Dissertation

Investor Sentiment and Attention in Capital Markets

A (Social) Media Perspective

Quoc-Thai Ton

Dissertation

Vom Fachbereich Rechts- und Wirtschaftswissenschaften
Technische Universität Darmstadt (D17)
Zur Erlangung des akademischen Grades Doctor rerum politicarum (Dr. rer. pol.)

Investor Sentiment and Attention in Capital Markets

A (Social) Media Perspective

vorgelegt von

M. Sc. Quoc-Thai Ton
geboren am 21. August 1985 in Dillenburg, Deutschland

Tag der Einreichung: 11.09.2018
Tag der mündlichen Prüfung: 20.12.2018
Erstgutachter: Prof. Dr. Dirk Schiereck
Zweitgutachter: Prof. Dr. Alexander Benlian

Darmstadt 2018 (D17)

Ton, Quoc-Thai: “Investor Sentiment and Attention in Capital Markets – A (Social) Media Perspective”
Darmstadt, Technische Universität Darmstadt

Tag der Einreichung: 11.09.2018
Tag der mündlichen Prüfung: 20.12.2018

Darmstadt 2018 (D17)

Ich danke

meiner Familie

für ihren steten Glauben an mich und

meiner Karina und Ihrer Familie

für die bedingungslose Unterstützung zu jeder Zeit

Zusammenfassung

Die vorliegende Dissertation untersucht den Einfluss sozialer und traditioneller Medien auf den Kapitalmarkt. Im Vordergrund der empirischen Analysen steht hierbei die Stimmung der Investoren (nachfolgend Investor Sentiment genannt), die z. B. direkt durch Beiträge auf sozialen Plattformen, aber auch mithilfe innovativer Datenbanken und der digitalen Textanalyse traditioneller Printmedien gemessen werden kann. Mit dieser Ausrichtung werden implizit die Annahmen der traditionellen Finanztheorie in Frage gestellt und neue empirische Erkenntnisse mit den Erklärungsansätzen der verhaltensorientierten Finanztheorie (auch bekannt als Behavioral Finance) in Verbindung gebracht.

Eine der grundlegenden Elemente der traditionellen Finanzmarkttheorie stellt die Effizienzmarkthypothese dar. Die Verfügbarkeit von Informationen ist in dieser Betrachtung eine Grundvoraussetzung für die Funktionsfähigkeit effizienter Märkte. In solchen Märkten werden neue Informationen bezüglich einer Anlagemöglichkeit vom Kapitalmarkt schnell und exakt verarbeitet. Der neue Preis der Anlagemöglichkeit spiegelt damit jene neue Information und ihren fundamentalen Wert unmittelbar wider (Fama, 1969; 1970). Verschiedene Beobachtungen haben jedoch in der Vergangenheit gezeigt, dass Preisbewegungen am Kapitalmarkt nicht immer eindeutig auf rationale Informationen zurückzuführen sind. Über- und Unterreaktionen von Anlagepreisen auf Nachrichten oder ein vermeintliches Muster in vergangenen Gewinnentwicklungen schrieben dem Forschungszweig der verhaltensorientieren Finanztheorie seit den 1990ern daher eine zunehmende Bedeutung zu.

Eine wichtige Rolle spielte hierbei die sich verändernde Verfügbarkeit und der teilweise leichtere Zugang zu Informationen für sowohl institutionelle Investoren aber auch für Kleinanleger. Abbildung 1-1 (S. 3) beschreibt zum Beispiel die Reichweite traditioneller Printmedien zwischen 1970 und 2017 in den USA. Von mehr als 60 Millionen US-Haushalten in den 1970er sank die Zahl der mit traditionellen Printmedien erreichten US-Haushalte auf nur noch rund 30 Millionen in 2017. Dieser Trend ist jedoch nicht nur in den USA, sondern auch in anderen entwickelten Ländern zu beobachten. Auf der anderen Seite hat sich mit der Etablierung des Internets und der damit einhergehenden digitalen Dynamik ein paralleler Trend entwickelt. Abbildung 1-3 (S. 5) beschreibt die Entwicklung der weltweiten „Social Media“-Nutzer seit 2010.

Waren es 2010 noch rund 1 Milliarde globale Nutzer, werden bis 2021 rund 3 Milliarden Nutzer sozialer Netzwerke erwartet. Dies beeinflusst nicht nur die Gesellschaft, sondern auch eine spezielle Zielgruppe der vorliegenden Arbeit: die Investoren.

Die Art und Weise der Informationssammlung, -verarbeitung und -verbreitung hat sich mit den letztbeschriebenen Trends für Investoren in den letzten Jahrzehnten stark verändert (Puppis et al., 2017). So wird die Aufmerksamkeit für einzelne Anlagemöglichkeiten und das vorherrschende Investor Sentiment nachhaltig durch den vernetzten Einsatz digitaler Medien beeinflusst. Hieraus ergeben sich vier grundlegende Forschungsfragen, die auch die empirischen Analysen dieser Dissertation durchweg begleiten:

1. Welche Rolle nimmt Investor Sentiment in Kapitalmärkten ein? Spiegelt Investor Sentiment vergangene Marktdaten wider oder beeinflusst Investor Sentiment zukünftige Preisentwicklungen?
2. Welche Bedeutung haben (soziale) Medien für den Kapitalmarkt im alltäglichen Umfeld und in besonderen Unternehmenssituationen, wie beispielsweise Gewinn- oder Übernahmeankündigungen?
3. Welche Unternehmen sind besonders sensibel gegenüber Investor Sentiment?
4. Inwiefern stabilisieren Arbitragegeschäfte rationaler Investoren Kapitalmärkte in der Gegenwart von „Noise Tradern"?

Um diese Fragen bestmöglich beantworten zu können, ist die vorliegende Dissertation wie folgt strukturiert und aufgebaut: Das erste Kapitel leitet den Leser in die Relevanz der Thematik und die führenden Forschungsfragen der Dissertation ein. Das zweite Kapitel legt den theoretischen Grundstein dieser Arbeit und beschreibt grundlegende Ansätze der traditionellen als auch verhaltensorientierten Finanztheorie. Insbesondere werden verhaltensorientierte, theoretische Modelle beschrieben, die in Ihrer Gesamtheit ausgewählte Marktanomalien zu erklären zu versuchen. Zudem werden psychologische Ansätze und Konzepte dargelegt, die begründen warum Investoren teilweise irrational handeln und mit ihrem unbestimmten Handeln Unsicherheiten in den Markt tragen und Anlagepreise von ihrem inneren, tatsächlichen Wert abweichen lassen. Literaturüberblicke zum Thema Investor Sentiment im Bezug zu traditionellen Printmedien und sozialen Medien runden das Kapitel ab.

Das dritte Kapitel umfasst die erste empirische Arbeit dieser Dissertation und untersucht primär den Einfluss sozialer Medien auf den Kapitalmarkt. Hierfür greift diese Arbeit insbesondere auf mehr als 4,5 Mio. Beiträge zurück, die auf dem führenden, finanzbezogenen Internetforum HotCopper in Australien im Zeitraum zwischen Januar 2008 und Mai 2016 veröffentlicht wurden. Mithilfe umfassender empirischer Methoden (z. B. Eventstudien, vektorautoregressive Modelle, „Impulse-Response"-Funktionsanalysen oder multivariate „fixed-effects" Regressionen) zeigt diese Studie die Finanzmarktrelevanz von Beiträgen in sozialen Medien. Zum Beispiel zeigen die Ergebnisse, dass positives Investor Sentiment unmittelbar und signifikant mit abnormalen Unternehmensrenditen korreliert. Dieser Effekt ist jedoch nach einem Monat nicht mehr zu beobachten. Arbitragegeschäfte, die besonders von informierten Anlegern durchgeführt werden, wirken diesem Effekt gleichzeitig nur teilweise entgegen. Diese Beobachtungen deuten darauf hin, dass Internetbeiträge von Kleinanlegern den Markt kurzfristig beeinflussen, auch wenn die positiven Beiträge vermeintlich keine werthaltigen Informationen beinhalten und der Markt kurzzeitig auf diese Informationen überreagiert.

Die Folgen negativer Beiträge in Internetforen lassen jedoch andere Rückschlüsse zu. Negatives Investor Sentiment ist stark signifikant mit zukünftigen abnormalen Unternehmensrenditen im Zeitraum von einem Monat korreliert. Zudem deutet eine zunehmende Übereinstimmung von negativem Sentiment auf, über der Erwartung hinaus, negative Gewinnankündigungen hin. Beide Ergebnisse befürworten damit die Werthaltigkeit negativer Investorenbeiträge in sozialen Medien. Auf die Frage, inwiefern Beiträge in sozialen Medien sich auf Marktschwankungen auswirken, findet diese Arbeit mehrdeutige Ergebnisse. Insgesamt zeigen Granger-Tests und die Reaktionen einer „Impulse-Response"-Funktion eine bilaterale Beziehung zwischen Rendite-Volatilitäten und der Anzahl von Internetbeiträgen. Jedoch zeigt sich in diesem Zusammenhang, dass Kleinanleger in Internetforen mit ihrer Anzahl an Beiträgen stärker auf Marktschwankungen reagieren als umgekehrt. Zusammen zeigen die Ergebnisse der ersten empirischen Studie die ökonomisch, signifikante und asymmetrische Bedeutung von Investor Sentiment in sozialen Medien für den Finanzmarkt.

Im vierten Kapitel der Dissertation erweitern wir unsere empirischen Analysen und untersuchen den Einfluss traditioneller und sozialer Medien auf Kursreaktionen von Übernahmezielen vor offiziellen Übernahmeankündigungen. Die Literatur hat in der Vergangenheit oftmals bereits 2 Monate vor einer offiziellen Übernahmeankündigung einen Kursanstieg des Übernahmeziels beobachtet (z. B. Keown und Pinkerton, 1981). Dieses Phänomen wird auch als „Target Run-up“ bezeichnet. Erklärungen hierfür bieten zum einen die Insider-Hypothese (Insiderinformationen werden im Vorfeld ausgenutzt) oder die Markterwartungshypothese (auf Basis öffentlicher Informationen erwartet der Markt ein bevorstehendes Übernahmeangebot). Die zweite empirische Arbeit untersucht 2.765 Übernahmeangebote in Australien im Zeitraum von Januar 2008 und August 2015. Dabei greifen wir auf mehr als 15 Tsd. Zeitungsartikel, 80 Tsd. Beiträge im Internetforum HotCopper, diverse Analystenempfehlungen, sowie auf Daten bezüglich der relativen Suchhäufigkeit auf Google zurück, die in den Vorzeitraum der Übernahmeankündigung gefallen sind. Dadurch prüfen wir im Speziellen den unterschiedlichen Einfluss der Aufmerksamkeit unterschiedlicher Investorengruppen (institutionelle und individuelle Investoren) auf „Target Run-ups“.

Die Ergebnisse lassen den Rückschluss zu, dass Target Run-ups kleinerer Wachstumsunternehmen, die in der Vergangenheit operativ schwache Margen aufwiesen, im engen Zusammenhang mit Beiträgen auf dem sozialen Medium HotCopper stehen. Ähnliche kleinere Wachstumsunternehmen ohne Medienberichterstattung erfahren hingegen keinen signifikanten Anstieg der Aktienpreise im Vorfeld einer Übernahmeankündigung. Target Run-ups größerer Unternehmen sind dagegen besonders sensibel gegenüber Analystenempfehlungen. Diese Ergebnisse stehen mit der Beobachtung im Einklang, dass kleinere Unternehmen nicht durch Analysten gedeckt werden. Soziale Medien füllen in dieser Hinsicht diese Lücke. Google-Suchanfragen nach den Übernahmezielen zeigen auf der anderen Seite keinen ökonomisch signifikanten Zusammenhang. Insgesamt befürworten die Ergebnisse der zweiten empirischen Arbeit die Markterwartungshypothese. Soziale Medien tragen in dieser Hinsicht zu einer höheren Markteffizienz bei und füllen teilweise Informationslücken, die beispielsweise aufgrund ineffizienter Ressourcenverteilung oder zu kostenaufwendigen Recherchearbeiten existieren.

Das fünfte Kapitel schließt die empirische Arbeit der Dissertation ab und untersucht den Zusammenhang zwischen Stimmungen in Printmedien (News Sentiment nachfolgend) und Kapitalmärkten. Im „Asset Pricing“-Kontext prüfen wir den Einfluss aggregierter News Sentiment Indizes auf den Querschnitt von Unternehmensrenditen. Im Fokus der Asset Pricing-Literatur steht insbesondere die Ermittlung von Risikoprämien, die Unternehmensrenditen erklären sollen. Eine zentrale Frage dieser dritten empirischen Arbeit ist demzufolge, ob bestimmte Renditen im Zusammenhang mit dem eingebrachten Risiko stehen, oder jene Renditen als Resultat irrationaler Marktbewegungen hervorkommen. Mithilfe der Datenbank von RavenPack News Analytics berechnen wir auf Basis von mehr als 120 Millionen klassifizierten Nachrichtenartikeln im Zeitraum von 2000 und 2017 monatlich aggregierte News Sentiment Indizes und testen ihre Zusammenhänge mit Unternehmensrenditen. Dabei bilden wir in unserer Analyse monatliche „Zero-Investment“-Portfolios, die Unternehmen mit einem im Vormonat durchschnittlich positiven (negativen) News Sentiment kaufen (verkaufen). Die Ergebnisse zeigen, dass jene Portfolios eine jährliche Rendite von 7,5% erzielen, auch wenn wir in unseren Regressionen um die bekannten Risikofaktoren Markt, Unternehmensgröße, Momentum, Liquidität, Profitabilität und Investitionen kontrollieren. Die Resultate werden insbesondere vom positiven News Sentiment beeinflusst. Das hieraus resultierende Premium bezeichnen wir in dieser Arbeit als „premium on optimism“. Eine Begründung könnte in der generell positiven Berichterstattung im untersuchten Zeitraum liegen. Die Wahrscheinlichkeit, dass positive Nachrichten erscheinen, ist insbesondere höher, wenn in den vergangenen Monaten bereits positiv über jenes Unternehmen berichtet worden ist. Die Ergebnisse der dritten empirischen Studie sprechen insgesamt für die Sichtweise, dass News Sentiment als Risikofaktor angesehen werden kann.

Die Ergebnisse dieser Arbeit haben damit weitreichende Implikationen für Unternehmen, Investoren, Regulatoren sowie die weitergehende Forschung in diesem Umfeld. Unternehmen müssen in der heutigen Zeit lernen, Bewegungen in (sozialen) Medien frühzeitig zu antizipieren und mit vermeintlichen Falschmeldungen umzugehen. Eine stärkere inhaltliche und kommunikative Auseinandersetzung der Investor Relations-Abteilung mit diesem Thema könnte diesem Problem Rechnung tragen. Zudem können Unternehmen gezielte Kommunikationsstrategien für besondere Unternehmensereignisse entwickeln, um einer womöglich negativen öffentlichen Wahrnehmung frühzeitig entgegenzuwirken. Falschmeldungen und volatile Märkte sind zudem für Regulatoren von besonderer Bedeutung. Die Identifizierung von manipulativen Aktivitäten oder die Stabilisierung von Märkten im Umfeld mehrdeutiger Informationen sind hier vom speziellen Interesse. Gerade in aktuellen Zeiten der digitalen und vernetzten Kommunikation nimmt diese Aufgabe daher eine essentielle Rolle ein. Umso wichtiger ist daher das Verständnis um die Finanzakteure und deren Aktionen für einen effizienteren Kapitalmarkt. Die Ergebnisse dieser Arbeit stellen schlussendlich für die weitere Forschung neue Anknüpfungspunkte. Die asymmetrische Rolle von Investor Sentiment und deren dahinterliegenden Mechanismen sind weiterhin umstritten. Insbesondere bleiben aktuelle Forschungen bezüglich der Langzeitauswirkungen von direkt gemessenem Investor Sentiment Antworten schuldig. Diese Arbeit stellt daher mit ihren Ergebnissen eine fundierte Basis für zukünftige empirische Arbeiten. Auch konnte diese Arbeit nicht vollständig erklären, in welchen Situationen unterschiedliche Investorengruppen zu bestimmten Medien greifen und hierauf basierend Investitionsentscheidungen treffen. Eine Lösung könnte beispielsweise eine Intraday-Betrachtung unter Berücksichtigung einzelner Medieninstrumente darstellen.

Zusammenfassend zeigt diese Dissertation, dass Investor Sentiment ein wichtiger Bestandteil von heutigen Finanzmärkten geworden ist und ihre Bedeutung auch in der traditionellen Finanztheorie nicht mehr vernachlässigt werden darf.

Abstract

This dissertation examines the impact of social and traditional media on capital markets. The empirical tests focus on investor sentiment which, for example, can be captured by postings on social media platforms, innovative news databases and the textual analysis of traditional media press. The research direction of this dissertation implicitly questions the assumptions stated by the traditional finance theory. Our new empirical findings and their explanations are, hence, closely linked with the behavioral finance theory.

The Efficient Market Hypothesis constitutes one of the fundamental pillars of the traditional finance theory. In this concept, the availability of information is the basic requirement for the functionality of efficient capital markets. New information is quickly and correctly incorporated into an asset's price. The new price of an asset, therefore, immediately reflects the updated fundamental value (Fama, 1969; 1970). However, various studies have recently shown that stock market movements are not always associated with rational information about an asset's value. The observation of over- and underreaction of asset prices to news signals or distinctive return patterns gave reason for the gaining importance of the behavioral finance theory since the 1990's.

The changing availability and the easier access to information for institutional and individual investors play an important role in this recent development. For example, Figure 1-1 (p. 3) depicts the circulation of US newspapers between 1970 and 2017. The number of households covered by traditional media press decreased from more than 60 million to around 30 million households in 2017. The establishment of the internet, on the other hand, parallelly accelerated the digital development in the media landscape. Figure 1-3 (p. 5) describes the global development of social media users since 2010. The number of social media users is expected to increase from 1 billion users in 2010 to around 3 billion users in 2021. This development not only affects the society but also a specific focus group of this dissertation: the financial investors.

The way investors gather, process, and disseminate information also experienced a significant change in recent decades (Puppis et al., 2017). In this connection, the development of investor attention and sentiment for individual assets is sustainably impacted by the digitalization of media channels. Consequently, we derive for fundamental research questions, which accompany the empirical analyses of this dissertation:

1. What role does investor sentiment play in financial markets? Do investors solely follow the market, or do beliefs of investors predict future returns or other market variables?
2. How does (social) media relate to financial markets in the general daily context and specifically around news events, such as earnings or M&A announcements?
3. What kind of firms are more sensitive to investor sentiment than others?
4. Does arbitrage stabilize financial markets against noise traders?

The following structure of this dissertation aims to answer these questions in the best possible way: The first chapter introduces the reader to the relevance of the topic and the leading research questions of the dissertation. The second chapter lays the theoretical foundation and describes the fundamental concepts of the traditional and also the behavioral finance theory, which aims to comprehensively explain selected market anomalies. Also, we summarize selected psychological concepts that help to explain irrational actions of investors, which potentially cause market volatility and asset prices to deviate from their fundamental value. Literature reviews on investor sentiment in close relationship with traditional and social media complete the second chapter.

The third chapter encompasses the first empirical work of this dissertation and primarily explores the impact of social media on capital markets. The empirical analysis falls back to more than 4.5 million posts on the leading Australian financial internet message board HotCopper between January 2008 and May 2016. The findings suggest that social media activity is price relevant for capital markets. Positive investor sentiment, for example, is in this connection contemporaneously and significantly correlated with a stock's abnormal return. However, the effect diminishes after one month. Arbitrage of presumably informed investors only partially countervail this effect. Postings by individual investors on social media, hence, cause capital markets to overreact to potentially non-relevant information in the short-term.

