-buy-microsoft-now-or-wait-2-months	2012-10-11
http://seekingalpha.com/article/941951-tuesday-options-brief-ko-dd-armh	2012-10-23
http://seekingalpha.com/article/945211-3-stocks-to-protect-your-portfolio-now-and-2-names-to-avoid	2012-10-24
http://seekingalpha.com/article/947401-warning-signs-avoid-this-telecom-giant	2012-10-24
http://seekingalpha.com/article/1013011-4-reasons-to-buy-home-depot-now	2012-11-16
http://seekingalpha.com/article/1024591-6-gilded-stocks-to-avoid	2012-11-22
http://seekingalpha.com/article/1026491-3-high-quality-stocks-for-long-term-value-investing-and-3-to-avoid	2012-11-25
http://seekingalpha.com/article/1086451-buy-citigroup-sell-bank-of-america	2012-12-29
http://seekingalpha.com/article/1103361-earnings-beat-but-alcoa-is-still-a-nightmare	2013-01-09

Table 59: Stocks referenced by investor sentiment annotations in the entire corpus, totaling 638 documents. One reference is actually a stock index. The table also reports whether a stock was part of the DJIA stock index as of 2009-06-08 (according to (S&P Dow Jones Indices LLC, 2013)) in column DJIA. Furthermore, the number of blog documents with respect to a stock is reported in column Nod.

Stock	DJIA	Nod	Stock	DJIA	Nod
3M Company	y	10	Kinder Morgan Inc.	n	1
Activision Blizzard Inc.	n	2	Knight Capital Group	n	1
Advanced Micro Devices	n	1	Kohlberg Kravis Roberts & Co.	n	1
Alcatel-Lucent	n	2	LinkedIn Corp.	n	1
Alcoa Inc.	y	16	L'Oreal Group	n	1
Allianz AG	n	1	Lowe S Cos Inc.	n	5
Amazon.com Inc.	n	2	Mastercard Inc.	n	2
American Express Co	y	13	The McClatchy Company	n	1
Anheuser-Busch	n	1	McDonald's Corp.	y	17
Annaly Capital Management Inc.	n	1	Merck & Co Inc.	y	13
Apple Inc.	n	3	MGM Resorts International	n	1
Ariba	n	1	Microsoft Corp.	y	41
AT&T Inc.	y	13	Molycorp Inc.	n	1
Balchem Corp.	n	1	The Mosaic Company	n	1
Bank Of America Corp.	y	21	The Madison Square Garden Company	n	1
Barnes & Noble Inc.	n	1	Mylan Inc.	n	1
Beiersdorf	n	1	NeuroMetrix Inc.	n	1
The Boeing Company	y	18	Nokia Corp.	n	1
Bombardier	n	1	Novartis AG	n	1
BP plc	n	1	Novo Nordisk	n	1
Caterpillar Inc.	y	19	Oracle Corp.	n	1
Cemex	n	2	Pandora Media Inc.	n	1
Century Aluminum Company	n	1	PepsiCo Inc.	n	2
Chevron Corp.	y	19	Pfizer Inc.	y	17
Chipotle Mexican Grill Inc.	n	2	Phillips	n	2
Cigna Corp.	n	2	Procter And Gamble Co The	y	13
Cisco Systems Inc.	y	21	Provident Financial Services	n	1
Citigroup Inc.	n	1	Renren Inc.	n	1
Clearwire	n	1	Reynolds American Inc.	n	1
The Coca-Cola Company	y	10	Rio Tinto plc	n	1
ConocoPhillips	n	1	Rockwood Holdings Inc.	n	1
Danone	n	1	Royal Caribbean International	n	1
Darden Restaurants Inc.	n	1	SanDisk Corp.	n	1
Deere & Co	n	1	Sanofi-Aventis	n	2
Deutsche Bank AG	n	1	SAP AG	n	2
Dr Pepper Snapple Group Inc.	n	1	Siemens AG	n	1
Du Pont Ei De Nemours	y	9	Sirius XM Holdings Inc.	n	1
Exide Technologies	n	1	S&P 500	n	2
Extorre Gold Mines Ltd.	n	1	Sprint Nextel Corp.	n	1
Exxon Mobil Corp.	y	16	Stericycle Inc.	n	1
Fifth Third Bancorp.	n	1	SunTrust Banks Inc.	n	1
Ford Motor Co	n	2	Telefonica Brasil SA	n	2
General Electric Co	y	25	Tesoro Corp.	n	1