However, negative investor sentiment expressed in internet message boards provides a differentiated picture. Negative investor sentiment is significantly related with the next month's abnormal returns. Also, an increasing rate of agreement on negative investor sentiment before

earnings announcements forecasts negative earnings surprises. Both findings support the information hypothesis that negative internet message board postings contain value-relevant information. The question whether social media activity induces market volatility remains ambiguous. The Granger-tests and the reactions of the impulse-response functions show a bilateral relationship between return volatility and the number of internet message board postings. However, we find in this context that individual investors react more sensitive to market volatility on social media than the other way around. In summary, the results of the first empirical work provide evidence for the economic significance of investor sentiment measured on social media and its asymmetric role in capital markets.

We extend the empirical analysis in the fourth chapter of this dissertation and investigate the impact of traditional and social media on target price run-ups before bid announcements. The literature previously documented an increase in the target stock price two months prior to the official bid announcement (e.g., Keown and Pinkerton, 1981). This phenomenon is also referred to as the target run-up. One group of researchers find explanations within the insider hypothesis (leakage of insider information prior to the bid announcement). Another group argues based on the market expectation hypothesis (the market anticipates publicly available information to predict upcoming mergers). Our second empirical work considers 2,765 bid announcements in Australia between January 2008 and August 2015. We use more than 15 thousand news articles, more than 80 thousand posts on the internet message board HotCopper, analyst recommendations, and Google search queries to analyze their relationship with target run-ups before official bid announcements. Thus, we specifically examine the varying impact of investor attention of different investor groups (institutional and individual investors) on target run-ups.

The results let us conclude that target run-ups of smaller, unprofitable, and growth firms are significantly related with social media coverage on HotCopper. On the contrary, similar firms that lack media coverage do not experience a significant target run-up prior to a bid announcement. Target run-ups of larger capitalization stocks are, on the other hand, more sensitive to analyst recommendations. The results are consistent with the anecdotal evidence that smaller firms are usually less covered by analysts. Social media closes the information gap for small firms in this perspective. Google search inquiries for target firms are not found to be significantly related to target run-ups. The overall findings of the second empirical work support the market expectation hypothesis. In this regard, social media contributes to the increase of market efficiency and partially closes informational blind spots for smaller firms which might exist due to inefficient allocations of resources or costly information sourcing for smaller firms.

The fifth chapter comprises the last empirical work of this dissertation and explores the relationship between media press sentiment and capital markets. We specifically examine the impact of aggregated news sentiment indices on the cross-section of returns in the asset pricing context. The literature around asset pricing especially focuses on the determination of risk premia that help to explain stock returns. A central question of our third empirical work is, therefore, whether stock returns are associated with their underlying risk or whether these returns are just a result of irrational market movements in the spirit of the behavioral finance theory. We calculate monthly aggregated news sentiment indices based on more than 120 million unique classified news articles from the Ravenpack News Analytics database between 2000 and 2017. Thus, we construct monthly zero-investment portfolios that go long on (sell) stocks which exhibit on average positive (negative) news sentiment in the previous month. The portfolio yields an annual return of 7.5% even if we control for widely-accepted risk factors, such as market, size, momentum, liquidity, profitability, and investments. The results are mainly driven by positive news sentiment. Hence, we refer this premium to the "premium on optimism". One possible explanation could be the persistent positive news coverage in the respective time period. The probability of the publication of good news is in particularly higher if a firm experienced positive news in the prior months. The total results of our third empirical work support the view that news sentiment reflects a risk factor.

The overall results of this dissertation have several implications for firms, investors, regulators and researchers in the field of behavioral finance. Firms must learn today to early anticipate crowd movements on (social) media and to deal with putatively fake news. The investor relations department of a firm must engage in this topic more sophistically content-wise and in the communicative interaction with its stakeholders. Selective communication strategies for specific firm events are required to early prevent a potentially negative public perception of the

firm. Fake news and volatile markets are also gaining in importance for regulators. The identification of manipulative activities or the stabilization of financial markets in the presence of ambiguous information is of special interest for regulators. This task is even more relevant in the time of increased digitalization of media channels and the networks behind them. The more important is, hence, a better understanding of the stakeholders in financial markets and their actions for the functionality of efficient markets. Finally, the results of this dissertation create new connection points for future research. The asymmetric role of investor sentiment and its underlying mechanism are still controversial and elusive. Current studies especially fail to shed light on the long-term impact of investor sentiment on capital markets. This dissertation, hence, provides a substantiated baseline for future empirical work. Also, this work could not fully answer the question in which situation investors specifically use different media channels for information sourcing and dissemination. An intraday-based analysis on various media channels could provide new answers to this question.

In summary, this dissertation shows that investor sentiment is an integral part of today's financial markets and its important role cannot be anymore neglected by advocates of the traditional finance theory.

Table of Content

Table of Content I
List of Figures IV
List of Tables.. V
List of Tables in Appendix..VII
List of Abbreviations and Frequently Used Symbols.. VIII
1. Introduction...1
2. From Traditional Finance Theories to Behavioral Finance 9
2.1. Traditional Finance Theory and Efficient Markets 10
2.2. Market Demand and Arbitrage of Rational Investors 12
2.3. Asset Pricing and Return Patterns 16
2.3.1. Asset Pricing Theory...16
2.3.2. Return Patterns 19
2.3.2.1. Patterns Based on Valuation Parameters 19
2.3.2.2. Patterns Based on Firm and Valuation Parameters....................................21
2.4. The Transition to Behavioral Finance 23
2.4.1. Limitations on Arbitrage in Imperfect Markets 24
2.4.2. Noise Trading in Efficient and Inefficient Markets 27
2.4.3. Investor Sentiment...29
2.5. Psychological Background on Investor Beliefs and Preferences 35
2.5.1. Beliefs...35
2.5.2. Preferences 39
2.6. Market Anomalies 41
2.6.1. Market Over- and Underreaction 43
2.6.2. Momentum 45
2.6.3. Excess Volatility and Risk 46
2.7. Information, Agents, and Media in Behavioral Finance 47
2.7.1. Analyst Coverage and Herding Behavior...50
2.7.2. Investor Sentiment and Traditional News Media...51
2.7.3. Increasing Social Interaction with Digitalization of Information Channels ..56
2.7.4. Investor Sentiment and Social Media...58
2.8. Summary...65

3. Agreeing to Disagree: Informativeness of Sentiments in Internet Message Boards 67
3.1. Introduction ... 67
3.2. Data and Research Design ... 72
3.2.1. Internet Stock Message Board Sentiment and Agreement ... 72
3.2.2. Financial Data ... 74
3.2.3. Sample Characteristics and Summary Statistics ... 77
3.3. Event Study ... 78
3.4. Predictability of Investor Sentiment for Abnormal Returns ... 83
3.4.1. Predictability of Message Board Sentiment and Abnormal Returns ... 92
3.4.1.1. Sentiment and Abnormal Returns at the Aggregate Level ... 92
3.4.1.2. Sentiment and Abnormal Returns at the Individual Level ... 93
3.4.2. Informed Short Selling Against Positive Noise Traders ... 100
3.4.3. Volatility and Message Board Activity ... 102
3.4.4. Predictability of Message Board Activity for Market Volatility ... 103
3.4.4.1. Message Volume and Volatility at the Aggregate Level ... 103
3.4.4.2. Message Volume and Volatility at the Individual Level ... 105
3.5. Fundamental Information in Message Boards around Company Events ... 107
3.5.1. Portfolio Analysis ... 111
3.5.2. Regressions on Earnings Surprises ... 113
3.6. Conclusion ... 116
4. Investor Sophistication and Attention in Target Price Run-Ups ... 118
4.1. Introduction ... 118
4.2. Data and Research Design ... 122
4.3. Event Study with M&A Announcements and Merger Signals ... 126
4.4. Multivariate Regression Analysis of Pre-Bid Run-Ups ... 133
4.5. Net Buyers or Sellers for Attention-Grabbing Events ... 137
4.6. Search Activity of Unsophisticated Individual Investors ... 139
4.7. Short-Term Target Markups around Bid Announcements ... 144
4.8. Discussion on Investor Behavior around Bid Announcements ... 146
4.9. Conclusion ... 149
5. A News Sentiment Risk Factor in the Mass Media Zoo ... 151
5.1. Introduction ... 151
5.2. News Media Tone and Investor Sentiment ... 153

5.3. Literature Review 156
5.3.1. Media Sentiment and Stock Returns 156
5.3.2. Sentiment and the Cross-Section of Stock Returns 158
5.4. Data Design and Descriptives 159
5.4.1. Sample Data 159
5.4.2. News Sentiment 161
5.5. RPNA Sentiment and US-Market Indices 164
5.6. News Sentiment and the Cross-section of Stock Returns 167
5.6.1. Return Differentials in Univariate Analysis 168
5.6.2. Factor Spanning Regressions for Long-Short Portfolios 171
5.6.3. Robustness Tests with Quintiles-Based Long-Short Portfolios 175
5.7. In Explanation for the News Sentiment Effect 179
5.7.1. Overreaction and Return Reversal 179
5.7.2. Limits in Arbitrage and Noise Trader Risk 183
5.7.3. Persistence of Sentiment and Crowd Momentum 185
5.8. Time Variation of Loadings and Systematic Risk 189
5.9. Conclusion 193
6. Overall Summary and Conclusion 194
References 204
Appendix 220

List of Figures

Figure 1-1. Daily Circulation of US Newspapers....3
Figure 1-2. Unique Visitors of US Newspaper Websites....4
Figure 1-3. Development of Worldwide Social Media Users (in Billion) 5
Figure 1-4. US-Users of Online Investing/Stock Trading Services Between 2008 and 2017.5
Figure 1-5. Unique Monthly HotCopper Users....6
Figure 1-6. Google Searches per Year....7
Figure 2-1. Overview of Theoretical Components of Behavioral Finance 9
Figure 2-2. Price Distribution of a Security 14
Figure 2-3. Theoretical Effects of Investor Sentiment on Stocks....33
Figure 2-4. Utility (Value) Function....40
Figure 2-5. Beliefs, Preferences and Market Anomalies....42
Figure 2-6. Relationship between Media, Analysts, Firms, and Investors....48
Figure 2-7. Classification of Social Media....57
Figure 3-1. Message Postings (in thsd.): 2003-2008 vs. 2008-05/2016....72
Figure 3-2. Abnormal Message Board Activity (min. 10/20 Buy/Sell Messages)....79
Figure 3-3. Abnormal Message Board Activity (min. 30/40 Buy/Sell Messages)....80
Figure 3-4. Impulse-Response-Functions – Abnormal Returns (Impulse : Response)....99
Figure 3-5. Impulse-Response-Functions – Volatility (Impulse : Response) 107
Figure 3-6. Average Cumulated Agreement Index Relative to the Event Date t 111
Figure 3-7. Average Unexpected Earnings (SUEAF) by Agreement/Sentiment Quartiles 112
Figure 4-1. Average CARs around M&A Announcements 128
Figure 4-2. Average CARs around M&A Announcements – Change of Control....132
Figure 4-3. Google Ticker Search Volume around M&A Announcements....142
Figure 4-4. Abnormal Google Ticker Search Volume around M&A Announcements....142
Figure 5-1. Impact of Media Tone on Stock Prices....156
Figure 5-2. Aggregated News Sentiment and Index Developments 165
Figure 5-3. Time Horizon Analysis of News Sentiment Effect – Long-Short Portfolio....182
Figure 5-4. Time Horizon Analysis of News Sentiment Effect – Long and Short Legs....182

List of Tables

Table 2-1. Proxies for Investor Sentiment 34
Table 2-2. Literature on News Media and Sentiment 52
Table 2-3. Relation between Social Media and Financial Market Activity 60
Table 3-1. Example of HotCopper Messages 75
Table 3-2. Summary Statistics: On Firm/Trading Level 76
Table 3-3. Event Study Results: “Buy”-Events 81
Table 3-4. Event Study Results: “Sell”-Events 82
Table 3-5. Message Board Variables and Abnormal Returns – All Stocks 86
Table 3-6. Message Board Variables and Abnormal Returns – Only Large Stocks 87
Table 3-7. Message Board Variables and Abnormal Returns – Only Small Stocks 88
Table 3-8. Message Board Variables and Abnormal Returns – All (Min. 10 Mes.) 89
Table 3-9. Message Board Variables and Abnormal Returns – Large (Min. 10 Mes.) 90
Table 3-10. Message Board Variables and Abnormal Returns – Small (Min. 10 Mes.) 91
Table 3-11. Lag-Order Selection Statistics for VAR – Aggregate Level 93
Table 3-12. Vector Autoregressions at Aggregate Level – Abnormal Returns 94
Table 3-13. Panel Vector Autoregression at Individual Level – Abnormal Returns 95
Table 3-14. Short Selling and Sentiment Regressions 101
Table 3-15. Regressions on Volatility 102
Table 3-16. Lag-Order Selection Statistics for VAR – Aggregate Level 103
Table 3-17. Vector Autoregression at Aggregate Level – Volatility 104
Table 3-18. Panel Vector Autoregressions at Individual Level – Volatility 106
Table 3-19. Summary Statistics 110
Table 3-20. Message Board Activity as Predictor of Earnings Surprise (SUEAF) 114
Table 3-21. Message Board Activity as Predictor of Earnings Surprise (SUEHIST) 115
Table 4-1. M&A Announcement Sample Distribution 124
Table 4-2. Financial Descriptives of Sample 125
Table 4-3. Average and Median CAR-Differences between Subsamples 127
Table 4-4. Average CARs around M&A Announcements 129
Table 4-5. Median CARs around M&A Announcements 131

Table 4-6. Fixed-Effects Regressions on Target Price Run-Ups..135
Table 4-7. Regressions on Target Run-Ups including Ownership Disclosure....................138
Table 4-8. Google Descriptives 142
Table 4-9. Fixed-Effect Regression on Target Price Run-Ups – Google Merger Signals ..143
Table 4-10. Target Markups around Bid Announcements 145
Table 4-11. Timing of First Media / Internet Attention 147
Table 4-12. Comparison of Attention Across Media and Internet Outlets..........................148
Table 5-1. Summary Statistics for Total RPNA Stories...161
Table 5-2. Summary Statistics for Media Sentiment...163
Table 5-3. Correlation between Sentiment and Market Variables 166
Table 5-4. LAG-Order Selection Statistics for VAR (Russell 3000).................................166
Table 5-5. Excess Market Return Prediction with Aggregate Sentiment Score.................167
Table 5-6. Stock Returns and Sentiment: Predictive Return Differentials..........................169
Table 5-7. Spearman Rank Correlations of Risk Factors 172
Table 5-8. Factor-Based Time-Series Regressions for Zero-Investment-Portfolios 173
Table 5-9. Spanning Regression by Firm Characteristics 174
Table 5-10. Stock Returns and Sentiment: Predictive Return Differentials by Quintiles ...177
Table 5-11. Time Series Regressions for Zero-Investment-Portfolios by Quintiles 178
Table 5-12. Variation of Formation and Holding Periods..181
Table 5-13. Stock Returns and Sentiment: Short Sale Volume..184
Table 5-14. Spanning Regression for Short Sale Volume Subsample 185
Table 5-15. Unconditional Sentiment Transition Matrices 186
Table 5-16. Conditional Sentiment Transition Matrices 188
Table 5-17. Spanning Regressions in Time Periods...191
Table 5-18. Fama-MacBeth Regressions: Cross Section of Individual Stock Returns.......192

List of Tables in Appendix

Table A-1. Summary Statistics on Firm/Trading Level - 5 Days Formation Period..........220
Table A-2. Summary Statistics on Firm/Trading Level - 10 Days Formation Period.........221
Table A-3. Summary Statistics on Firm/Trading Level - 30 Days Formation Period.........222
Table A-4. Summary Statistics on Firm/Trading Level - 60 Days Formation Period.........223
Table A-5. Lag-Order Selection Statistics for VAR – Individual Level............................223
Table A-6. Summary Statistics..224
Table A-7. Summary Statistics..224
Table A-8. Message Board Activity as Predictor of Earnings Surprise (SUEAF).............225
Table A-9. Message Board Activity as Predictor of Earnings Surprise (SUEHIST)..........226
Table A-10. Average CARs Around Bid Announcements – Takeover Subsample............227
Table A-11. Description of Main Variables...228
Table A-12. Pearson Correlation for Factors..229
Table A-13. Stock Returns and Sentiment: Predictive Return Differentials.......................230
Table A-14. Factor Time Series Regressions - Only Positive Sentiment...........................231
Table A-15. Factor Time Series Regressions - Only Negative Sentiment.........................232
Table A-16. Time Series Regressions by Firm Characteristics...233
Table A-17. Stock Returns and Sentiment: Predictive Return Differentials by Quintiles ..233

List of Abbreviations

AAM	Alliance for Audited Media
AAR	Average abnormal returns
ACAR	Average cumulative abnormal returns
AMEX	American Stock Exchange
AR	Abnormal return
ASIC	Australian Securities and Investments Commission
ASX	Australian Securities Exchange
CAPM	Capital Asset Pricing Model
CAR	Cumulative abnormal return
CMA	Conservative-Minus-Aggressive
CSS	Composite sentiment score
DJIA	Dow Jones Industrial Average
EPS	Earnings per share
ETF	Exchange-traded fund
FCA	Financial commentary and analysis
FF	Fama-French
HC	HotCopper
HML	High-Minus-Low
I/B/E/S	Institutional Brokers' Estimate System
IMB	Internet message boards
IPO	Initial public offering
LIQ	Liquidity
LTM	Last twelve months
M&A	Merger and Acquisitions
MBP	Microblogging platforms
MOM	Momentum
NASDAQ	National Association of Securities Dealers Automated Quotations
NYSE	New York Stock Exchange
P/E	Price/Earnings

PNM	Positive-Minus-Negative
PS	Pastor-Stambaugh
RMW	Robust-Minus-Weak
RPNA	Ravenpack News Analytics
S&P	Standard & Poor’s
SEC	Securities and Exchange Commission
SIRCA	Securities Industry Research Centre of Asia-Pacific
SMB	Small-Minus-Big
SUE	Standardized unexpected earnings
SUEAF	Standardized unexpected earnings surprise based on analyst forecasts
SUEHIST	Standardized unexpected earnings based on historical time series information
TRNA	Thomson Reuters News Analytics
US	United States of America
VMA	Vector moving-average

Investor Sentiment and Attention in Capital Markets

A (Social) Media Perspective

1. Introduction

This dissertation aims to shed light on the elusive link between the traditional finance theory on efficient markets and well-documented market anomalies, such as stock price over- and underreaction to news signals, propagated by advocates of behavioral finance. In this dissertation, behavioral finance theory complements the traditional perspective in which the market processes information quickly and efficiently. As financial research comprehensively examines theories and concepts about return patterns, risk-adjusted asset pricing models, and market anomalies, no behavioral research stream could definitely claim to explain the variances of future returns unambiguously.

Within the connection of efficient market theories and the concept of behavioral finance, this dissertation mainly attempts to address three distinctive goals in different empirical setups. First, we evaluate how investor sentiment disseminated on internet message boards convey value-relevant or noisy information in financial markets. By extending existing literature but also applying new approaches and broader test samples, we test the informativeness of the crowd in a dynamic digital environment. Second, we empirically evaluate whether media attention of financial investors is related to wealth effects associated with merger and acquisition announcements (M&A or bid announcements forth on). The combination of distinctive media channels (incl. traditional media press, social media and internet search queries) allows us to differentiate the impact of attention spent by different investor types, such as (un-)sophisticated individual and institutional investors, on financial market activity.

Third, we extend the traditional asset pricing theory by behavioral elements and introduce a news sentiment measure for investor sentiment, which presumably captures a behavioral risk-factor in the theory of capital asset pricing.[1] In this connection, prior literature mainly relied on investor sentiment proxies, constructed from market output variables due to only limited possibilities to directly measure investor sentiment in a real-time setting. Hence, the application of self-disclosed and extracted investor sentiment from social media platforms and media press releases enables us to analyze direct measures of investor sentiment and its relation to financial markets. The previously mentioned goals of this dissertation follow four specific research questions, which are all somehow addressed throughout the dissertation:

1. What role does investor sentiment play in financial markets? Do investors solely follow the market or do beliefs of investors predict future returns or other market variables?
2. How does (social) media relate to financial markets in the general daily context and specifically around news events, such as earnings or M&A announcements?
3. What kind of firms are more sensitive to investor sentiment than others?
4. Does arbitrage stabilize financial markets against noise traders?

These are only a few but the most important questions this dissertation aims to answer.

The basic foundation of the traditional finance literature had been established in the 1950s and 1960s with the seminal work on portfolio theory by Markowitz (1952) and pioneering, theoretical models on asset pricing (e.g., Sharpe, 1964; Lintner, 1965). In all these theories, the access to information is elementary for the functionality of efficient markets (Fama, 1970). The general efficient market theory asserts that available information about an asset is quickly processed by financial markets and thus reflected in its returns. However, the availability and accessibility to new (it is the question whether this information is found to be exact) information have changed rapidly with new technologies and the digitalization of the media. For example, Figure 1-1 demonstrates meticulously, how traditional and especially printed newspaper circulation diminished dramatically in the United States with the rise of the internet.

[1] Traditional asset pricing models refer, for example, to risk premiums associated with market risks, size effects (smaller firms generate higher returns compared to larger firms), or valuation effects (value stocks experience higher returns than growth firms). See Fama and French (1993).

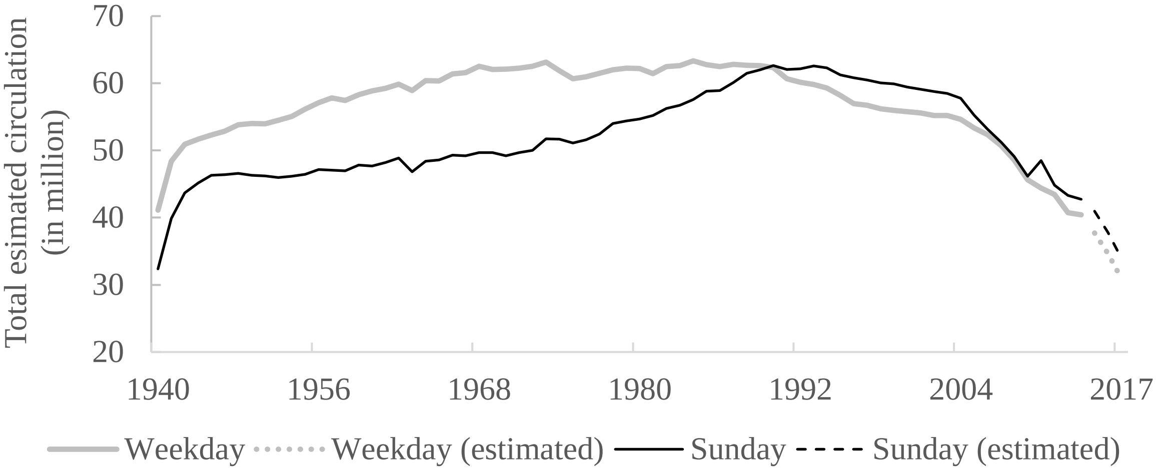

Figure 1-1. Daily Circulation of US Newspapers

This figure describes the total circulation (print and digital) of US newspapers in week- and Sundays. The data only includes newspapers that report numbers to the Alliance for Audited Media (AAM). Source: http://www.journalism.org/fact-sheet/newspapers/

In 2017, only 31 to 34 million US newspapers were circulated daily compared to the peak of around 63 million in the 1970's. On the other hand, digital coverage of US newspapers gauged from 8 to nearly 12 million unique daily visitors only in the period between 2014 and 2017 as shown in Figure 1-2. This corresponds to an increase of more than 40% of monthly unique visitors of US newspaper websites and should not only reflect a local but global development. As a consequence, the mechanisms of information dissemination not only changed for the broader population but also for a distinctive group of interest for this dissertation: the financial investors.

Nowadays, individual but also institutional investors gather information from internet search queries, have easier online access to financial databases, and exchange investment opinions or results on financial analysis on online investment platforms. Figure 1-3 depicts the global development of general social media users between 2010 and 2021. During this period, it is expected that the number of social media users will triple from 1 to 3 billion users worldwide, yielding an annual compounded growth rate of 11%. This trend not only reflects the general dynamics of social interaction but also points to the conclusion that financial markets are also affected by the connectivity of its users. To further underpin the importance of social media in

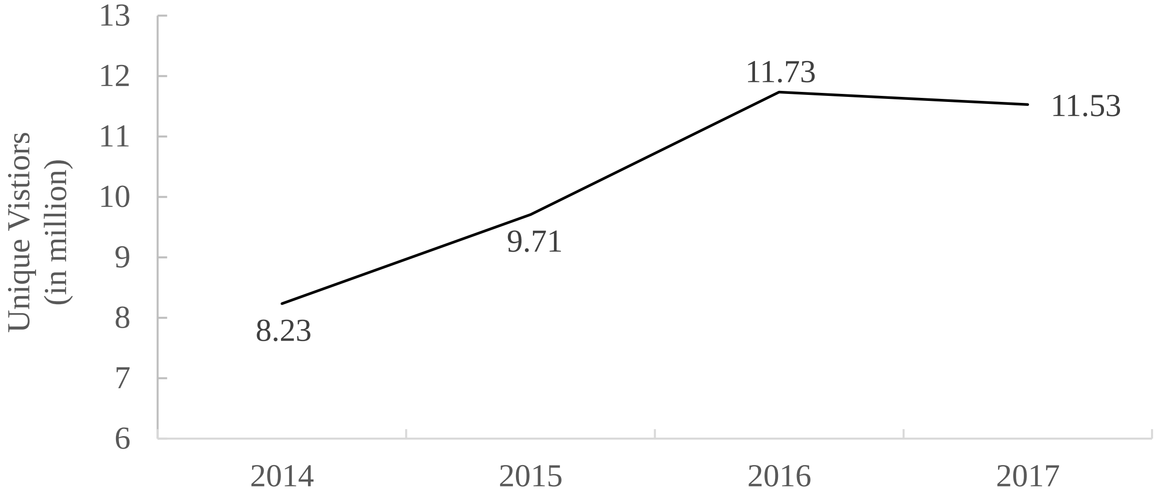

Figure 1-2. Unique Visitors of US Newspaper Websites

This figure presents the average monthly unique visitors for the top 49 US newspapers between 2014 and 2017. The average is based on the period between October and December. Source: http://www.journalism.org/fact-sheet/newspapers/

financial markets, we show the number of US people using online investing or stock trading services in Figure 1-4. Starting from 11.6 million people in the US in spring 2008, the number increased by almost 50% to 15.8 million in summer 2017, corresponding to a compound annual growth rate of 3%. Another global example of the increasing number of financial investors entering financial, social media platforms is shown in Figure 1-5. HotCopper is one of the leading financial internet message boards embedded in the highly regulated financial market of Australia. Internet board discussions on stocks, derivates or foreign exchanges regionally focus on Australia but also cover international markets. HotCopper was able to triple its unique monthly visitor number from 200 to nearly 600 thousand between June 2014 and July 2016. We will further explain the relationship of internet message board activities this board and the Australian financial market in section 3. Another trend, which we mentioned earlier, is the facilitated access to information in particular for individual (or retail) investors. Internet search engines enable its users to quickly and efficiently find information within a short amount of time. In 2012, Google recorded more than 1.2 trillion internet search queries on the global level (see Figure 1-6). All the presented numbers comprehensively underline the increasing importance of (social) media and internet trends and its sustainable impact on financial markets.

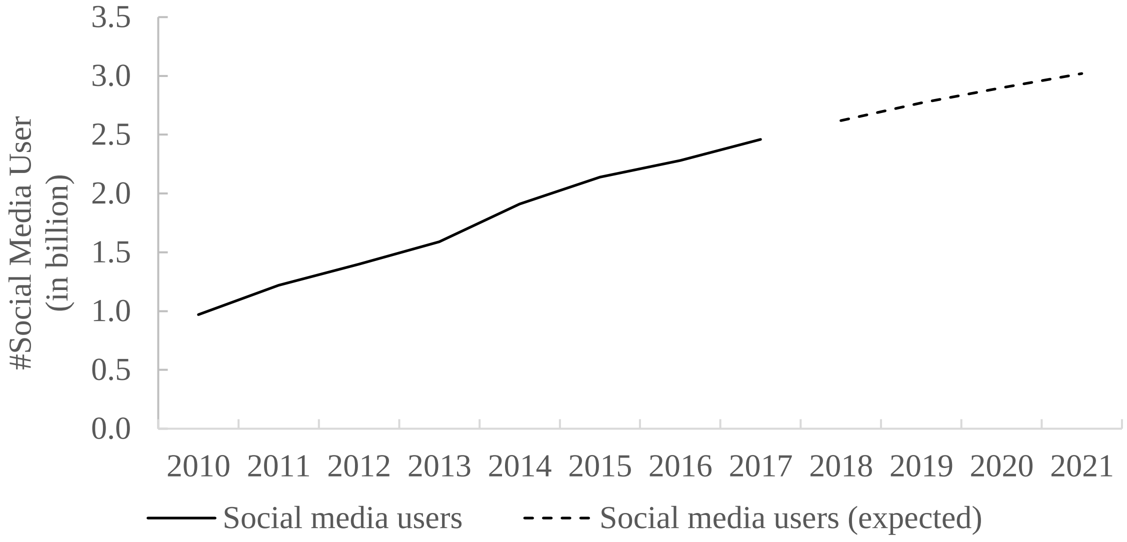

Figure 1-3. Development of Worldwide Social Media Users (in Billion)

The figure depicts the number of worldwide social media users between 2010 and 2021 with projections starting in 2017. Internet users who access a social media site at least once a month are considered for this statistic. Source: Worldwide, eMarketer on https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/

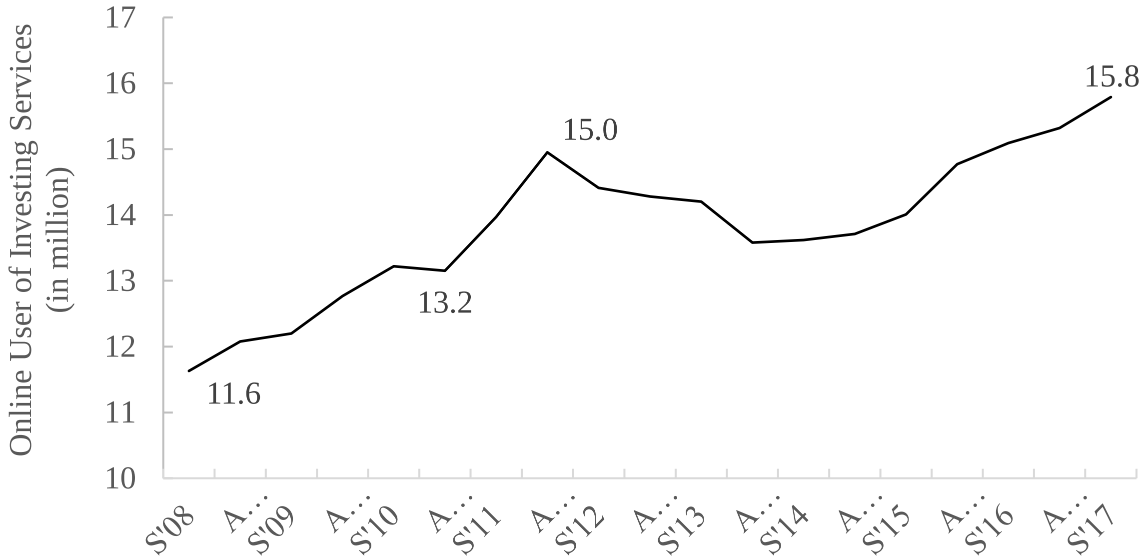

Figure 1-4. US-Users of Online Investing/Stock Trading Services Between 2008 and 2017

This figure describes the number of people living in US households who used an online investing/stock trading service in the last 12 months in the period between 2008 and 2017 (S=Spring, A=Autumn). Source: Nielsen Scarborough on https://www.statista.com/statistics/228118/people-in-households-with-an-online-investing-stock-trading-service-usa/

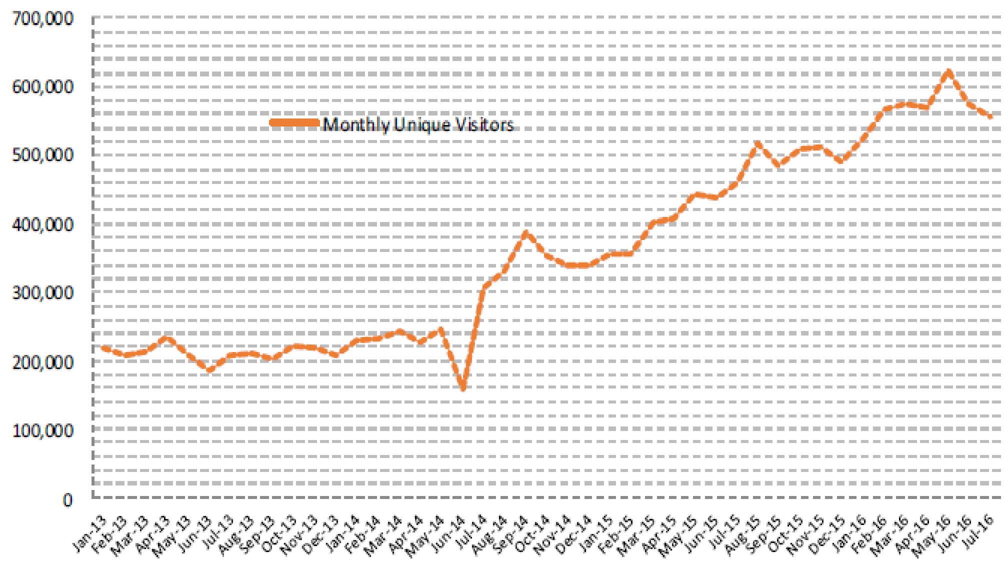

Figure 1-5. Unique Monthly HotCopper Users

This figure illustrates the number of monthly unique visitors of the website HotCopper.
Source: http://www.asx.com.au/asxpdf/20160913/pdf/43b4y2tn62t46v.pdf

Consequently, it is the goal of this dissertation to provide insights on investor sentiment and attention deducted from media and internet channels and its relation to financial market activities. Our results offer a variety of implications for financial practitioners, researchers, and regulators regarding the relevance of information disseminated via social media platforms and news media. We summarize our main contributions and findings of the dissertation in the following structure:

Chapter 2 provides an overview of traditional finance theories and creates a link to the concepts of behavioral finance. To do so, chapter 2 comprehensively describes the basic ideas of the efficient market theory, the noise trading theory on irrational and informed market participants, and psychological concepts in finance, which together help explaining well-documented market anomalies, such as stock price overreaction (e.g., De Bondt and Thaler, 1985) and return momentum (e.g., Jegadeesh and Titman, 1993).

Chapter 3 examines the informativeness of positive and negative investor sentiment expressed on the Australian internet investment platform HotCopper. Several studies are discordant, whether investor sentiment on social media platforms are related to capital market activities (stock returns, trading volume, volatility).

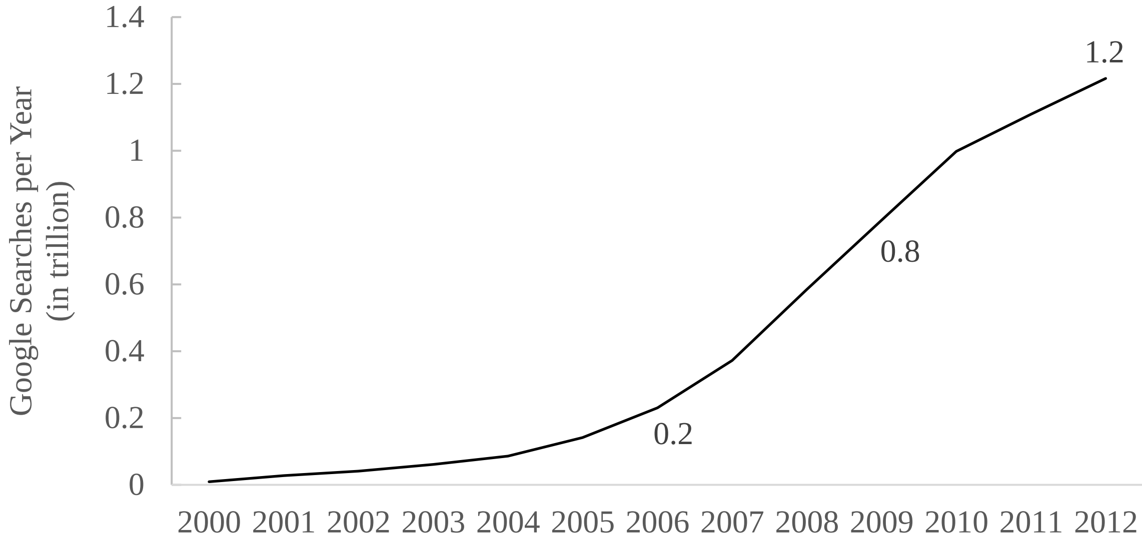

Figure 1-6. Google Searches per Year

This figure depicts the number of Google search inquiries per year. Source: http://www.internetlivestats.com/google-search-statistics/

Since prior literature claim that either online investor sentiment provides no predictive power (Antweiler and Frank, 2004) or that only negative sentiment has the power to explain stock returns in the future (Chen et al., 2014), we demonstrate in our work that positive sentiment leads to overreaction in the short-term. However, short selling activity of informed traders partly mitigates overreaction. Then again, negative investor sentiment seems to convey value-relevant information. Also, we provide evidence that internet message boards possess predictive power before earnings announcements and thus convey fundamental information. We demonstrate that the divergence (convergence) of opinions predict lower (higher) earnings surprises at the date of the earnings announcement.

In chapter 4, we evaluate how investor attention, directly measured as the coverage of traditional news media, internet social media, and internet search queries, affect the well-documented phenomenon of target price run-ups before bid announcements. Controlling for the attention of (un-)sophisticated individual and institutional investors, our results show that dedicated internet investment platforms contribute to identifying run-ups of small enterprise M&A targets. The fundamental characteristics of these firms covered only in internet message boards do not economic meaningfully differ from other small firms (market capitalization, market-to-book, EPS, equity ratio, EBITDA) that receive no (social) media attention. Contrarily, analyst recommendations primarily influence investment decisions from institutional investors who usually cover large stocks. Altogether, the results are consistent with the market expectation hypothesis around M&A announcements.