Stock	DJIA	Nod	Stock	DJIA	Nod
General Motors Company	n	1	Textron Inc.	n	1
Gentiva Health Services	n	1	Tootsie Roll Industries Inc.	n	1
Gilead Sciences Inc.	n	1	Total Sa	n	1
Goldman Sachs Group Inc.	n	1	Transocean Ltd.	n	3
Google Inc.-Cl A	n	1	The Travelers Companies Inc.	y	3
Green Mountain Coffee Roasters Inc.	n	2	Unilever plc	n	1
Groupon Inc.	n	3	United Technologies Corp.	y	10
Heineken	n	1	UnitedHealth Group Inc.	n	11
Hewlett-Packard Co	y	22	Universal Corporation	n	1
Home Depot Inc.	y	13	VeriFone Systems Inc.	n	1
Human Genome Sciences Inc.	n	1	Verizon Communications Inc.	y	19
Iconix Brand Group Inc.	n	1	Visa Inc.	n	2
Infineon Technologies AG	n	1	Vivendi	n	1
Inhibitex Inc.	n	2	Vonage Holdings Corporation	n	1
Insmed Inc.	n	1	Wal-Mart Stores Inc.	y	18
Intel Corp.	y	31	The Walt Disney Company	y	20
International Business Machines Corp.	y	20	WellPoint Inc.	n	2
Invesco Mortgage Capital Inc.	n	1	Wells Fargo & Co	n	3
Itron Inc.	n	1	Whirlpool Corp.	n	1
JC Penney Co Inc.	n	2	Whole Foods Market Inc.	n	1
Johnson & Johnson	y	11	Xyratex Ltd.	n	1
JPMorgan Chase & Co	y	18	Zillow Inc.	n	1

A.3 Investor Sentiment Classifier Implementation

The classifier designed in Section 3.2 uses a vector of numerically weighted features derived from textual blog documents by a document-vector transformation (see Section 3.2.1). The supervised machine learning algorithm is a linear SVM (see Section 3.2.2).

Table 60: Natural language processing resources used for classifying the sentiment orientation of investor sentiment in blog documents. The resources make up the classifier pipeline. All resources are part of GATE version 6.1 (Cunningham et al., 2011).

No	Processing resource	Task
1	ANNIE English Tokenizer	Identification of tokens, i.e., the basic units (e.g., words) of a textual document (Cunningham et al., 2011).
2	ANNIE Sentence Splitter	Identification of sentences that are required for the next step (Cunningham et al., 2011).
3	ANNIE POS Tagger	Identification of the parts of speech (e.g., verb, noun, etc.) of each token (Cunningham et al., 2011). The parts of speech are required for the next step (Cunningham et al., 2011).
4	GATE Morphological Analyzer	Identification of the roots (i.e., lemmas) (Cunningham et al., 2011).
5	Flexible Gazetteer	Identification of the stock for which a blog document was queried and retrieved using the labels provided in Table 61 (see Appendix A.4 for details).
6	Batch Learning using a Java version of LibSVM[3]	Performs (1) the document-vector transformation using the lemmas identified in a blog document, and (2) SVM-based training and application of a classifier of the sentiment orientation of the text of a blog document using the vector representation (Cunningham et al., 2011).

Each step described in Section 3.2 was realized by software tools in GATE[4] version 6.1 (Cunningham et al., 2011) called processing resources (see Table 60). GATE ("general architecture for text engineering") is a Java-implemented software framework for natural

[3] http://www.csie.ntu.edu.tw/~cjlin/libsvm/ (cited in (Cunningham et al., 2011, p.362))
[4] http://gate.ac.uk, retrieved 2014-05-16

language processing (Cunningham et al., 2011). Most of the processing resources used are part of the "ANNIE" information extraction system (Cunningham et al., 2011, pp.113–130), which is shipped with GATE. Together, the processing resources make up a classifier "pipeline" that executes the processing resources in the order given in Table 60. Default parameters of the processing resources were used except otherwise noted in Section 3.2 and 3.3.2.