Chapter 5 completes the third empirical analysis of the dissertation and explores the relationship between news sentiment and cross-sectional returns. More specifically, we examine the media tone of more than 120 million unique US news stories between 2000 and 2017 and its relation to the cross-section of stock returns. Our results provide evidence that an equally-weighted long-short portfolio of stocks sorted by the tone of the news media coverage (in other words news sentiment) earn significant returns of 7.5% per year even after controlling for market, size, book-to-market, momentum, liquidity, profitability, and investment factors. Separating the effect of positive and negative media tones reveals that results are mainly driven by positive media tone which we refer to as a “premium on optimism.”

The last chapter 6 concludes the empirical findings of this dissertation and summarizes the main findings and implications for researchers and practitioners who deal with the topic of investor sentiment and attention in financial markets.

2. From Traditional Finance Theories to Behavioral Finance

The traditional (neoclassical) finance theory rests on the efficiency of capital markets and leaves no room for irrational explanations of stock movements associated with investor sentiment (e.g., Baker and Wurgler, 2006) or noise driven transactions (e.g., De Long et al., 1990). Both concepts are referred to as non-informational and thus irrational. However, critiques on the efficient market perspective increased with the gauging number of observed market anomalies which could not be explained rationally. For example, the overreaction of stock returns (e.g., De Bondt and Thaler, 1985) refers to positive short- but negative long-term autocorrelation of stock returns, resulting in non-rational return reversals after immediate price reactions. It is noticeable at this point that there exists no uniform or comprehensive theory which jointly explains all aspects and market anomalies observed by advocates of behavioral finance.

Hence, we summarize the different behavioral research streams and concepts in Figure 2-1. In this dissertation, explanations for the occurrence of market anomalies are mainly categorized into three concepts or approaches: 1) Behavioral Biases, 2) Investor Sentiment and 3) Noise Trading.[2]

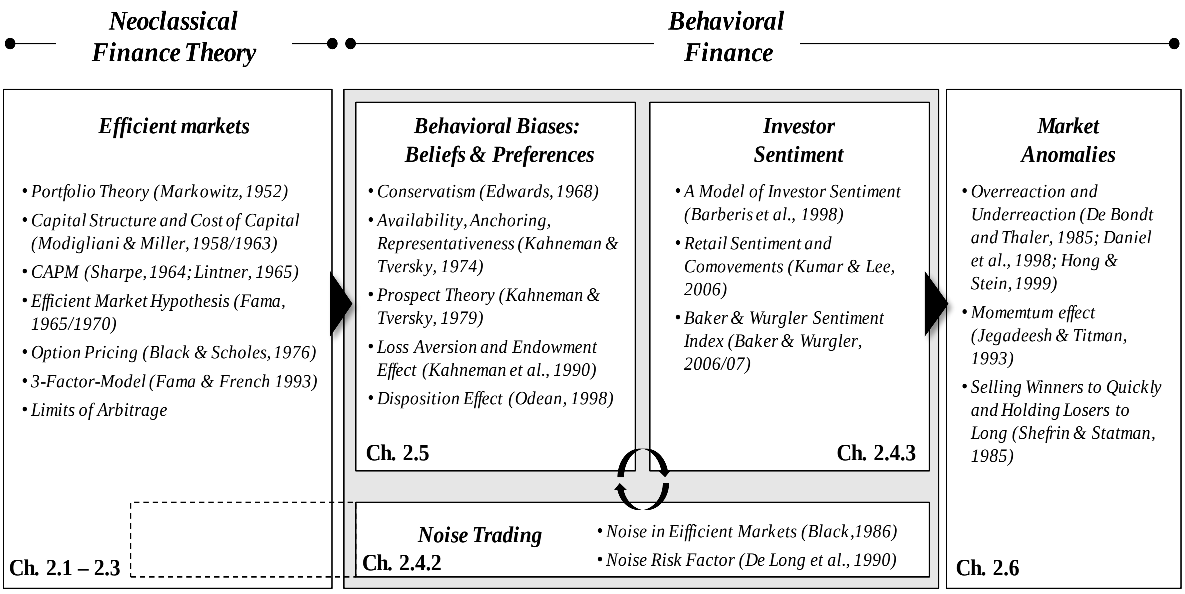

Figure 2-1. Overview of Theoretical Components of Behavioral Finance

[2] The sequence of the three theories should not reflect the order of importance.

Some researchers perceive the behavioral finance view as a short-term phenomenon. Early on, Graham (1965) shared the view that stock markets act like voting systems in the short-term but become a weighting system in the long-term. In this connection, the fundamental value of a firm should hold in the long run and overcome short-term behavioral shocks (Malkiel, 2003). It is consequently a goal of this dissertation to provide further clarity on the short- and long-term role of behavioral finance in capital markets. We provide details and implications on the underlying theories and concepts of traditional and behavioral finance in the following sections to set the theoretical foundation for later analysis and discussions in this dissertation.

2.1. Traditional Finance Theory and Efficient Markets

In the early 1950's, Markowitz (1952) set the foundation for the traditional finance theories with his seminal work on portfolio selection and the interaction of expected returns and associated risks. Modigliani and Miller (1958) further extended the basic foundation by providing theoretical evidence on the independence of a firm's capital structure from its capital costs. The Capital Asset Pricing Model (CAPM), mostly attributed to Sharpe (1964) and Lintner (1965), combines both cornerstones of traditional finance theory and sets a framework on the relationship between risk, capital structure and expected stock returns. Even until now, the widespread research on asset pricing and the associated explanation of cross-sectional return patterns uses the CAPM as the starting point of the so-called risk-factor models.

The random walk theory propagated by Fama (1965) dissents the assumption that informed investors might exploit return patterns. In this context, price changes today are random deviations from previous prices and therefore independent. Furthermore, Fama (1970) formulates, according to the random walk assumption, the efficient market hypotheses in which stock prices reflect all available information so that price changes are unpredictable, similar to the information content (such as news or company announcements) itself (Malkiel, 2003). In an efficient market, security prices accurately provide information on a firm's resource allocation, firms can make decisions on production and investments, and investors can arbitrarily choose between stakes of firm ownership in the form of securities (Fama, 1970). The efficient market hypothesis (EMH) formally tests three different states of market anticipation of information.

First, the weak form posits that information incorporated into security prices only reflect historical prices. In the second form, the semi-strong form tests how capital markets process publicly available information (e.g., earnings announcements). Lastly, the strong-form tests whether investors possess unique access to information that is not available to another group of investors (Fama, 1970).

Also, Fama (1970) states three sufficient conditions that help capital markets to adjust to new information efficiently. An ideally frictionless market does not contain transaction costs for securities, public information is freely available to all market participants, and there is a collective agreement on the implications of new information for future return adjustments. For an efficient market, it is not necessary that these conditions are all fulfilled as long as investors consistently and rationally outperform irrational investors. For example, disagreement amongst investors does not necessarily imply market inefficiency as long as one group of investors can consistently make better investment decisions based on available information. Transaction costs, information that is only limited to a selected group of investors, and disagreement among investors are thus no reasons for market inefficiency but potential sources for market inefficiency (Fama, 1970). All of these three sources of market inefficiency exist in the real world and are hence subject to many empirical tests on market efficiency.

Shleifer (2000) discusses EMH from the perspective of rationality. According to him, in an efficient market, investors act rationally and hence securities should be valued rationally. Irrational investors who trade on non-informational news or events may exist, but their trades are somewhat random and cancel each other, not affecting security prices in the end. Furthermore, irrational investors are met by rational arbitrageurs who trade on superior information and thus eliminate non-informational trades which might affect security prices.

In another groundbreaking work by Black and Scholes (1972), the authors examine whether buying undervalued and selling overvalued price contracts (options) with all available information would result in excess returns according to an efficient market.[3] They find that the options market seems not to be efficient. However, transaction costs hinder traders to exploit the mispricing in options markets.

[3] The option contract gives the right to buy or sell another asset at a pre-determined price within a specific period of time (Black and Scholes, 1972).

Empirical tests on the semi-strong form of market efficiency gained in popularity with so-called event studies, as propagated by Fama et al. (1969). The methodology in particular tests how stock returns react to given news events and if stock prices quickly and correctly adjust over a period. Such events might, for example, include corporate news announcements, takeover announcements or stock splits. The term quickly mainly refers to the fact that stale news read by investors do not move stock prices and traders could not exploit this kind of information. The resulting price changes should on average reflect the fundamental value of the news. There should be no under- or overreaction of stock returns to specific news stories. As a consequence, stock prices should not react to non-informational news (Fama et al., 1969).

In summary, the efficient market hypothesis constitutes the quick and exact reflection of security prices to fundamental information and the non-reaction to noise (Shleifer, 2000). The next section describes in more detail the market mechanism of supply and demand for rational investors in efficient markets. The better understanding of market mechanisms in financial markets provides the foundation for discussions on deviations of rational market behavior and hence creates the link to trading patterns observed in the behavioral finance literature.

2.2. Market Demand and Arbitrage of Rational Investors

The demand and supply of investors determine security prices. In the controversy discussion on traditional and behavioral finance theories, the literature offers numerous distinctions of investor groups. For example, Wurgler and Zhuravskaya (2002) describe two general types of investors, arbitrageurs, and non-arbitrageurs, who differ in two main dimensions. First, arbitrageurs have homogeneous and accurate beliefs about the fundamental value of an asset in the long-term, whereas non-arbitrageurs on average disagree in their beliefs on the fundamental value. Second, arbitrageurs invest in arbitrage portfolios, which require no upfront capital. This portfolio is also commonly known as a zero-net-investment portfolio.

Black (1986) offers another distinction from the informational perspective. He divides investor groups into information and noise traders. The former group bases their trading decisions on fundamental information about an asset, considering the fact that information traders cannot be sure if their information is noisy (non-informational) or correct. Noise traders, on the other hand, trade even though objectively their action might be considered as irrational. They might mistakenly evaluate their noisy information as fundamental, or they just want to trade. Furthermore, Shleifer (2000) refers irrational to unsophisticated investors and denominates rational traders as smart investors.

Altogether, most categorizations of investor groups have in common that one group profits from informational advantages since information asymmetry and disagreement amongst investors exist even in efficient markets. The efficient market theory described in the previous section relies on the informed group of investors who drive asset prices to its fundamental value. To make use of the informational advantage, informed investors can fall back to a financial instrument, called arbitrage. Sharpe and Alexander (1990) define arbitrage as the purchase and sale of a similar security at the same time in different markets. In this connection, the arbitrageur generates profits from different prices for the essentially similar security in different markets. Since efficient markets, in the end, must fulfill the law one price, arbitrage is commonly seen as one of the central elements to enforce the law of one price. Assets of similar risks must, therefore, yield similar expected rates of returns. Hence, over- or undervaluation of assets create profit opportunities and investors would arbitrage on these opportunities (Scholes, 1972). In its basic theory, arbitrage only exists in perfect capital markets, is risk-free and requires no upfront capital. In reality, however, market frictions induce costs related to arbitrage and create risks (Mitchell et al., 2002). We discuss these limitations at a later stage in this dissertation.

In the following, we describe the general market mechanisms of supply and demand for capital markets and the impact of arbitrage on security prices in the market equilibrium. This theoretic foundation on how security prices settle in the equilibrium is necessary to create the link to behavioral finance, where security mispricing deviate from the traditional understanding of finance theory.

Miller (1977) describes the price mechanism of financial markets in a simple two-period model. According to Miller (1977), investors strive to maximize the net present value of their investment and decide to invest if the expected returns of the investment exceed the returns of a risk-free one-year government bond. Given the risk of an investment, investors have heterogeneous beliefs on the expected returns of the investment. The curve *ABC* plotted in Figure 2-2 depicts the cumulative distribution of the number of investors with different beliefs

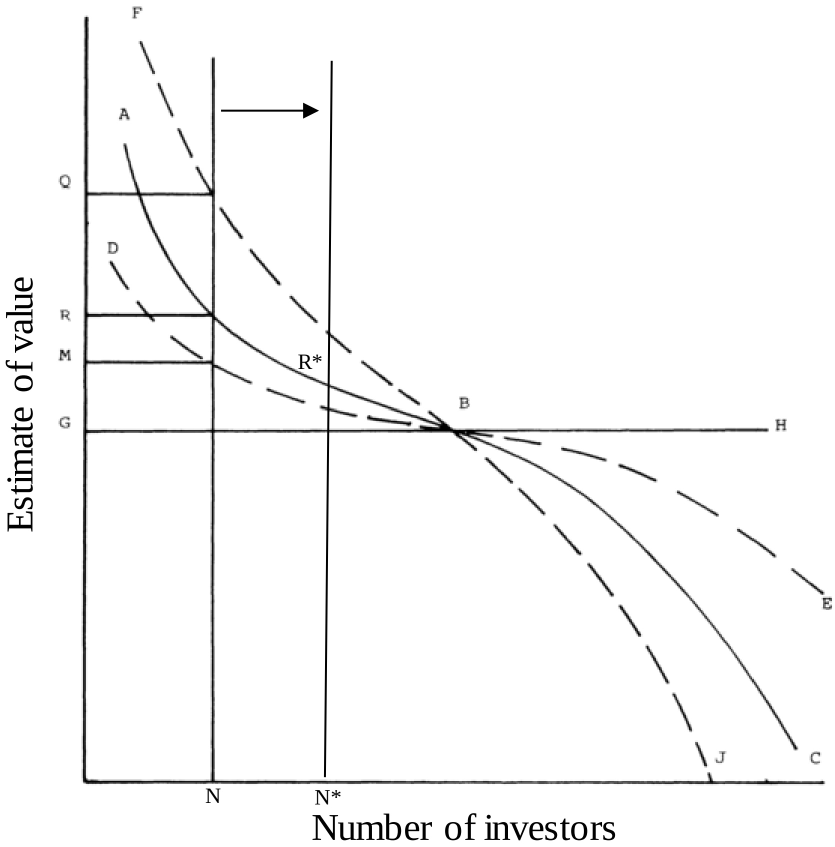

Figure 2-2. Price Distribution of a Security

The figure describes the cumulative distribution of the quantitative number of investors with beliefs above a certain value for the proceeds from their investments. Alternatively, one can interpret the figure as the number of shares investors are willing to hold given a security price. Curve ABC describes the base scenario for the demand curve. A greater (lower) divergence in investor opinion results in replacement of the curve ABC to FBJ (DBE) (Miller, 1977, p. 1152).

in the value of the investment's proceeds. It is evident that *N* investors with the highest evaluations (beliefs) will own the shares at price *R* if investors are only able to buy one single share from the total universe of *N* shares. If the selling price would fall below *R*, then more than *N* investors would be interested in the stock and thus bid up the price back to *R*. In the case that the selling price exceeds *R*, less investors would be interested in the seemingly overvalued security and seek to sell the security, driving prices downwards to *R*. The curve ABC, thus, reflects the demand curve for the security. The vertical line depicts the supply curve and is defined by the number of available shares. The intersection of the demand and supply curve determines the security price *R.* Holding the number of investors constant, a surge (decrease) of divergence in opinions increases (lowers) the market clearing price from *R* to *Q* (*M*). In the extreme case that all investors share the same opinion, the demand curve would result in a straight line *GBH*. In this situation, the market clearing price is equal to the average security price evaluation of all existing investors (Miller, 1977).[4]

What implications does arbitrage have on the market clearing price? Short selling activity constitutes one particular case of arbitrage. Short sales allow investors to sell shares of stocks that they do not own by borrowing a stock from the owner and agreeing to the owner of the stock to compensate any dividends paid by this stock. The investor further commits himself to redeem the borrowed stock upon request and a predetermined date. The effect of short selling on the supply curve is comparable to the effect of money supply by banks. A bank borrows currencies and agrees to pay back the amount upon request. The bank simultaneously lends the currency to a third party. Hence, short sales increase the supply of stocks by the number of open short positions (Miller, 1977). Consequently, short sales move the vertical supply curve horizontally to *N**, resulting in a lower market clearing price of *R** as shown in Figure 2-2. If arbitrageurs, in this case short sellers, possess access to perfect substitute securities and compete against each other for profits, the price of a security should converge to its fundamental value in the long-term (Shleifer, 2000). Short sales only mitigate excessive speculation for price increases and are thus not beneficial for price declines. The market imperfection that investors cannot reinvest the proceeds from short sales (the proceeds are deposited with the lender of the stock or an associated broker as a security) prevent short sale activity to contribute in market efficiency if investors share different opinions on future expected returns on a security (Miller, 1977).

Another implication of arbitrage is that irrational or noise traders diminish from the market in the long-term. If irrational investors purchase overpriced securities and sell undervalued securities on average, then one must expect these group of investors to lose money in the long-term. As a result, competitive selection and arbitrage trades banish irrational traders from capital markets in the long run (Shleifer, 2000). Without any doubt, arbitrage activity exists in real capital markets, but limitations seem to hinder arbitrage and thus the existence of perfectly

---

[4] For more details, please refer to Miller (1977), p. 1152 ff.

efficient markets. One issue is that individual securities do not have perfect substitutes. Accordingly, arbitrageurs who trade on misevaluated securities and hedge the opposite position with non-sufficient substitutes bear the so-called "arbitrage risk" that both return profiles do not cancel each other out. Invalidating the hedging mechanism, an arbitrageur trades more passively in the presence of arbitrage risk (Wurgler and Zhuravskaya, 2002).

It remains open to debate whether the exploitation of price differences can be linked to the existence of mispricing in inefficient markets (market anomalies) or whether resulting profits must be treated as a fair compensation for risk-bearing. Hence, exploiting return patterns is not itself evidence for market efficiency. Earning profits from such an investment strategy, therefore, might just be a result of risk-taking (Shleifer, 2000). One possibility to model a fair relationship between risk and return was established with the well-accepted Capital Asset Pricing Model (CAPM forth on), commonly ascribed to Sharpe (1964) and Lintner (1965). However, measuring and modeling risk is up to now a controversy yielding hundreds of different risk factors. The vast number of risk factors enticed Cochrane (2011) to coin the famous term "factor zoo".

2.3. Asset Pricing and Return Patterns

Pricing mechanisms, arbitrage activity and few potential limitations of arbitrage[5] in capital markets were explained in the previous section. This section of the dissertation, consequently, addresses basics of asset pricing theory and the drivers of well-known return patterns.

2.3.1. Asset Pricing Theory

The asset pricing theory aims to provide explanations for asset prices associated with uncertainty. Consequently, implicitly high rates of return should follow low prices. More generally, the theory explains why certain assets earn more profits than others. In this context, the valuation of assets must account for two dimensions: time effects and the risk of its underlying payments. At this juncture, the price of time is reflected by the pure interest rate, usually represented by the risk-free interest rate. The price of risk translates into the additional expected return per unit risk born (Sharpe, 1964).

[5] Further details on limitations of arbitrage are explained in Section 2.4.1.

There are two competing views on the role of asset pricing: the normative and the positive. From the normative perspective, asset pricing theory determines the true value of an asset so that investors can determine which assets might be mispriced. This practically creates opportunities for the investor to trade and earn adequate risk-adjusted profits. The positive perspective, on the other hand, sees the world as it is so that asset prices are assumed to be accurate. Deviations of asset pricing models must, therefore, be erroneous so that corrections of the models need to be applied (Cochrane, 2001). The normative perspective has prevailed especially for practitioners since this view allows for derivations of asset pricing theory-based investment strategies.