A.4 Portfolio Simulation Dataset Retrieval

The datasets of investor sentiment (document scores) from blog documents used for this work's portfolio simulation (described in Section 4.1) were retrieved from the Seekingalpha and Blogspot websites using Google's Search API. Access to the API[5] was granted on the basis of the University Research Program for Google Search. This program was discontinued in 2012. Therefore, the documentation is no longer available from Google's website. However, it can be retrieved in the version of, e.g., 2011-10-16 from the Internet Archive[6]. The requests and retrievals were conducted throughout 2011 and partially in 2012. The requests to the API were restricted to a specific website, a specific stock, and a specific date in the period 2007-01-01 until 2011-12-31. Finally, requests were restricted to English language documents (by setting the URI-parameter "lr=lang_en").

The websites of Seekingalpha and Blogspot were specified by the respective URI in the search query string by the "site:URL"-part. The stock was specified by the search terms in Table 61 in the search query string. The date was specified in the search query by the "daterange:startdate-enddate"-part[7]. The enddate was specified to be the desired publication date of a blog document. The startdate was specified to be one day before the enddate as the date format uses the Julian calendar[7], which starts at 12:00 p.m. (i.e., noon)[8], thus a certain enddate would not include the first half of the day. The specified daterange actually refers to the time period Google added the blog document to its index. The publication date of a blog document is assumed to be the enddate used in the query for the search API in this thesis. This is considered a conservative estimate in this thesis because a blog document can be only crawled once it was published and the enddate is the most recent date in the crawl time period queried. For each day in the overall time period studied in this thesis in Section 4, a separate search request and retrieval of results was conducted.

The response to each request to the search API delivered usually multiple pages with 10 result links per page (like on the webpage www.google.com (retrieved 2014-05-16) on displaying search results for a user's query). All blog documents to which the result links of

[5] using the URI https://research.google.com/university/search/service? (followed by some parameters)
[6] under http://web.archive.org/web/20111016085811/http://research.google.com/university/search/docs.html, retrieved 2014-05-16
[7] http://jwebnet.net/advancedgooglesearch.html#advDateRange, retrieved 2014-05-16
[8] e.g., http://quasar.as.utexas.edu/BillInfo/JulianDatesG.html, retrieved 2014-05-16

the first 10 pages referred to were downloaded. In case less than 10 pages were returned by the API, all blog documents referred to in the result links were downloaded.

The article text (comprising the title, body, and publication date) of these blog documents was extracted by a boilerplate remover. For boilerplate removal, the Java library boilerpipe[9] was used in version 1.2.0 with the default configuration. Blog documents for which no article text could be extracted were discarded. Blog documents with exactly the same URI as one of the previously downloaded blog documents were also discarded.

The text of each downloaded blog document was verified to refer to the stock in the query. For this, a set of labels corresponding to each stock (see Table 61) was used respectively. At least one of those labels was required to appear in the text of a blog document. Other documents were discarded. If the file size of the text of a blog document exceeded 100 KB, it was also discarded.

Table 61: DJIA stocks and terms used to search for and retrieve blog documents via Google's Search API. The stocks constituting the DJIA are those of 2009-06-08 (according to (S&P Dow Jones Indices LLC, 2013)). The labels were used to verify that the blog documents actually refer to the respective stock.

Stock	Ticker symbol	Google API search term	Labels
3M Company	MMM	3M	"3M's", "3M Company", "3M Co (MMM)", "3M (MMM)", "3M (3M)", "MMM co", "3M Co", "MMM Company", "3M", "3M Company (MMM)", "3m (MMM)", "MMM", "Minnesota Mining and Manufacturing"
Alcoa Inc.	AA	Alcoa	"Alcoa's (NYSE: AA)", "Alcoa (NYSE:AA)", "Alcoa, Inc. (AA)", "Alcoa inc", "Alcoa's", "Alcoa (AA)", "Alcoa (NYSE: AA)", "Alcoa Inc.", "AA", "Alcoa, Inc.", "Alcoa Inc. (AA)", "Alcoa", "AA's", "Alcoa incorporated"
American Express Company	AXP	American Express	"American Express Co.", "American Express Co", "American Express (AXP)", "American Express", "AXP", "American Express Company", "Amex", "American Express Company (AXP)"
AT&T Inc.	T	AT&T	"AT&T incorporated", "(T)", "AT&T Inc. (NYSE:T)", "AT&T Inc.", "AT&T Mobility (NYSE:T)", "T", "AT&T Inc (T)", "AT&T's (T)", "AT&T Mobility", "ATnT", "AT&T Inc. (T)", "AT&T (T)", "AT&T inc", "AT&T", "AT&T Inc. (NYSE: T)"
Bank of America Corporation	BAC	Bank of America	"Bank of America, Inc (BAC)", "Bank of America Corp", "Bank of America Merrill Lynch (BAML)", "Bank of America (BAC)", "Bank of America (NYSE: BAC)", "Bank Of America", "Bank of America Corporation (BofA)", "BofAML", "B of A", "Bank of America (BofA)", "BAML", "BoA", "Bank of America, BofA", "Bank of America's (BAC)", "BankOfAmerica", "Bank of America Corporation (BAC)", "BAC", "Bank of Amerika Corporation", "Bank of America Corp.", "Bank of America Corp. (BAC)", "Bank of America Corporation", "BofA", "BofA's", "Bank of America", "Bank of America Merrill Lynch", "Bank of America's"

[9] see http://code.google.com/p/boilerpipe/, retrieved 2014-02-22

Stock	Ticker symbol	Google API search term	Labels
The Boeing Company	BA	Boeing	"Boeing Co (BA)", "Boeing", "Boeing (NYSE: BA)", "Boeing Co. (BA)", "Boeing Company (BA)", "Boeing Co.", "Boeing (BA)", "Boeing co", "BA", "(NYSE: BA)", "The Boeing Company (BA)", "Boeing's", "Boeing Company", "The Boeing Company", "Boeing company"
Caterpillar Inc.	CAT	Caterpillar	"CAT's", "Caterpillar Incorporated", "(NYSE: CAT)", "CAT", "Caterpillar's (CAT)", "Caterpillar Inc. (CAT)", "Caterpillar (CAT)", "Caterpillar", "Caterpilar Inc (CAT)", "Caterpillar Inc.", "Caterpillar's", "Caterpillar, Inc.", "Caterpillar Inc", "Caterpillar (NYSE: CAT)"
Chevron Corporation	CVX	Chevron	"Chevron Corp.", "Chevron Corp", "Chevron Corporation (CVX)", "Chevron's", "Chevron Corp. (CVX)", "CVX", "Chevron", "Chevron Corp.'s (CVX)", "Chevron (CVX)", "Chevron Corporation"
Cisco Systems Inc.	CSCO	Cisco	"Cisco's (CSCO)", "Cisco's (CSCO)", "CSCO", "Cisco Systems, Inc. (CSCO)", "Cisco Systems (Nasdaq: CSCO)", "Cisco Systems Inc. (CSCO)", "Cisco Systems, Inc.", "Cisco Systems inc", "Cisco Systems", "Cisco Systems incorporated", "Cisco Systems (CSCO)", "Cisco Systems Inc (CSCO)", "Cisco", "Cisco (CSCO)", "Cisco's", "Cisco Systems Inc."
The Coca-Cola Company	KO	Coca-Cola	"Coca Cola", "Coca Cola's (KO)", "Coca-Cola (KO)", "Coca Cola's", "Coca-Cola (NYSE: KO)", "Coca Cola (KO)", "Coke", "Coca-Cola Co. (NYSE: KO)", "The Coca-Cola Corporation (KO)", "Coca Cola Co", "Coca-Cola Company (KO)", "Coca-Cola Co", "Coca-Cola Co.", "The Coca-Cola Corporation", "The Coca-Cola Company (KO)", "Coca-Cola's", "Coke's", "KO", "Coke (NYSE: KO)", "Coca-Cola Company", "Coca-Cola", "Coca Cola Company", "(NYSE: KO)"
E.I. du Pont de Nemours and Company	DD	DuPont	"DD", "E. I. du Pont de Nemours and Co", "EI DuPont de Nemours & Co.'s (DD)", "DuPont (DD)", "EI DuPont de Nemours & Co. (DD)", "EI DuPont de Nemours & Co.'s", "EI DuPont de Nemours & Co.", "E. I. du Pont de Nemours and Co (DD)", "DuPont Co.", "Du Pont", "DuPont", "Du Pont De Nemours"
Exxon Mobil Corporation	XOM	Exxon	"Exxon Mobil Corporation (XOM)", "Exxon Mobil Corp", "Exxon-Mobil", "Exxon (XOM)", "Exxon Mobil Corporation (NYSE: XOM)", "Exxon Mobil (XOM)", "ExxonMobil", "Exxon Mobil Corporation", "Exxon Mobil", "Exxon-Mobil (XOM)", "ExxonMobil's", "ExxonMobil (NYSE: XOM)", "ExxonMobil (XOM)", "Exxon Mobil Corp. (XOM)", "Exxon Mobil Corp.", "Exxon", "XOM"
General Electric Company	GE	General Electric	"General Electric", "General Electric Co (GE)", "General Electric Company (GE)", "GeneralElectric", "GE", "General Electric Company", "General Electric Co", "GE (GE)", "GE's", "General Electric Co. (GE)", "General Electric (NYSE:GE)", "General Electric (GE)", "General Electric Co."
Hewlett-Packard Company	HPQ	Hewlett Packard	"Hewlett Packard Company", "Hewlett-Packard (HPQ)", "Hewlett-Packard's (HPQ)", "HPs", "HP (HPQ)", "Hewlett-Packard Co", "Hewlett-Packard's", "Hewlett-Packard Company (HPQ)", "HP's", "Hewlett-Packard", "HPQ's", "Hewlett Packard (HPQ)", "HP", "Hewlett-Packard Co. (NYSE: HPQ)", "Hewlett Packard (HP)", "HewlettPackard", "Hewlett-Packard Co.", "Hewlett Packard", "Hewlett-Packard Company", "HPQ", "Hewlett Packard Company (HPQ)", "HP's (HPQ)", "H-P's", "Hewlett-Packard Co. (HPQ)"