The beginning of asset pricing theory is most commonly credited to Sharpe (1964) and Lintner (1965) summarized with the Sharpe-Lintner-CAPM. As of now, the application of the CAPM is widely-spread, such as for the estimation of a firm's cost of capital and the performance evaluation of a managed portfolio. The CAPM signifies that a firm's risk must be measured relative to an efficient market portfolio, which as a matter of principle must not only consist of financial assets but also consumer durables or real estate (Fama and French, 2004). The CAPM builds upon the portfolio theory credited to Markowitz (1952). In his model of portfolio choice, the investor selects a "mean-variance-efficient" portfolio so that either the portfolio 1) minimizes the variance of the portfolio return at an expected return or 2) maximizes the expected return at a given variance (Markowitz, 1959). The Sharpe-Lintner-CAPM rests on two additional assumptions to the mean-variance-efficient portfolio selection. First, investors in a capital market fully agree on the statistical distribution of asset returns in the future. Secondly, investors have unrestricted opportunities to borrow or lend at a risk-free rate (Fama and French, 2004). The Sharpe-Lintner-CAPM finally describes the linear relationship between an expected return and risks born by the security. Thus, an investor expecting high returns must, therefore, accept higher risks expressed as the volatility of a security. The CAPM can be formally expressed as follows:

$$E(R_i) = R_f + \beta_{i,M}(E(R_M) - R_f) \qquad (1)$$

where $E(R_i)$ is the expected return for security i, R_f is the risk-free interest rate, $E(R_M)$ is the expected market return, and $\beta_{i,M}$ is the market beta of security i. The market beta of a security can be determined with the covariance of the return of security i divided by the variance of the market return as shown in the following equation:

$$\beta_{i.M} = \frac{cov(R_{i,}R_M)}{\sigma_M^2} \tag{2}$$

Hence, the market beta measures the sensitivity of a security's return to the variation in market returns. In other words, the market beta reflects the covariance risk of a security relative to the covariance risk of all securities of the market portfolio, which equals the variance of the market return. Referring back to the two dimensions of asset pricing, the price of time and risk, R_f, therefore, reflects the price of time whereas ($E(R_M)$ - R_f) denotes the price (or premium) per unit of market beta risk (Fama and French, 2004).

The unrealistic assumption of free borrowing and lending in the CAPM prompted Black (1972) to develop an extended CAPM model with limitations on risk-free borrowing and lending. He further assumes unrestricted short selling opportunities for investors of risky assets and that the market portfolio results from the weighted portfolio chosen by each investor. The baseline of his results implies that his extended model only differs to the CAPM regarding the treatment of the risk-free rate. The Black-CAPM requires the risk-free rate to be smaller than the expected market return, which is necessary for a positive premium for the market beta.

One of the major critiques on the CAPM is the empirically appropriate consideration of the market portfolio, which must not be limited to financial assets. Roll (1977) and Roll and Ross (1980) demonstrate the sensitivity of CAPM to financial securities but also all other individual assets. Roll and Ross (1980) formulate the following equation as the central conclusion of the so-called Arbitrage Pricing Theory (APT):

$$E_i = \lambda_0 + \lambda_1\beta_{i,1} + \cdots + \lambda_k\beta_{i,k} \tag{3}$$

where E_i is the expected return of asset i, λ_0 is the return for a riskless asset and $\beta_{i,j}$ (j = 1, …, k) is the coefficient vector for the risk factor premium λ_j. Roll and Ross (1980) argue that fundamental economic variables, such as the Gross National Product, must be a component of systematic risk and as such part of the model (3). However, the authors do not finally resolve the issue which economic variables must finally be included in such an asset pricing model to fully capture the systematic risk required to explain the expected returns of an asset.

An enduring challenge of asset pricing models is to consistently explain returns independent from time regimes. The validity of asset pricing models is prone to price fluctuations in the short run. Some researcher argue that returns indeed can be deducted from past behavior. If historical prices repeat themselves in patterns, they also should occur in the future, contradicting the random walk theory and the associated independence of successive price changes (Fama, 1965). Hence, one must differentiate between historical patterns which might predict future returns as propagated by behavioral finance theory or risk premiums which compensate the investor's risks born with the investment. The following section provides an overview of return patterns observed in the past which potentially contradict the understanding of the efficient market hypothesis.

2.3.2. Return Patterns

2.3.2.1. Patterns Based on Valuation Parameters

The empirical literature provided several indications that initial valuation parameters can predict future returns. Valuation ratios, such as dividend yields or price-earnings-ratios, are found to forecast future returns. In this connection, Fama and French (1988) and Campbell and Shiller (1988) formally conducted statistical tests whether dividend yields (ratio of dividends paid per stock to stock price) predict future returns.

Dividend Yields

Fama and French (1988) study the predictability of dividend yields for a New York Stock Exchange (NYSE) market portfolio for the period between one month and four years. In their study, only less than 5% of return variations could be explained by dividend yields in the period up to three months. However, dividend yields explain more than 25% of the return variances in the long run. In another study, Campbell and Shiller (1988) confirm the former results and find a higher explanatory power of dividend yields for returns in particular for the long run. Dividend yields explain a variation of returns of up to 27% in ten years. Both studies, therefore, jointly find implications for the return predictability of dividend yields and hence a confutation of the efficient market hypothesis.

In a more common asset pricing test design, Malkiel (2003) presented double sorted decile-based portfolios according to initial dividend yields and the following ten-year total returns of firms in the Standard & Poor's 500 stock index. He reports that investors earned higher future returns with purchasing portfolios of high dividend yield stocks and low returns with low dividend yield stocks. In his argumentation, the results do not refute the efficient market hypothesis since high (low) dividend yields are correlated with high (low) interest rates. Hence, initial dividend yields might just reflect the general macroeconomic conditions in which firms operate. He asserts that the predictability of dividend yields diminished since the mid-1980s. In summary, the literature provides no consistent proof of the traditional nor the behavioral finance theory regarding the return predictability of dividend yields.

Price-Earnings-Ratios

Similar implications on return predictability were reported for price-earnings (P/E) ratios of US firms. Basu (1977) documents higher returns for stocks with lower P/E ratios even controlling for transactions costs and taxes. He concludes that security prices are biased and the P/E ratio is a proxy for this associated bias. Campbell and Shiller (1988) find that up to 40 percent of future return variation can be explained with initial P/E ratios. Consentaneously, Malkiel (2003) describes in his paper how stocks with low P/E ratios on average outperformed high P/E ratio stocks in a ten-year horizon. However, in the joint consideration of P/E ratio and dividend yield implications on return predictability become self-contradictory as argued by Malkiel (2003). The next ten-year market returns following years of higher P/E valuations but low dividend yields were remarkably high in the late 1980s averaging nearly 17 percent. However, similar valuations in the early 2000s yield lower future market returns together with higher fluctuations. The interpretation of return predictability must, therefore, be evaluated cautiously when considering the variety of relationships in time (Malkiel, 2003).

2.3.2.2. Patterns Based on Firm and Valuation Parameters

Eventually, researchers document a flat relationship between returns and the systematic risk of market betas (Black and Scholes, 1972; Fama and MacBeth, 1973). This raises the question which factors really capture risk. Consequently, several prominent studies evaluated the validity of the capital asset pricing model and established two well-accepted risk factors known as the size effect and the valuation effect of book-to-market ratios. This section details the findings for these risk factors and summarizes the implications for the traditional and behavioral finance theory.

Firm Size

One of the most prominent effects researchers documented is the outperformance of small-capitalization compared to large-capitalization stocks (Malkiel, 2003). As one of the first researcher, Banz (1981) investigated the size effect between the period of 1936 and 1975. In his study, stocks of small firms earned on average higher risk-adjusted returns than stocks of large firms. Hence, he concludes a misspecification of the CAPM and yet cannot fully explain whether size itself reflects a risk factor or just serves as a proxy for another unknown risk factor. One possible explanation for the size effect is the lack of information available for smaller stocks. Thus, investors claim a compensation for an estimation risk on return distributions since only little information is available for smaller stocks (Klein and Bawa, 1977). In another study, Keim (1983) provides further empirical evidence on the validity of the size effect. He reports consistent negative correlations between abnormal returns and size in the period between 1963 and 1979. However, nearly fifty percent of the premium can be attributed to a seasonal effect in January. Following the most prominent asset pricing test setting, Fama and French (1993) sorted stocks according to market capitalization and returns into deciles with a broader dataset between 1963 and 1990. Decile-portfolios constructed for smaller stocks generate on average higher monthly returns compared to larger stock portfolios. In this connection, the critical question is the extent to which higher returns of small stocks are related to return patterns that create opportunities for investors to earn excess risk-adjusted returns. If the capital asset pricing model accurately captures the total systematic risk of a stocks market beta, then the additional size effect might be referred to a market anomaly embedded in an inefficient capital market (Malkiel, 2003). Furthermore, double-sorted portfolios by market betas and market capitalization imply resilient size effects but flat market betas. As a consequence, Fama and French (1993) argue that size may better capture systematic risk than market betas, and their findings are consistent with the efficient market hypothesis. Possible concerns on the validity of the size effect are, for example, addressed by the survivorship bias, which neglects the performance of smaller firms that went bankrupt. Researchers would, thus, only measure the returns of firms that survive (Malkiel, 2003). Altogether, it remains discussable in the literature whether size may serve as a proxy for risk or explains returns in excess to risk-adjusted expectations consonant to the behavioral finance theory.

Value Stocks and Growth Stocks

Another popular pattern discussed in the literature is the above average performance of stocks with higher book-to-market ratios also coined with the term “value” or “value stocks”. The associated overpayment of “growth stocks” (stocks with lower book-to-market ratios), therefore, results in lower future returns by the view of behaviorists (Malkiel, 2003). The positive relationship between average returns of US stocks and the ratio of book-to-market were documented by Stattman (1980) and Rosenberg et al. (1985). In addition to the market risk and size effect, Fama and French (1993) created the well-established Fama-French three-factor model in which the market premium, size and book-to-market ratio jointly measure the extent of systematic risk. Seeking further evidence on the validity of the three-factor model, Fama and French (1998) extended their empirical tests from the US to a global study and find confirmation on the universal explanatory power of their model.

As a whole, many findings in the literature question the validity of the standard capital asset pricing model in its purest form. The general implications, however, do not allow to finally draw the conclusion on market inefficiency but intensify discussions on the role of risk and excessive risk-adjusted returns within historical return patterns. It is indisputable that the mispricing of securities does occur in reality even if only in the short-term. The next sections provide theoretical explanations (market frictions and psychological approaches) for the occurrence of mispricing independent from the question if capital markets are efficient in the long run.

2.4. The Transition to Behavioral Finance

As Shleifer (2000) pointed out, “At the most general level, behavioral finance is the study of human fallibility in competitive markets” (Shleifer, 2000, p. 23). This theory not only acknowledges the existence of irrational, confused, or biased investors in capital markets. Beyond that, the theory describes the ubiquitous interaction of those and rational investors in a complex financial market setting. Behavioral finance, therefore, investigates the outcome of such interactions on asset prices or other financial market-related performance dimensions, such as return volatility (Shleifer, 2000).

Asset price deviation from the fundamental value, in fact, exist as a result of market imperfections. Two commonly named foundations of behavioral finance theory, and thus the explanations for deviations of asset prices from fundamentals, are 1) limited arbitrage and 2) the concept of investor sentiment (Shleifer, 2000). Non-perfect substitutes, direct and indirect costs of arbitrage, time and capital restrictions of rational investors, or portfolio specialization are widely discussed reasons for limited arbitrage activity. De Long et al. (1990) and Shleifer and Vishny (1997) moreover theoretically attribute limitations on arbitrage to the basic concept of the noise trading theory. They argue within a theoretical framework that noise trader risks prevent rational arbitrageurs from trading against mispriced positions and thus converging asset prices to fundamental values.

Closely linked to the former theories is the concept of investor sentiment or in other words the concept on how investors form their beliefs in future asset prices (Barberis et al., 1998; Shleifer, 2000). More specifically, Baker and Wurgler (2006) define investor sentiment as either the propensity to speculate or investor optimism/pessimism about stocks in the general context. The behavioral finance theory relies on both dimensions to explain the mispricing of assets. In financial markets with unlimited arbitrage, informed investors could quickly counteract noise traders and cause otherwise unjustified prices to converge to their fundamental values. With the absence of investor sentiment, arbitrageurs would ascertain the efficiency of financial markets. In combination, both theories of limited arbitrage and investor sentiment enable researchers to forecast future stock returns (Shleifer, 2000). We further describe the above-mentioned foundations of behavioral finance theory in more detail in the next sub-sections.

2.4.1. Limitations on Arbitrage in Imperfect Markets

According to the efficient market hypothesis arbitrage against mispricing does not require capital upfront and entails no risk (Shleifer, 2000). In a perfectly competitive market, each arbitrageur might take an infinitesimal small position against the mispricing and pushes prices towards the fundamental value. Capital constraints, therefore, do not exist for such small arbitrage positions and arbitrageurs would be risk-neutral (Shleifer and Vishny, 1997).

The real world, however, looks different. Capital constraints cause arbitrageurs to act risk-averse, and arbitrage trades are as a matter of fact risky. In a seminal work on limitations of arbitrage, Shleifer and Vishny (1997) formalized a theoretical model on limited arbitrage consisting of three time periods and three market participants: irrational noise traders, rational arbitrageurs, and rational investors who allocate their funds to arbitrageurs. In this model, arbitrageurs are specialized in trading only whereas investors do not trade on their own but delegate portfolio management decisions to arbitrageurs. As of now, noise traders are associated with non-informational transactions.[6] A misperception of noise traders causes prices to deviate from the fundamental value of an asset. Arbitrageurs, on the other hand, are aware of the true fundamental value of the asset. The valuation of arbitrageurs and noise traders converge in time period 3, assuming that there is no fundamental risk in the long run. The model asserts that arbitrageurs aim to maximize their profits (which in this model equals the total funds in time period 3) according to the following statement:

$$EW = (1-q)\left\{\alpha\left(\frac{D_1 * V}{p_1} + F_1 - D_1\right) + (1-\alpha)F_1\right\} + q\left(\frac{V}{p_2}\right) * \left\{\alpha\left(\frac{D_1 * p_2}{p_1} + F_1 - D_1\right) + (1-\alpha)F_1\right\} \quad with\ \alpha \geq 1 \qquad (4)$$

where *EW* denotes the arbitrageur's third period funds, *q* is the probability that the noise trader's misperception intensifies in period 2, D_1 is the investment value of arbitrageurs in an asset in t = 1, *V* is the fundamental value of the asset, p_t is the price of the asset at time *t*, F_t is the limited cumulative amount under management for the arbitrageurs, and α describes the sensitivity of assets under management against historical performances. If arbitrageurs lose

[6] Details on noise trader theory follow in the next section.

money on their positions, investors will not provide more funds and α takes the value of 1. α is greater than 1 if investors withdraw money as a consequence of poor historical performances. An important assumption is, hence, the allocation of funds based on historical performances of arbitrageurs.

The main implication of the model is that performance-based arbitrage, which is the sensitivity of investors to refuse/provide more capital or to even withdraw capital from the fund under management as a result of poor historical performances, causes inefficiencies in particular in extreme situations. This is the case when prices significantly deviate from its fundamentals and arbitrageurs are fully invested. Facing the risk of high short-term losses, arbitrageurs withdraw from the market when arbitrage would be most profitable. Arbitrageurs, thus, face limitations in the moment of their best opportunities.[7]

The former model on limited arbitrage includes dimensions such as the specialization of arbitrageurs and capital constraints as potential reasons for limited arbitrage. Furthermore, the literature offers several different reasons for limitations on arbitrage which potentially cause market inefficiencies. One explanation refers to the lack of perfect substitutes in the real world. A riskless hedge of assets requires perfect securities, implying a certainty that both, the relative prices of the underlying asset and the hedge, converge. Yet, imperfect markets are characterized by statistical likelihoods as opposed to certainty (Shleifer, 2000). Arbitrage costs are other prominent obstacles that deter arbitrage opportunities. Direct trading costs, for example, incur for short selling activity associated with arbitrage. To borrow a stock, the short seller must pay a fee to the stock lender or an intermediary. In addition, indirect costs incur if short positions are closed due to the recall of the stock by the lender. Furthermore, stock borrowers must post collaterals if the price of the shorted stock increases and payouts are due if brokers call out the margin-call. Another indirect cost can be attributed to costs related to finding a stock lender in non-centralized shorting markets. Hence, arbitrage is costly and risky in an imperfect financial market (Jones and Lamont, 2002). Other reasons for limited arbitrage are found in the special-

[7] Please refer to Shleifer and Vishny, 1997, p. 38 ff, for more details on the theoretical model. The authors describe in four distinctive propositions the implications of specific extreme situations on the ultimate arbitrageur's goal to maximize the funds under management in period 3.

ization of arbitrageurs and time constraints. An arbitrageur's portfolio might lack diversification as a result of high specialization, which causes the arbitrageur to bear idiosyncratic risks. If prices further deviate from their fundamental value, then specialized arbitrageurs are not able to diversify that risk. Additionally, if prices deviate temporarily and arbitrageurs do not own enough capital to engage in further arbitrage trades, they are forced to close the position and realize losses (De Long et al., 1990; Shleifer and Summers, 1990 and Shleifer and Vishny, 1997). Facing the losses reduces the willingness of the arbitrageur to invest additional amounts (Mitchell et al., 2002). Furthermore, institutional and cultural inhibitions could also potentially deter arbitrage since underlying guidelines might prohibit short selling or partly limit the extent of arbitrage activity (Jones and Lamont, 2002).

The literature documents several implications of limited arbitrage for financial markets. For example, one might assume that high volatility in financial markets attract arbitrageurs because mispricing would frequently exist in such an environment. However, arbitrageurs rather avoid short positions in highly volatile financial markets in particular in the presence of high fundamental risks. If risk-adjusted excess returns (often called alpha) do not increase proportionally to the volatility, arbitrage becomes less attractive. In addition, higher volatility increases the likelihood of losses, which, given time and capital constraints, deters arbitrage for risk-averse arbitrageurs (Shleifer and Vishny, 1997). Moreover, the literature argues that stocks with high short selling costs often tend to be smaller growth firms with higher market-to-book valuations (Jones and Lamont, 2002). This goes hand in hand with the view on overpriced stocks which tend to be owned by only a few optimistic investors. Arbitrageurs, or more specifically short sellers, avoid buying overvalued stocks (Miller, 1977). Based on their theoretical model, Diamond and Verrecchia (1987) conclude that the absence of trade is a bad signal for financial markets. In their argumentation, abnormal low trading activity may be a result of situations in which informed traders with bad news face short selling constraints and thus cannot trade based on their information. Engelberg et al. (2012), on the other hand, find evidence that short selling activity and advantages from such trades substantially result from an arbitrageur's ability to analyze publicly available information. However, arbitrageurs rarely anticipate informative news events. The authors do not finally answer the question whether this weak anticipation results from limited arbitrage or other factors.

In summary, systematic and idiosyncratic risk matter both for professional arbitrageurs. Arbitrageurs or informed investors are in reality often reflected by only a few highly specialized investors, refuting the view on a perfectly competitive market with homogenous views on asset prices. In imperfect markets, hedging is not costless and risk-free. The classical view on the positive correlation between risk and returns thus changes with limited arbitrage. The antagonists of arbitrageurs in financial markets are the so-called noise traders. They are mostly referred to the group of uninformed investors. The underlying noise trading theories and concepts are detailed in the next section.

2.4.2. Noise Trading in Efficient and Inefficient Markets

Noise trading induces a particular risk that, in addition to transaction costs or imperfect substitutes, deters arbitrage. The literature offers several definitions for noise or noise traders. In one of the fundamental work on noise, Black and Scholes (1972) interpret noise traders as investors who falsely trade on information which they believe is correct. Shleifer (2000), on the other hand, states that noise traders conduct transactions based on their erroneous beliefs on future distributions of returns on risky assets. Yet, most definitions have in common that noise traders are considered irrational and objectively uninformed.