Stock	Ticker symbol	Google API search term	Labels
The Home Depot Inc.	HD	Home Depot	"Home Depot Inc", "HD's", "Home Depot (HD)", "The Home Depot", "The Home Depot, Inc (NYSE: HD)", "Home Depot Incorporated", "Home Depot, Inc.", "Home Depot, Inc. (HD)", "Home Depot", "The Home Depot (HD)", "HD", "The Home Depot, Inc. (HD)"
Intel Corporation	INTC	Intel	"Intel Corp. (INTC)", "Intel", "Intel's (INTC)", "INTC's", "Intel Corp.", "Intel's (INTC)", "Intel (NASDAQ:INTC)", "Intel (INTC)", "INTC", "Intel Corporation (NASDAQ:INTC)", "Intel Corporation (INTC)", "Intel's", "Intel Corporation", "Intel Corp (NASDAQ: INTC)", "Intel Corp", "Intel's (NASDAQ: INTC)"
International Business Machines Corporation	IBM	IBM	"IBM", "International Business Machines Co (IBM)", "International Business Machines", "IBM Corp.", "IBM (IBM)", "Big Blue", "Computing-Tabulating-Recording Co.", "International Business Machines (IBM)", "IBMs", "IBM's", "IBM Corporation", "International Business Machines Corporation (NYSE: IBM)", "IBM Corp. (IBM)", "International Business Machines Corp.", "IBM´s", "IBM Corp", "International Business Machines Corp (IBM)", "International Business Machines Corporation (IBM)", "IBM-NYSE", "International Business Machines Corp", "International Business Machines", "International Business Machines' (IBM)", "International Business Machines Corporation", "International Business Machines Corp. (NYSE: IBM)"
Johnson & Johnson	JNJ	Johnson & Johnson	"Johnson & Johnson", "J&J", "Johnson and Johnson (JNJ)", "JohnsonAndJohnson", "JNJ", "Johnson & Johnson (JNJ)", "Johnson & Johnson's", "Johnson & Johnson's (JNJ)", "Johnson & Johnson's", "Johnson & Johnson's (JNJ)", "Johnson&Johnson (JNJ)", "Johnson&Johnson", "Johnson and Johnson", "J&J's"
J.P. Morgan Chase & Company	JPM	JP Morgan	"J.P. Morgan Chase", "JPMorgan Chase & Co. (JPM)", "JP Morgan Chase & Company", "JPMorgan Chase", "J P Morgan Chase (JPM)", "J.P. Morgan Securities Inc.", "JPMorgan's", "JPMorgan Chase & Co (JPM.N)", "JPMorgan Chase(JPM)", "JPM", "JPMorgan (JPM)", "JPMorgan Chase (JPM.N)", "JPMorgan Chase & Co (JPM)", "JPM.N", "J.P. Morgan (JPM)", "JP Morgan Chase", "JP Morgan", "JPMorgan Chase (JPM)", "JP Morgan (JPM)", "JPMorgan Chase & Co.", "J.P. Morgan", "JP Morgan Chase & Co", "J P Morgan Chase", "JP Morgan Chase & Co.", "J.P. Morgan Securities LLC", "JP Morgan (NYSE: JPM)", "JP Morgan's", "JPMorgan", "JP Morgan's (JPM)", "JPMorgan Chase & Co. (NYSE: JPM)"
Kraft Foods Inc.	KFT	Kraft	"Kraft (KFT)", "Kraft", "Kraft Foods Inc (KFT)", "Kraft Foods Inc-Class A (KFT)", "Kraft Foods", "Kraft Foods Incorporated", "KFT", "Kraft Foods Inc"