In one of the first theoretical studies on formal models of informed and uninformed trading, Grossman (1976) developed a pricing model which first included noise. In this model noise ultimately prevents informed traders to observe the true fundamental value of an asset. He concludes that a market equilibrium, where prices reflect all aggregated and available information, might break down in the presence of noise. In another study, Black (1986) offers economic reasonings for noise traders and their role in financial markets. In his argumentation, noise does exist in different dimensions. Noise provides substantial liquidity to a financial market, but in interaction with arbitrageurs, prices would ultimately be pushed back to their fundamental value. Noise trading is in this perspective a complementary element in efficient markets and a foundation for liquidity in financial markets.

Another view on noise is the risk perspective of noise. Mispricing resulting from noise trading might occur in the short-term but diminishes in the long run as long as informed traders exploit the arbitrage opportunities. However, there is a risk for arbitrageurs that a noise trader's misperception persists or even increases before prices return to the mean. For example, if the noise trader's optimism drives up prices, arbitrageurs should (short-)sell this asset, assuming that prices reverse in the future. Yet, there is a risk that noise traders become even more optimistic and push prices further away from its fundamentals. Fearing additional losses or receiving pressure from their investors, arbitrageurs liquidate their positions and realize losses. The fear of loss, hence, limits the original amount invested by risk-averse arbitrageurs. This risk of additional losses is subsumed under the noise trading risk (De Long et al., 1990).

In a constitutional work on noise trader risk, De Long et al. (1990) developed a theoretical model on asset prices as a function of exogenous variables, which in its purest form is stated as follows:

$$p_t = 1 + \frac{\mu(\rho_t - \rho^*)}{1+r} + \frac{\mu\rho^*}{r} - \frac{(2\gamma)\mu^2\sigma_p^2}{r(1+r)^2} \tag{5}$$

where p_t is the asset price at time t, μ is the share of noise traders present in the model, r is the dividend of the asset, ρ_t is a random variable reflecting the misperceived expected price of the risky asset, ρ^* is the measure of the average bullishness of the noise traders, γ is a coefficient describing the absolute risk-aversion of investors, and σ_p^2 denotes the variance of the noise trader's misperception of expected returns per unit of the risky asset.[8] The model in its simplest form provides three main implications of noise trading for financial markets. First, a shift of noise trader's opinions induces fluctuations of prices and hence volatility. Second, the average misperception of noise traders is unequal to zero. Deviations of beliefs by noise traders, therefore, cause a mispricing of assets. Lastly and the probably most important implication, noise trading creates risks. The uncertainty over the noise trader's belief in the next period makes an otherwise riskless asset risky, drives prices down and increases future expected returns. In this connection, a total risk-aversion of zero would imply that investors do not sell their assets in case of overpricing. As a result, prices remain high and expected future returns are low (De Long et al., 1990).

[8] Please refer to De Long et al. (1990), p. 707 ff, for details on the model.

Based on this model, noise trading creates risks. Hence, if the noise trader's belief follows a random walk, then prices revert to the mean in the long run. An instationary process of noise trader's beliefs, however, results in a persistent deviation of the asset price with continuous risk premiums for noise trading.

Noise trading is commonly associated with transactions based on non-information. The behavioral perspective, therefore, argues that noise traders base their decisions on false or inaccurate information (Black, 1986). Bloomfield et al. (2009) distinguish in their study between two types of noise traders. The first group are "liquidity traders" who's trades are triggered by random liquidity shocks (for example, a fund's investor requires liquidity for some unknown reasons and recalls funds without no economic reason). The second group consists of "uninformed traders" who trade despite having any advantageous information. We do not differentiate between both groups and refer all non-informative trades to noise traders forth on.

In summary, noise trading and its associated risk have several implications for financial markets. With the presence of noise traders, prices of assets are excessively volatile and not correlated to the variance of its fundamentals. If asset prices react to noise temporarily, then asset prices should revert to the mean in the long run. Yet, a persistence in noise trader risk might force capital constrained investors to withdraw from the market. The expected mean reversion of asset prices additionally changes the logic to traditional investment strategies that propagate the buy-and-hold-strategy. The noise trading theory creates room for the so-called contrarian investment strategy, where the timing of investment decisions is essential. With this strategy, arbitrageurs invest in times when noise traders are bearish and reduce their exposure when noise traders are bullish (De Long et al., 1990). One can infer that bullishness is one essential element in the concept of noise trading and in behavioral finance. Bullishness is highly connected to the concept of "investor sentiment". Consequently, we discuss the theory and implications of investor sentiment for financial markets in the next section.

2.4.3. Investor Sentiment

The previous section described a formalized model on limited arbitrage that gives implications on a noise trader's misperception of asset prices for financial markets. Yet, the model missed to further explain how the misperception of asset prices, or investors beliefs, evolve over time. In order to fill this putative gap, Barberis et al. (1998) introduced a theoretical model of investor sentiment. The model presented in this section connects empirical evidence on stock return predictability and seminal psychological concepts to jointly explain the formation of investor beliefs and its return predictability. The observed patterns considered in the model are the phenomena of under- and overreactions in financial markets. Underreactions describe the slow incorporation of particular news into asset prices. Hence, the average returns of firms gradually adjust to fundamentals in subsequent periods. Overreaction, on the other hand, results from a string of good (bad) news after which investors become too optimistic (pessimistic) and overvalue (undervalue) a stock. Furthermore, the model integrates the psychological concepts of conservatism and the representativeness heuristic to form a comprehensive behavioral-based financial model.

Before explicating the model, we first need to clarify the ambiguous understanding of investor sentiment. There are several different definitions and understandings of investor sentiment in the literature. In the end, most definitions refer to the preferences and beliefs of investors in confirmation of psychological evidence rather than the rational normative model. The belief in heuristics instead of rational information is thus defined as investor sentiment (Shleifer, 2000). Baker and Wurgler (2006) offer two different definitions of investor sentiment. In their first possible definition, investor sentiment is an investor's tendency to speculate. A second more general definition describes investor sentiment as the optimism or pessimism about an investment. The terms optimism or pessimism are commonly related to bullish and bearish sentiment, respectively (Brown and Cliff, 2004). The understanding of bullishness, however, is ambiguous and often spuriously used in the literature. Brown and Cliff (2004), therefore, assert that bullishness is a measurement of the discount from an intrinsic value of an asset. A bullish (bearish) investor, thus, assumes an undervaluation (overvaluation) based on current prices. In this definition, a bullish (irrational) investors would not only expect absolute positive returns, but also a relative outperformance of returns compared to rational investors. The measurement of the discount on the intrinsic value is hardly feasible so that the term bullish (bearish) is commonly associated with an expected price increase (decline) (Brown and Cliff, 2004). This dissertation follows the later understanding of bullishness as the directional interpretation of investor sentiment.

To formally describe the formation of investor sentiment, Barberis et al. (1998) developed a parsimonious model to explain asset prices from the behavioral perspective as mentioned before. The model consists of one investor and one asset. This investor reflects the consensus belief of all investors even if they in total have heterogeneous beliefs. The investor, thus, affects prices and returns. The future earnings of the asset are independent of previous returns and follow a random walk. However, the investor does not anticipate the independence of returns in time. Instead, he believes in two different states in which a firm's earning can move. The first state assumes a mean-reversion of earnings. The second state considers an instationary process, where earnings follow an autocorrelated trend. A constant transition probability exists between both states so that the investor does not adapt these probabilities based on past experiences. Earnings are more likely to stay in one state than switching to the other. The investor regularly and rationally updates his beliefs based on the observed earnings. For example, if negative earnings follow previous negative earnings, then investors rather belief that they find themselves in a (negative) trending state. However, if positive follow negative earnings, investors raise the likelihood of being in a reverting state according to state 1. In this model, the price of an asset is simply the discounted value of all future expected earnings:

$$P_t = E_t\left\{\frac{N_{t+1}}{1+\delta} + \frac{N_{t+2}}{(1+\delta)^2} + \cdots\right\} \tag{6}$$

where P_t is the price of the asset, N_t reflects the earnings, and δ is a discount factor. The authors show that in a state switching model, as believed by the model's investor, the pricing function can be expressed as follows:

$$P_t = \frac{N_t}{\delta} + y_t(p_1 - p_2 q_t) \tag{7}$$

where y_t is a shock to earnings at time t, p_1 and p_2 are constants which depend on model-specific transition parameters, and q_t is the probability that the shock at time t was induced by state 1.[9] State 1 assumes a mean-reversion to fundamental values of asset prices. State 2, on the contrary, assumes a trend in earnings. According to the random walk hypothesis, however, the

[9] Please refer to Barberis et al. (1998) for details on the model specification and proof of the model.

expected return should follow $E(N_{t+j}) = N_t$. In equation (7), $\frac{N_t}{\delta}$ denotes the fundamental value of the asset whereas $y_t(p_1 - p_2 q_t)$ equals the deviation from the intrinsic value as assumed by the investor and the true fundamental value. The implications of this model can be summarized as follows. The investor's belief in state 1 (mean-reversion), which implies a higher value of q_t, relates to the psychological concept of conservatism, given p_1 is small in relation to p_2. The investor underrates the importance of the news event even though the information is of statistical importance. This ultimately leads to an underreaction of the investor to news, such as corporate earnings announcements. On the contrary, if the investor tends towards state 2 (trend), which implies a lower value of q_t, then his decision is closely connected to the psychological concept of the representativeness heuristic, given p_1 is large in relation to p_2. The second scenario is equivalent to an overreaction of investors to news. An increase in the asset price would be followed by another price increase and the investor would consequently overrate past information (Barberis et al., 1998).[10]

The presented model provided theoretical explanations for the formation of beliefs and the interaction of psychological and financial concepts. Yet, the literature has shown that different types of stocks are more sensitive to investor sentiment than others. Baker and Wurgler (2007) investigate the sensitivity of stock types against broader waves of investor sentiment and find that in particular "stocks of lower market capitalization, younger, unprofitable, high-volatility, non-dividend paying, growth companies or stocks of firms in financial distress" (p. 130) to be more sensitive to shocks in investor sentiment. The findings of their study can be summarized in Figure 2-3. Stocks that are harder to value and to arbitrage on are ordered towards the right on the x-axis. These are particular smaller, younger, more volatile, unprofitable growth firms. Easier to value and to arbitrage stocks are ordered towards the left. An example are regulated utility stocks, which are fairly transparent to financial markets with long-lasting earnings histories (Baker and Wurgler, 2007). The y-axis describes the valuation of the stock, with $P*$ equaling the fundamental value on a stock. According to this figure, high valuations are related to high levels of investor sentiment especially for stocks which are difficult to value

[10] Please refer to Barberis et al. (1998) for details on models.

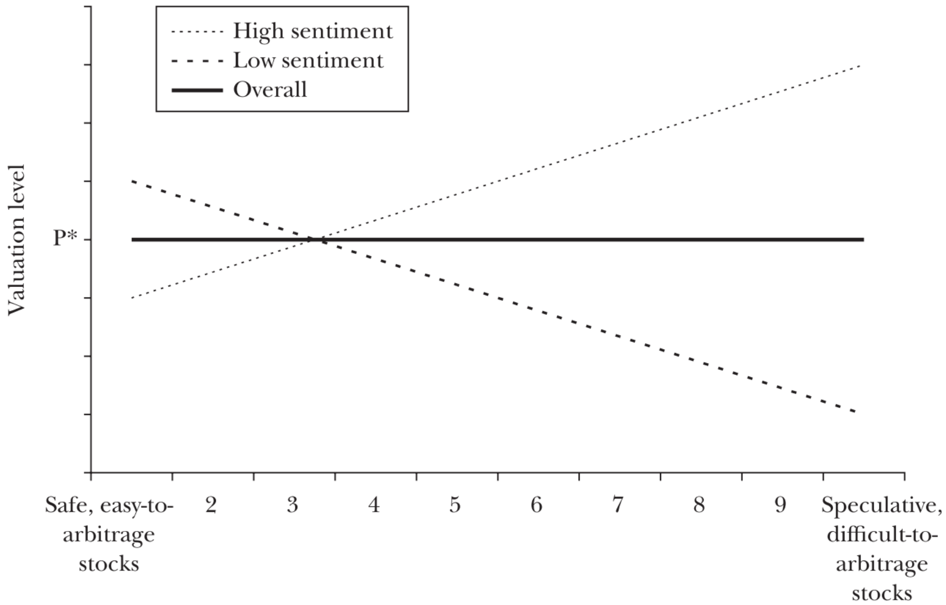

Figure 2-3. Theoretical Effects of Investor Sentiment on Stocks

This figure describes the interaction between investor sentiment, the stock valuation, and the difficulty on stock valuation and arbitrage. P* denotes the fundamental value of the stock (Baker and Wurgler, 2007, p. 133)

and arbitrage. The market is assumed to price stocks correctly in the absence of noise or in this case investor sentiment (Baker and Wurgler, 2007).

As stressed out by Baker and Wurgler (2007), "… the question is no longer, as it was a few decades ago, whether investment sentiment affects stock prices, but rather how to measure investor sentiment and quantify its effects" (p. 130). A variety of literature dealt with the question of how to accurately capture investor sentiment from direct or indirect measures. Table 2-1 summarizes the most commonly discussed proxies or measures for investor sentiment in the behavioral finance literature. In narrowing down the variety of sentiment measures for empirical tests, this dissertation mainly focuses on investor sentiment extracted from (social) media and internet investment platforms for several reasons. A survey on investor sentiment is most likely to be the most direct measure on investor sentiment. A major weakness identified by economists is the potential gap between how panelists respond and how they actually act (Baker and Wurgler, 2007). Moreover, survey-based investor sentiment proxies are often only measured in monthly, quarterly or yearly frequency. Explanatory power on micro-trading-level is, hence, highly limited (Da et al., 2015). Market-based measures, such as the IPO volume or

Table 2-1. Proxies for Investor Sentiment

This table summarizes common proxies for investor sentiment discussed in the behavioral finance literature. This extended table includes a summary provided by Baker and Wurgler (2007).

Proxy / Measure	Hypothesis / Implications	Related studies
Investor surveys	Asking investors about their opinion on trends in financial markets, e.g., - direct (randomly-selected households) - indirect (Consumer Confidence Index)	Brown and Cliff (2005); Qiu and Welch (2006); Lemmon and Portniaguina (2006)
Investor mood	Extracting sentiment from exogenous changes in human emotions, e.g., - market returns in different seasons (fall, winter) - losses in major sports events	Kamstra et al. (2003); Edmans et al. (2007)
Retail investor trades	Development of sentiment index based on micro-level trading data from retail investor transactions	Kumar and Lee (2006)
Mutual fund flows	Sentiment implications from fund flows, e.g., fund pulled out from government funds and reinvest in growth stock funds	Frazzini and Lamont (2008)
Trading volume	Irrational investors add liquidity to the market in particular when they are optimistic	Scheinkman and Xiong (2003); Baker and Stein (2004)
Dividend premium	Firms pay dividends when dividends are at premiums (premium is the difference between average market-to-book ratios of dividend payers and nonpayers)	Baker and Wurgler (2004)
Closed-end fund discount	Low sentiment is related with higher closed-end fund discounts (which is the difference between the net asset value of a closed-end fund and its market price)	Zweig (1973); Lee et al. (1991); Neal and Wheatley (1998)
Option implied volatility	Expected volatility (e.g., calculated by back-solving of Black-Scholes-Formula for option pricing) is inversely related with investor sentiment. The Market Volatility Index ("VIX") is often referred to capture market pessimism	Whaley (2000)
IPO volume and returns	First-day returns of IPOs and the general demand for IPOs assumed to be sensitive to investor sentiment	Baker and Wurgler (2006)
Equity issues over total new issues	Share of total equity offerings over total equity and debt issued by all firms is positively related to sentiment	Baker and Wurgler (2000)

Textual analysis / Self-disclosed sentiment	Sentiment extraction from (social) media and internet investment platforms to directly capture investor sentiment	Antweiler and Frank (2004); Tetlock (2007); Das and Chen (2007); Chen et al. (2014)
Search-based measure	Measurement of search frequency and sentiment related words via internet search engines	Da et al. (2015)

IPO first-day returns, on the other hand, bear the disadvantage to explain output variables (e.g., returns) with other output variables (e.g., IPO first-day returns) instead of exogenous input variables (Qiu and Welch, 2006; Da et al., 2015). The application of investor sentiment directly extracted from (social) media outlets alleviates the above concerns.

This section comprehensively described the theoretical concepts on the formation of beliefs, its implications for financial markets, and further provided an overview of empirical measures of investor sentiment. So far investor sentiment was mostly discussed as a given outcome from psychological driven actions by irrational investors. The next section further explains the psychological concepts applied in behavioral finance to complete the picture on the behavior of irrational investors in imperfect markets.

2.5. Psychological Background on Investor Beliefs and Preferences

The theories on limited arbitrage and investor sentiment relied primarily on the assumption of irrationality, which ultimately causes deviations from fundamental values. This dissertation discussed before how behavioral financial related theories consider the formation of beliefs and that investors act on noise as if it was real information. The following section provides cognitive psychological backgrounds on how investors actually form their beliefs and how preferences influence the decision-making process of investments.

2.5.1. Beliefs

The formation of investor beliefs is the foundation of the concept of investor sentiment. In the following. we summarize selected and the from our view most relevant psychological concepts for this dissertation in the financial context.

Overconfidence

One of the most robust implications on the psychology of people's judgement is that people are overconfident (De Bondt and Thaler, 1995). This phenomenon exists in two types. First, people set too narrow confidence intervals in their judgement of an estimate. For example, an assumed 98% confidence interval typically only include the true estimate in about 60% of the cases (Alpert and Raiffa, 1982). Second, people set poor probabilities on the occurrence of an event. Events that people are certain about to occur happen only in 80% of all cases, and perceived impossible events occur about 20% of the time (Slovic et al., 1977). These observations apply to many fields of job profiles, such as physicians, engineers, attorneys or investment bankers (Daniel et al., 1998). In particular, experts tend to be more overconfident than comparably inexperienced persons (Griffin and Tversky, 1992). In the finance perspective, overconfident investors believe in their superior ability to value a security compared to their true ability. Hence, they underestimate the resulting forecast error. The concept of overconfidence is adapted by seminal theories of behavioral finance, such as the investor sentiment model by Barberis et al. (1998) and Daniel et al. (1998) or excess volatility as propagated by Shiller (1981) as discussed in previous sections.

Representativeness

According to Kahneman and Tversky (1972), representativeness is a heuristic where "… the subjective probability of an event, or a sample, is determined by the degree to which it: (i) is similar in essential characteristics to its parent population: and (ii) reflects the salient features of the process by which it is generated" (p. 430). The first statement translates into the implication that selected events are seen as typical or representative, ignoring the laws of probability in the overall process. For example, firms with a history of consistent growth are perceived as future growth stocks even though there is only a small likelihood that companies keep growing (Barberis et al., 1998). This bias is also referred to as the "base rate neglect" (Barberis and Thaler, 2003). Hence, people underestimate the true probability that an event or sample belongs to a defined population. The second observation reduces to another bias, called the "sample size neglect." When estimating the likelihood of a process, experimental subjects miss accounting for the sample size. In a financial context, a financial analyst with four positive historical recommendations is in this in view perceived as talented despite the limited and unrepresentative number of overall observations (Barberis and Thaler, 2003). In summary, people see patterns in a random walk process. The heuristic of representativeness is, thus, embedded within the concept or market anomaly of overreaction which we discuss in detail in section 2.6.1.