Stock	Ticker symbol	Google API search term	Labels
McDonald's Corporation	MCD	Mc Donald's	"MCDonald's", "McDonald's Corp", "McDonald's Corp. (MCD)", "McDonalds (MCD)", "McDonald's Corporation", "Mickie D's", "McDonald", "McDonald's Corp.", "(NYSDE: MCD)", "McDonald's Corporation (MCD)", "McDonald's", "McDonalds Corp", "McDonald's Corp. (MCD)", "McDonald's (MCD)", "McDonalds", "MCD", "McDonald's (MCD)", "McDonald's (NYSDE: MCD)", "McDonald's Corporation (MCD)", "McDonald's", "McDonald's (NYSE:MCD)", "McDonald's Corp.", "McDonald's Corporation", "McDonald´s", "McDonalds Corporation", "MCD-NYSE", "MCD´s"
Merck & Company Inc.	MRK	Merck	"MRK", "Merck (MRK)", "Merck", "Merck & Co Inc.", "Merck & co", "Merck & Co. Inc. (MRK)", "Merck & co inc", "Merck & Co. (MRK)", "Merck (NYSE: MRK)", "Merck and co inc", "Merck and Company, Inc. (MRK)", "Merck's", "Merck and co", "Merck & Co. Inc.", "Merck and Company, Inc.", "Merck & co incorporated", "Merck & Co Inc. (MRK)", "Merck & Co.", "Merck & Co., Inc."
Microsoft Corporation	MSFT	Microsoft	"Microsoft Corporation", "Microsoft Corp", "Microsoft (Nasdaq: MSFT)", "MSFT's", "Microsoft (MSFT)", "Microsoft", "Microsoft's (NASDAQ:MSFT)", "Microsoft's", "Microsoft MSFT", "MSFT", "Microsoft's (MSFT)", "Microsoft Corporation (MSFT)", "Microsoft (Nasdaq:MSFT)"
Pfizer Inc.	PFE	Pfizer	"Pfizer Inc.", "Pfizer's", "Pfizer Inc. (PFE)", "Pfizer Inc", "Pfizer", "Pfizer's (PFE)", "Pfizer's (PFE)", "Pfizer (NYSE:PFE)", "Pfizer incorporated", "Pfizer (PFE)", "Pfizer Inc (PFE)", "Pfizer inc", "PFE"
The Procter & Gamble Company	PG	Procter & Gamble	"Procter&Gamble", "PG", "(NYSE: PG)", "P&G", "Procter & Gamble Co. (PG)", "Procter&Gamble (PG)", "Procter and Gamble", "Proceter and Gamble Co", "Procter & Gamble (PG)", "P&G (PG)", "Procter & Gamble", "P&G's", "Procter & Gamble (NYSE: PG)", "Procter and Gamble (PG)", "Procter & Gamble Co"
The Travelers Companies Inc.	TRV	Travelers companies	"travelers", "The Travelers Companies Inc. (TRV)", "The Travelers Companies Inc.", "travelers companies", "travelers cos incorporated", "travelers cos", "Travelers", "travelers cos inc", "TRV", "The Travelers Companies (TRV)", "The Travelers Companies", "Travelers'"
United Technologies Corporation	UTX	United Technolo-gies	"United Technologies Corporation (UTX)", "United Technologies", "United Technologies Corp (UTX)", "United Technologies (UTX)", "United Technologies Corporation", "United Technologies Corp", "UT", "United Technologies (NYSE:UTX)", "UTX"
Verizon Communica-tions Inc.	VZ	Verizon	"Verizon Communications Inc.", "Verizon's", "Verizon Wireless", "Verizon Communications Incorporated", "Verizon Communications' (VZ)", "(VZ)", "Verizon (NYSE: VZ)", "Verizon Communications (VZ)", "Verizon Communications Inc. (NYSE: VZ)", "Verizon Communications Inc (VZ)", "VZ", "Verizon Communications Inc", "Verizon Communications Inc. (NYSE:VZ)", "Verizon (VZ)", "Verizon Wireless (NYSE:VZ)", "Verizon (NYSE:VZ)", "Verizon Communications", "Verizon", "Verizon Communications Inc. (VZ)"