Conservatism

The psychological concept of conservatism goes back to Edwards (1968). He defines conservatism as the slow update of general models in response to new evidence. Hence, individuals only slowly change their beliefs in the presence of new events or information. Edwards (1968) tests a subject's reaction to new evidence against a well-defined piece of evidence where the true normative value is known. In his findings, test subjects overemphasize the base rates (compared to the underestimation of base rates as seen by representativeness) in relation to the sample size (Barberis and Thaler, 2003). In this connection, an investor who is subject to conservatism underestimates the full information of earnings announcements (overemphasis on prior base rates) because he only partly believes in the new information and still holds on his prior estimate of the earning. As a result, valuations only slowly adjust to the full information content of the earnings announcement. Another interpretation is that conservative individuals are overconfident about historical information (Barberis et al., 1998). This phenomenon is connected to the market anomaly of underreaction discussed in section 2.6.1.

Anchoring

In many situations, individuals form estimates based on initial, random values and then adjust their estimates away from it. Tversky and Kahneman (1974) observed in their experiments that the adjustments of the estimates are most of the time insufficient and that individuals don't significantly depart away from the initial value. This observation was referred to as the "anchoring" effect. In their experimental setup, subjects were asked to estimate the percentage of African nations in the United Nations. Before estimating, the test subjects were faced with random values between 0 and 100, determined by a "wheel of fortune". It turned out that the test subjects tend to cling to the initial value. Those, receiving an initial value of 10 estimated the percentage with 25% while the other group facing an initial value of 60 estimated the percentage with 45% (Tversky and Kahneman, 1974).

Belief Perseverance

A similar phenomenon to conservatism and anchoring that once individuals formed their opinion, they obstinately hold to this belief for too long, is known as the "belief perseverance". Lord et al. (1979) explain this perseverance through two effects. First, individuals do not strive to look for new evidence that contradicts their own belief. Second, even in the case they find such evidence, they don't accept this new information in forming new beliefs. Thus, the difference between belief perseverance to conservatism and anchoring appears in the long-run neglection of new evidence compared to the slow update or insufficient adjustments, respectively. In connection with financial research, individuals propagating the Efficient Market Hypothesis may not deviate from their belief in the correctness of the model despite potentially overwhelming conflictive evidence (Barberis and Thaler, 2003).

Availability Bias

The availability bias is a judgmental heuristic, applied by individuals when estimating the probability of an event or the frequency of a class. In this situation, individuals link their estimation with available information that comes to their mind (e.g., experiences by acquaintances). As described by Tversky and Kahneman (1974), this estimation procedure is likely to be biased because not all information is retrievable or in memory at the same weight. For example, when individuals have to judge the probability of getting robbed in a larger city, they often recall their own experiences and information and form their estimate based on their limited own information instead of factual information (Barberis and Thaler, 2003).

Herding

The herding literature often includes theories on irrational, individual investors who trade based on sentiment (Nofsinger and Sias, 1999). The general definition of herding contains the tendency of many different independent agents to buy or sell the same security over a defined period of time, the herding interval. Such agents include portfolio managers, analysts, retail investors or corporate investment managers (Jegadeesh and Kim, 2010). Nofsinger and Sias (1999) differentiate agents between institutional and individual investor groups. Shleifer and

Summers (1990) argue that individual investors potentially herd based on the same signal (e.g., analyst recommendations). Individual investors are also more likely to be sensitive against fashionable trends in their investment decisions (De Long et al., 1990). The role of institutional investors in terms of herding is more ambiguous. One group advocates the contribution to market efficiency by informed institutional investors. In this case, herding will potentially move prices back to fundamentals instead of further away (Lakonishok et al., 1992; Froot et al., 1992). Another group shares the view that institutional herding causes large price movements of individual stocks that are not fundamentally justified (Nofsinger and Sias, 1999). Non-informational, institutional herding, therefore, potentially causes temporary bubbles (Dreman, 1979; Shiller et al., 1984). The beliefs and heuristics described above are the cognitive explanations for the formation of beliefs. The next section provides further insights on investor preferences that drive investment decisions.

2.5.2. Preferences

In the investment decision process, an investor must decide on how much he wants to consume, safe and what portfolio of assets he prefers to hold. Asset prices and trading decisions, therefore, rely on investor preferences. In this connection, the expected utility model is most commonly referred to the major theory on decision-making under risk, given the individual's preference. Von Neumann and Morgenstern (2007) developed the expected utility framework, which represents an individuals' utility function if the preferences satisfy four distinctive axioms (completeness, continuity, independence, and transitivity).

Due to the systematic violation of the axioms, found in the empirical studies on the expected utility framework, a number of non-expected utility theories evolved. These descriptive (how the world looks like) studies aimed to answer the questions that normative (how the world should look like) models failed in. In the financial literature, one of the most relevant descriptive theories of decisions under risk is the so-called "prospect theory" by Kahneman (1979). The prospect theory claims that investors tend to sell winners and hold on to losers which contradicts the suggestions of classical theories. The authors describe, in particular, two stages involved in the decision process. The first "editing" stage describes the framing of choices by

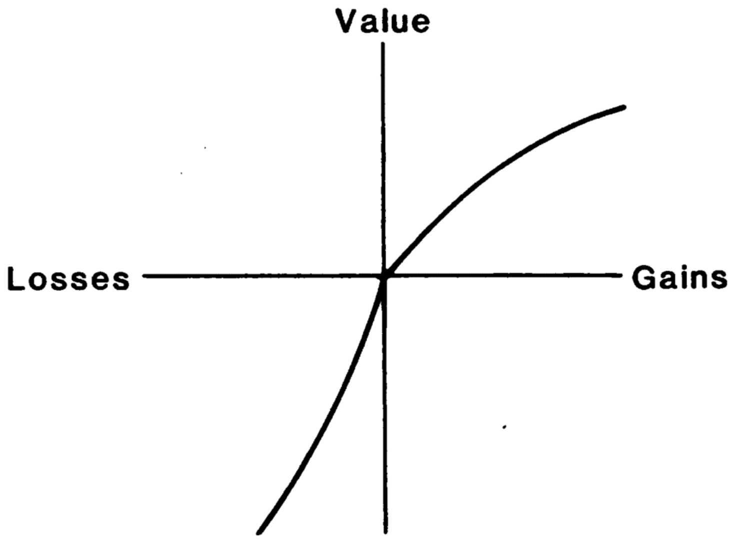

Figure 2-4. Utility (Value) Function
This figure shows the value function by Tversky and Kahneman (1981), p. 454.

decision makers regarding potential gains and/or losses in relation to a fixed reference point. In the second "evaluation" stage, decision-makers evaluate the choice based on an S-shaped utility function, which is shown in Figure 2-4. The concavity in the gains region and the convexity in the loss region displays the risk-aversion (risk-seeking) in the gains (loss) region (Kahneman, 1979). The prospect theory, shortly, concludes that an investor facing a price decline (the reference point) in his asset, chooses in an equiprobable situation of further losses or the limitation to the realized loss, the riskier option. In the hope that prices recover, the investor clings to loser stocks but is, on the other hand, disposed to sell winners (Shefrin and Statman, 1985). The greater sensitivity to losses than to gains is referred to "loss aversion" (Barberis and Thaler, 2003). Another closely related phenomenon is the disposition effect by Shefrin and Statman (1985). The authors suggest that investors tend to sell winners to quickly and hold on losers too long. The prospect theory is in their model an elementary foundation for this behavior next to the concepts of mental accounting, aversion to regret, and self-control.[11]

This section provided insights on how investor beliefs and preferences potentially lead to irrational trades. These irrational trades might occur in the form of patterns or so-called market anomalies that might repeat in history. The concepts on belief formation and preferences of

[11] Please refer to Shefrin and Statman (1985), p. 779ff, for details on the disposition effect.

decision making are, thus, the foundation for the existence of market anomalies. We briefly describe the most common market anomalies that are relevant for this dissertation in the following section.

2.6. Market Anomalies

The Efficient Market Hypothesis and the Capital Asset Pricing Model gave rise to the prominent view on rationally priced securities under uncertainty. Consequently, higher returns are inevitably related to higher compensation for risks. Price deviations that cannot be explained by systematic risk must, hence, result from model misspecifications. However, various studies in the finance literature have observed a number of corporate financing and return patterns, so-called anomalies, that earned higher returns than justified by the underlying systematic risk (Shleifer and Vishny, 1997). Fama and French (1998) explain these observations with chance deviations from rational pricing in the short run.

Daniel et al. (1998), yet, stress the disagreement amongst researchers on the interpretation of asset mispricing and associated return predictability out of such patterns. He summarizes the most common market anomalies as follows: “1. Event-based return predictability (public-event-date average stock returns of the same sign as average subsequent long-run abnormal performance); 2. Short-term momentum (positive short-term autocorrelation of stock returns, for individual stocks and the market as a whole); 3. Long-term reversal (negative autocorrelation of short-term returns separated by long lags, or "overreaction"); 4. High volatility of asset prices relative to fundamentals; 5. Short-run post-earnings announcement stock price "drift" in the direction indicated by the earnings surprise, but abnormal stock price performance in the opposite direction of long-term earnings changes.” (p. 1839-40).

These anomalies are potentially the result of preparatory, cognitive-influenced investment processes based on the beliefs and preferences discussed in Section 2.5. Ramiah et al. (2015) summarize the general links between the psychological concepts of beliefs and preferences with market anomalies in Figure 2-5. Common critiques on behavioral related explanations of asset prices enumerate the nearly unrestricted universe of behavioral patterns and the limited out-of-sample explanatory power of the models (Daniel et al., 1998). In search of the “correct” asset pricing model, a vast amount of literature evolved on risk factor and behavioral-related

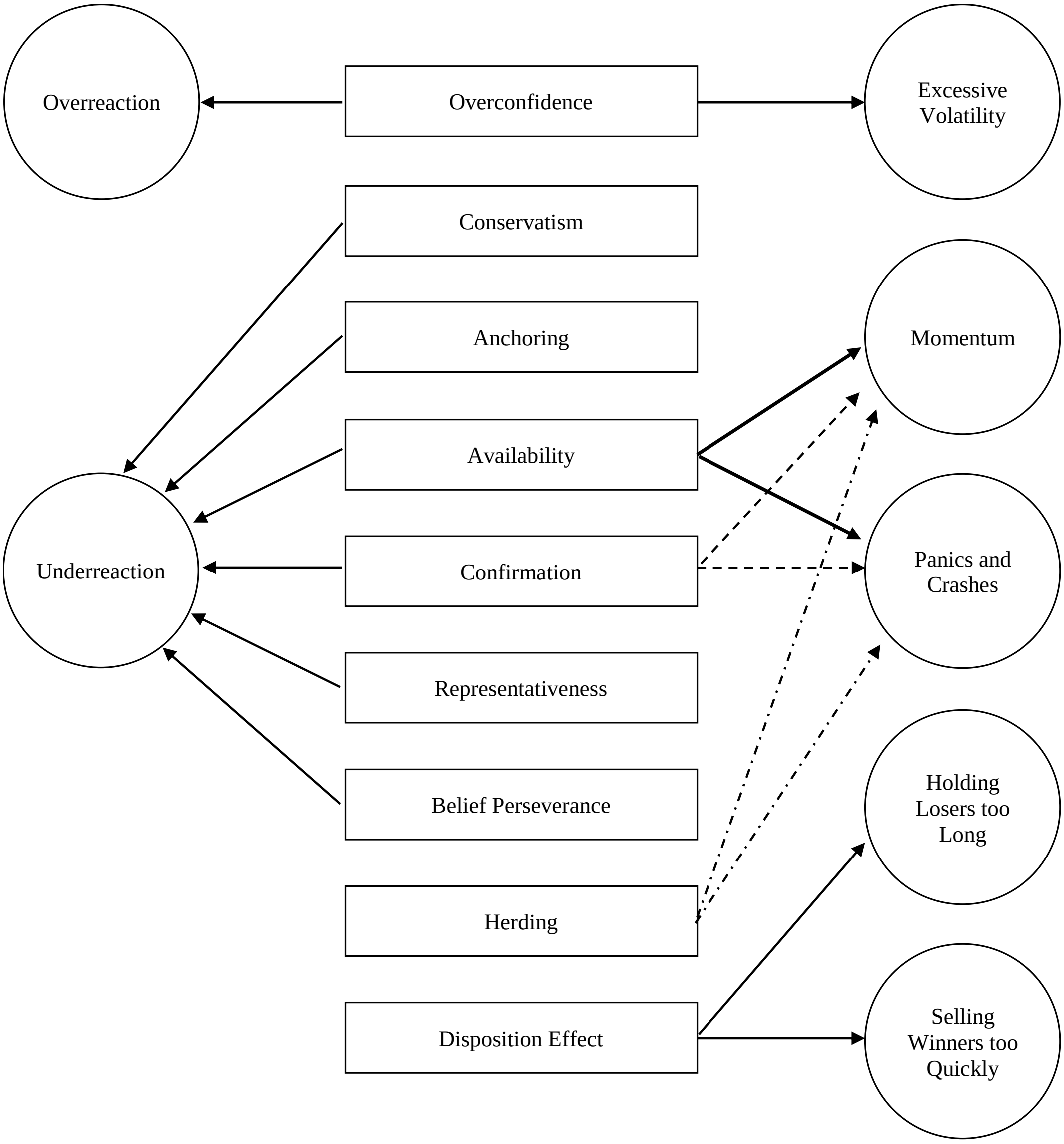

Figure 2-5. Beliefs, Preferences and Market Anomalies

This figure by Ramiah et al. (2015), p. 92, depicts the connections between individual psychological concepts and their resulting market anomalies. The rectangles contain the beliefs and preferences whereas the circles describe the market anomalies.

asset pricing models. As a result, Cochrane (2001) coined the term "factor-zoo" to describe the bulk of literature aiming to explain the cross-section of returns. In this section, we provide a brief overview of the most prominent and, for this dissertation, most relevant market anomalies: Over- and underreaction, excessive volatility, and momentum.

2.6.1. Market Over- and Underreaction

The literature documents time-invariant return patterns that several behavioral financial models comprehensively aim to explain: The over- and underreaction of financial markets. This section describes these anomalies in further details.

Overreaction

A number of studies documented the outperformance of loser stocks (stocks that performed poorly in the past) against winner stocks (stocks that performed well in the past) in the long-run. This observation is referred to as the so-called "overreaction" of asset prices resulting in a negative long-term correlation of returns (De Bondt and Thaler, 1985). In the general understanding, prices overshoot the fundamental value after a streak of good news and experience a correction of the mispricing in the long-term. De Bondt and Thaler (1985) argue in their influential work on overreaction that investors tend to overweight recent information and underrate the (base rate) data. Furthermore, the authors document a (lower) higher overreaction of prices when stocks experience (less) extreme returns in the formation period. De Bondt and Thaler (1987) find confirming results in a later study, where a zero-investment-portfolio consisting of 50 extreme losers outperforms another group of 50 extreme winners.

Several studies aimed to provide a unified behavioral theory on under- and overreaction of prices. In this connection, two of the most influential models were developed by Daniel et al. (1998) and Barberis et al. (1998). To formally describe overreaction, Barberis et al. (1998) defined overreaction as follows:

$$\begin{aligned} &E(r_{t+1} | z_t = G, z_{t-1} = G, \dots, z_{t-j} = G\ ) < \\ &\qquad E(r_{t+1} | z_t = B, z_{t-1} = B, \dots, z_{t-j} = B\ ) \end{aligned} \tag{8}$$

where G or B denotes either good or bad news z at the time period t , j is at least one or higher, and $E(r_{t+1})$ is the expected return in the subsequent time period. Hence, overreaction occurs when the average return following a series of public good news is lower than the average return following a series of bad news (Barberis et al., 1998). In other words, prices reach high levels because overly optimistic investors believe in good news in the future. Lower subsequent returns, however, are more likely since following news announcements are on average more pessimistic than the optimistic investor opinion. As stated in the previous section, Barberis et al. (1998) explain the overreaction with the representativeness bias presented by Tversky and Kahneman (1974).

Daniel et al. (1998), on the other hand, ascribe the overreaction to the self-attribution bias, where individuals become overly optimistic in the occurrence of public news that confirms their private signal or belief. Disconfirming public news, however, will rather be neglected in the investment decision process. Therefore, overreaction results, in this view, from public signals that confirm prior private signals of individuals.

Also, critics on the validity of the overreaction hypothesis arose. De Bondt and Thaler (1985) described in their own study that some part of the overreaction can be related to the seasonal January effect. Furthermore, Chopra et al. (1992) examine overreaction more distinctively and ascribe overreaction to the individual rather than to institutional investors. In another critical study, Conrad and Kaul (1993) criticize the methodology applied by De Bondt and Thaler (1985) and claim that buy-and-hold-returns are more appropriate than monthly rebalanced cumulative returns of loser-winners-zero-investment-portfolios. After the adaption of the method, Conrad and Kaul (1993) find no confirming evidence for the overreaction hypothesis.

Baytas and Cakici (1999) repeated the study by Conrad and Kaul (1993) for seven countries and generally find support for the overreaction hypothesis. The critics but also the confirming evidence on the overreaction phenomenon show the prevailing discrepancy amongst researchers on the validity of overreaction of asset prices to news.

Underreaction

Another market anomaly often jointly discussed with the term overreaction is the "underreaction" of prices to news announcements. Characteristically, underreaction of stock returns refers to the pattern of average stock returns on an event-date that have the same sign as the average subsequent abnormal performance in the long-run (Daniel et al., 1998). Another definition is the slow incorporation of news into asset prices in the short run (typically 1-12 months), corresponding to a short-term autocorrelation of returns (Barberis et al., 1998). More formally, underreaction can be stated as follows (same denotations as for overreaction):

$$E(r_{t+1}|z_t = G,) > E(r_{t+1}|z_t = B) \qquad (9)$$

The equation, therefore, formally describes that stocks underreact to good news, a mistake that is corrected with higher subsequent returns (Barberis et al., 1998). Such news can be related to a variety of events, such as stock splits, tender offers, analyst recommendations, seasoned issues of common stocks, dividend initiations and omissions, public announcements of previous insider trades or venture capital share distributions (Daniel et al., 1998).[12] The most prominent research on public news announcements in connection with market underreaction, however, deals with firm's earnings announcements and subsequent return drifts (e.g., Bernard and Thomas, 1989; Bartov, 1992; Narayanamoorthy, 2006).

Hong and Stein (1999) further decompose the underreaction of prices into two main groups, the 1) unconditional momentum-driven autocorrelation of returns and the 2) conditional autocorrelation of returns based on public news. First, the autocorrelation of returns, which is independent of public news events, seems to appear on a short-term horizon of three to twelve months. Jegadeesh and Titman (1993) observed this pattern in probably one of the most influential studies on return momentum. A reason for this unconditional short-term autocorrelation might be a slow and gradual incorporation of previously private information. Second, the conditional autocorrelation of momentum goes back to an initial public news event as described at the beginning of this section. The close association between underreaction and return momentum requires a further clarification which we attempt to provide in the next section.

2.6.2. Momentum

Momentum is commonly associated with positive short-term autocorrelation of returns for individual stocks or the market as a whole (Daniel et al., 1998). Jegadeesh and Titman (1993) demonstrated in their seminal work on momentum that a zero-investment-portfolio that goes long on winner stocks and shorts loser stocks earns significant returns within a six- to twelve-month period. The outperformance, however, diminishes more than half in the subsequent two- to three-year period. In developing a unified theory on underreaction, overreaction and momentum, Hong and Stein (1999) relinquished the cognitive-based foundation to explain the

[12] Please refer to Daniel et al. (1998), p. 1867ff, for further details on respective literature.

respective anomalies but introduced an interaction-based model (HS model forth on) in which different agents, news watchers, and momentum traders jointly evoke market anomalies.