Stock	Ticker symbol	Google API search term	Labels
Wal-Mart Stores Inc.	WMT	Wal-Mart	"Wal-Mart Stores Inc", "Wal-Mart Stores, Inc. (NYSE: WMT)", "Wal-Mart's", "Walmart's", "Wal-Mart Stores, Inc.", "Walmart", "Wal Mart Stores", "Wal-Mart Stores (WMT)", "WI", "WalMart", "(WMT)", "Wal-Mart Stores Inc. (WMT)", "Wal Mart", "Wal-Mart Stores", "Wal-Mart U.S.", "WMT", "Walmart (WMT)", "Wal-Mart (WMT)", "Wal-Mart Stores Inc.", "Wal-Mart", "Wal-Mart International"
The Walt Disney Company	DIS	Disney	"Disney", "Walt Disney Co.", "The Walt Disney Company (DIS)", "Walt Disney Company (DIS)", "Walt Disney Co. (DIS)", "Disney's", "Walt Disney", "The Walt Disney Company (NYSE: DIS)", "Disney's (DIS)", "DIS", "The Walt Disney Co. (DIS)", "Walt Disney Co (DIS)", "Disney (DIS)", "The Walt Disney Co.", "Walt Disney Company", "The Walt Disney Company", "Walt Disney Co", "Walt Disney (DIS)"

A.5 Market Data

To estimate the Fama-French model (see Section 2.1.2.2) in Section 2.1.4, which presents an example on the existence of abnormal returns, daily frequency market data representing the three factors were retrieved[10] from Kenneth French's website[11]. The daily factors were derived by French from data from the 2014-12 CRSP database according the file "F-F_Research_Data_Factors_daily.txt" contained in the zip archive. Also the "simple daily rate" of the 1-month U.S. Treasury bill returns from Ibbotson and Associates contained in the same file was used as proxy of the risk free interest rate for estimation of the model according Definition (2.4). Daily log returns for the eToys stock were calculated from closing prices from Datastream.

To estimate Carhart's (1997) model (see Section 2.1.2.3) in Section 4.3, the respective monthly frequency market data representing the three factors of the Fama-French model, on which Carhart extends, were used. The monthly Fama-French factors were retrieved[12], containing factors based on data from the 2014-12 CRSP database according the file "F-F_Research_Data_Factors.txt" contained in the zip archive. The monthly Fama-French factors dataset includes a proxy of the monthly "risk free" rate, i.e., 1-month U.S. Treasury bill returns from Ibbotson and Associates[12], which was used in the estimation of Carhart's (1997) model – as proposed by (Carhart, 1997, p.61). The market proxy returns contained in the Fama-French factors dataset[12] are monthly frequency returns of a value-weighted U.S. stock market proxy[11]. The monthly momentum factor data used in Carhart's model was retrieved also from Kenneth French's website [13].

[10] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/ftp/F-F_Research_Data_Factors_daily.zip, retrieved 2015-01-24
[11] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html, retrieved 2015-04-01
[12] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/ftp/F-F_Research_Data_Factors.zip, retrieved 2015-01-24
[13] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/ftp/F-F_Momentum_Factor.zip, retrieved 2015-01-24

To obtain monthly log returns for each stock in each period of the portfolio simulation in Section 4, close prices adjusted for dividends and stock splits were used for each day a position was opened or closed in the portfolio simulation. The daily price data for all stocks in the portfolio simulation was sourced from Yahoo! Finance[14].

[14] http://finance.yahoo.com/, retrieved 2014-06-03