The HS model is based on three main assumptions: 1. Newswatchers forecast future returns based on private signals on expected future developments of the asset's fundamental value, and they do not rely on past or current information, 2. Momentum traders only use information from past returns, and 3. Private information only diffuses gradually amongst the group of newswatchers. The mechanism of the HS model can simply be summarized as follows. The slow diffusion of information amongst newswatchers initiates underreaction of prices and never overreaction. Momentum traders who only apply simple trading strategies, in that they only trade based on information from past returns, anticipate the price change and accelerate the underreaction of prices towards its fundamental. However, excessive momentum trades lead to prices to overshoot compared to its long-term equilibrium. The HS model, thus, implies that early momentum traders can earn significant profits but momentum traders entering at a later stage might be exposed to the externality induced by the former group of momentum traders and hence realize losses (Hong and Stein, 1999).

There is a vast literature on momentum and its impact on cross-sectional returns (e.g., Carhart, 1997; Jegadeesh and Titman, 2002). This dissertation, however, focuses on the behavioral view and the link between a variety of behavioral biases and market anomalies. The HS model, therefore, helps us to clearly differentiate positive short-term autocorrelation caused by underreaction (gradual diffusion of information) and momentum (non-informational trades only based on past returns). The next sub-section describes another well-observed and controversially-discussed phenomenon in behavioral literature: excess volatility in financial markets.

2.6.3. Excess Volatility and Risk

A various number of studies investigated the relationship between risk and expected returns of individual stocks. In general, the literature measures risk as the covariance between a stock's return and another variable (French et al., 1987). Examples of risk measures applied in former studies are the covariance between a stock's return and the market return (Fama and MacBeth, 1973) or other factors resulting from multivariate time-series regressions of returns (Roll and Ross, 1980).

Other studies linked risk with excess volatility extracted from variations of dividends over time. Several studies argued that the stability of the present value of dividends in time implies excess volatility in the financial market in comparison to the present value suggested by the efficient markets model (Shiller, 1981; LeRoy and Porter, 1981). In this theory, the authors assumed dividends to fluctuate around a trend that is known. Based on this assumption, Shiller (1981) concludes that real dividends did not fluctuate sufficiently to justify the observed price variations in the aggregate market.

However, the till then traditional assumption on dividends experienced strenuous oppositions. Marsh and Merton (1986), for example, assert that dividends do not require to follow a known trend since stock issuance or repurchases can cause dividends to deviate from a trend. Furthermore, they argue if a firm's management uses dividends to smoothen the payout flow of their business, then stock prices should fluctuate more quickly. Yet, most studies on excess volatility have in common that they do not explain the cognitive or financial mechanisms and drivers behind the actual excess volatility. The excess volatility anomaly, therefore, remains more elusive than the formerly described anomalies, such as the over- and underreaction of prices.

We have learned about the most influential cognitive- and interaction-based theories in behavioral finance and the anomalies which traditional and behavioral researchers are still controversially discussing on. The models by Barberis et al. (1998), Daniel et al. (1998), and Hong and Stein (1999) set the theoretical foundations for many empirical studies in behavioral finance. These empirical studies, for example, investigate investor behavior in connection with traditional or social media. The next section of this dissertation focuses on the channels investors use to gather, process and disseminate information from. We, in particular, concentrate on the most common and trend-setting financial information channels that are exposed to digitalization: analyst recommendations, traditional news media and internet investment platforms as part of social media.

2.7. Information, Agents, and Media in Behavioral Finance

Information plays a crucial role in financial markets. Investors base their investment decisions on information. As such, information as an input variable induces changes in output variables,

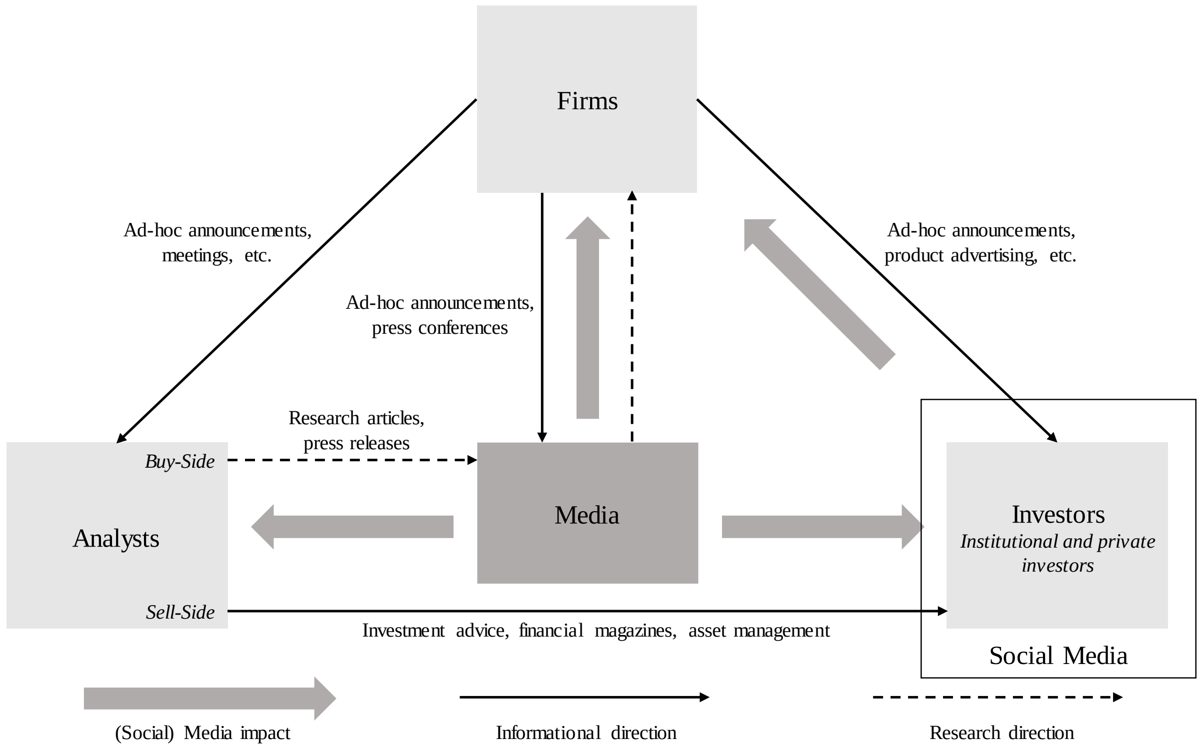

Figure 2-6. Relationship between Media, Analysts, Firms, and Investors

The figure describes the relationship between media and financial market stakeholders. The figure is based on Scheufele and Haas (2008), p. 99.

consequently measured as changes in stock prices. It is necessary, in this context, to clearly differentiate between information and news in the financial context. News disseminated via mass media provide potential information to the market. News, however, ultimately transform to information if the recipient of the news (the investor) is able to process the news. Useful news that contains a purpose and thus creates an advantage in knowledge is hence defined as information (Stanzel, 2007).

In financial markets, sources of information can be distinguished into two groups. The first group consists of agents, whereas the second group embraces the mass media. In this view, financial analysts and firms are considered as agents. Publicly listed firms must publish ad-hoc announcements in case of the occurrence of any price sensitive events. Furthermore, firms generally operate an investor relations department that ensures a transparent and conforming communication with its investors. In addition, the analyst (in particular sell-side) makes further information available for individual and institutional investors. Alternatively, investors receive information from the second group of sources: the mass media. The mass media receives, processes and disseminates information from different agents. In this connection, the impact of mass media on financial markets is inversely related to the investor's ease of access to agents. One might infer that especially irrational, uninformed investors refer to mass media in their investment decision process (Scheufele and Haas, 2008).

The term mass media comprises all media channels that reach a broad but an undefined number of audience with acoustical and/or optical statements (Maletzke, 1963). In this context, traditional media consists of mass media covering broadcast and press media, including books, magazines, newspaper, or television (Burkart, 2002). In this dissertation, the term traditional media will mainly relate to newspaper articles. Furthermore, mass media plays two distinct roles in financial markets. First, mass media acts as a pure intermediary of information. Second, mass media actively and independently affect financial markets (Scheufele and Haas, 2008). In its influencing role, mass media potentially creates or accelerates "social infection" amongst investors (Rapp, 2000). Consequently, one goal of this dissertations is, amongst others, to empirically disentangle the two roles of news media as a pure intermediary (in the spirit of the efficient market hypothesis) or the potential behavioral impact factor which causes prices to deviate from its fundamentals.

A newer stream of literature deals with the relationship between user-generated content on social media platforms with financial markets (e.g., Antweiler and Frank, 2004; Chen et al., 2014). Due to the additional complexity, this news channel will be separately discussed in Section 2.7.3 and forth on. The information flow between the different informational stakeholders in financial markets is summarized in Figure 2-6.

The previous sections explained thoroughly that investors, however, not solely decide based on true information. Additionally, it is not only the group of investors who might act irrationally. Analysts, as a group of information providers, might also be subject to cognitive biases. The following sections consequently examine cognitive biases and empirical findings related to analysts, news articles and social media and their impact on financial markets.

2.7.1. Analyst Coverage and Herding Behavior

Financial analysts are important agents in financial markets and primarily serve two different functions, the discovery of private information and the interpretation of public information. They are reputable providers of information and of high value for a broad base of investors (Ivković and Jegadeesh, 2004; Asquith et al., 2005; Chen et al., 2010). The confidence of investors in analyst recommendations is linked to the better abilities to interpret public information due to the better industry-specific knowledge, extensive training or longer experiences (Kim and Verrecchia, 1994). The superior abilities might help analysts to better understand implications of changes in accounting methods, one-time effects in the P&L, or changes of corporate strategies. Such information is particularly disclosed in earnings announcements (Chen et al., 2010). The higher complexity of companies, therefore, increases the usefulness of analysts in interpreting associated company reports. Complex companies are usually larger, have higher R&D costs, and exhibit high-growth opportunities (Chen et al., 2010).

Former studies often investigated the question whether analyst reports substitute company reports (negative correlation of returns) or rather complement/reinforce each other (positive correlation of returns). In the probably most influential studies in this area, Francis et al. (2002) and Frankel et al. (2006) rather find support for the reinforcement hypothesis and show that the information content of earnings announcements is positively correlated with that of analyst reports. The authors, however, point out that the ambiguous relationship between both reports cannot be solved, ultimately.

There are also other studies, however, that doubt the general usefulness of analyst recommendations in financial markets. The randomness of earnings outcome makes forecasts difficult for analysts. Therefore, the performance of analysts is rather benchmarked against his consensus with other financial analysts. In these situations, conformity of recommendations is found to protect the human capital of financial analysts (Scharfstein and Stein, 1990; Froot et al., 1992). Hence, analysts exhibit a tendency to publish recommendations that are in line with that of their colleagues (Olsen, 1996). This is also referred to as the herding bias as described in Section 2.5.1. As a consequence, the herding of analysts can lead to an overreaction. La Porta (1996) demonstrated, for example, that professional analysts are extremely bullish (bearish) on stocks that they are optimistic (pessimistic) about. He documents low subsequent returns for stocks with high-growth forecasts but inversely high follow-up returns for low-growth stocks. In his results, the overreaction of analysts, therefore, transmits to prices in financial markets.

In summary, herding potentially leads to an overreaction of analysts, which ultimately transmits to prices in financial markets. According to the efficient market hypothesis, stocks with high growth forecasts must earn high returns (Barberis et al., 1998). The empirical evidence, yet, showed that financial markets are imperfect, even only in a temporary time horizon. This section explained how behavioral biases of agents might impact on financial markets. The following sections provide further theoretical and empirical insights on investor sentiment expressed by traditional and social media.

2.7.2. Investor Sentiment and Traditional News Media

One of the first studies to explore the relationship between media press and financial market activity is attributed to Cutler et al. (1989). The authors examine different kinds of news and its explanatory power for aggregate stock returns. Their findings, however, do not support the hypothesis that the variance of stock price movements can be explained by news related to macroeconomic, political or world events. A number of studies followed, which not only analyzed the relation between financial markets and media press coverage (e.g., Chan, 2003; Fang and Peress, 2009) but also the media tone or sentiment expressed in these articles (e.g., Tetlock, 2007; García, 2013) or even both elements (Hillert et al., 2014). Shiller (2000) propagates the view that news media impacts on financial markets, even if the news is non-informative and just evoking a short hype. News content, thus, drives market sentiment in his opinion.

The most influential work on media sentiment is probably accounted to the work by Tetlock (2007). Tetlock (2007) states three distinctive hypothesis on the role of media sentiment in financial markets. First and the main hypothesis of his work, media sentiment and more specifically media pessimism serves as a proxy for investor sentiment. In this hypothesis, the timing of media sentiment is crucial due to the question of whether investor sentiment predicts media sentiment or reflects past media sentiment. In the former scenario, one might expect low returns following media pessimism in the short-run but high returns in the long-run. In the latter case, low returns follow media pessimism, but prices reverse to the fundamental value in the

Table 2-2. Literature on News Media and Sentiment

This table provides a selected literature overview of news media and sentiment related studies. News and returns describe the intertemporal relationship of news media and firm or market returns, where CT denotes the contemporaneous time period around the news release date (+/- 3 days), ST is the short-term horizon of up to 3 months, MLT is the mid- to long-term horizon between 3 and 36 months, + (-) describes a positive (negative) correlation in the time period, 0 characterizes a return reversal in the time period, +/0 depicts a partial return reversion in the time period, and N/A indicates a missing focus of the study on that time period.

			News and returns		
Authors and year	**Print media / Data sources**	**Time period**	**CT**	**ST**	**MLT**
Liu, Smith & Syed (1990)	Wall Street Journal "Heard on the Street" column	09/1982 - 09/1985	+	+	N/A
Barber & Loeffler (1993)	Wall Street Journal "Dartboard" column	10/1988 - 10/1990	+	+/0	N/A
Tetlock (2007)	Wall Street Journal "Abreast of the Market" col.	01/1984 - 09/1999	N/A	0	N/A
Tetlock, Saar-Tsechansky & Macskassy (2008)	Dow Jones News Service & Wall Street Journal	1980 - 2004	+	+	N/A
García (2013)	Two columns in New York Times	1905 - 2005	N/A	+/0	N/A
Hillert, Jacobs & Müller (2014)	45 national and local US newspapers	1989 - 2010	N/A	+	0
Hendershott, Livdan & Schürhoff (2015)	Thomson Reuters News Analytics (TRNA)	2003 - 2005	N/A	N/A	N/A
Bajo & Raimondo (2017)	Factiva database (majority of US newspapers)	01/1995 - 12/2013	+	N/A	N/A

long-run. In the second hypothesis, the media press sentiment reflects information that is not yet fully incorporated into prices. This effect is in the spirit of underreaction and the assumption of gradual diffusion of information amongst newswatchers according to Hong and Stein (1999). The last hypothesis states that media press sentiment only conveys stale information and hence has no impact on financial markets. The main limitation of this study is that the theory only applies to negative sentiment or media pessimism. In a more recent study, García (2013) documents also significant results for media optimism. The media sentiment effect reported in his study is most pronounced in times of economic recessions.

This dissertation frames the news media related topics around investor sentiment. The comprehensive literature on news-related market efficiency tests or the impact of media coverage (excluding the tone of the content) is, thus, out of scope of this dissertation. Consequently, we provide a brief but relevant literature overview of studies related to news media and investor sentiment. The main findings are summarized in Table 2-2.

Liu et al. (1990) examine the impact of low-cost analyst recommendations published in the "Heard-on-the-Street" column of the Wall Street Journal on stock returns. Their data overall consists of 852 recommendations, thereof 566 buy and 286 sell recommendations, in the time period between 1982 and 1985. Their findings indicate a symmetric impact of buy and sell recommendations on abnormal returns on the publication day. However, they also find significant abnormal returns in the two days before publication. Applying the event study methodology, the results yield a cumulative abnormal return of 3% in the three-day time window [-2,0] relative to the publication date. The absence of return reversals in their results implicate the informativeness of recommendations published in news media. Yet, this study only refers to second/hand information from analysts and, thus, does not consider the opinion of investors expressed in news media.

In the highly regarded study on the interaction of media sentiment and stock markets, Tetlock (2007) studies the media content of the Wall Street Journal column "Abreast of the Market". He specifically tests whether media pessimism predicts daily returns of the Dow Jones Industrial Average Index (DJIA). His US study covers the time period between 1984 and 1999 and extracts the fraction of negative words in a news column article with textual analysis based on the Harvard psychosocial dictionary. He documents an economic meaningful impact of media pessimism on the next day's market return. The effect, however, reverses after four subsequent days.

In a following but more extensive study, Tetlock et al. (2008) analyze whether textual analysis can be applied to not only predict stock returns but also earnings on the individual firm level. The authors fall back to a US data set (1980 – 2004) consisting of a variety of news articles about S&P 500 firms covered by the Dow Jones News Service and in the Wall Street Journal. Following the study by Tetlock (2007), the study focuses on the fraction of negative words in news articles based on the Harvard-IV-4 psychosocial dictionary. The data includes more than 80 quarters of earnings and 6,000 days of returns data. In contrast to the study by Tetlock (2007), the authors rather find new support for the information hypothesis. In this hypothesis, news articles convey value-relevant information that is not yet fully incorporated into stock prices. Their main finding suggests that news articles contain new information on firm earnings. In other words, news articles convey fundamental information and do not simply repeat stale information. Furthermore, the authors find weak evidence for a stock price underreaction where prices incorporate new information with a one-day delay. The predictive power for firm earnings and returns is even higher when specific news stories report on fundamental information. All in all, the findings by Tetlock et al. (2008) support the hypothesis that news media contributes to market efficiency.

In another study, García (2013) explores the interaction between the content of two columns in the New York Times and the aggregated US market return. His study falls back to news articles covering the time period between 1905 and 2005. The broad database counts 27,449 trading days in total, and media content was analyzed based on the Loughran and McDonald (2011) word dictionary. Other than the former presented studies, García (2013) stresses the symmetrical importance of positive and negative tones in news articles. The fraction of positive and negative words in news articles predict the next day aggregate return, followed by a partial reversal in the subsequent four days. Their results, in total, rather speak for the behavioral sentiment hypothesis than for the informational hypothesis with regard to the informational quality of the media content. The sentiment effect is found to be more pronounced in times of economic recessions, indicating a higher sensitivity to news in bad economic times.

Hillert et al. (2014) research on media coverage and sentiment from the momentum perspective. The authors refer to news articles published in 45 national and local US newspapers between 1989 and 2010, summing up to a total of more than 2.2 million news articles. In investigating on a buy-and-hold-portfolio that goes long on winner stocks (top 30% stocks with highest returns in the past 6 months) and shorts loser stocks (least 30%, respectively) with high media coverage, the authors find a significant momentum effect in the first 10 months which reverses afterwards. The portfolio return fully diminishes after a time horizon of 36 months. When portfolios are, furthermore, sorted by media sentiment (in addition to coverage and past returns), the momentum effect is found to be even stronger.

The former studies primarily studied the impact of news media sentiment on stock markets. Hendershott et al. (2015), however, turn the perspective and analyze whether institutional investors are already informed before the actual date of the news release. The authors, hence, examine the order flow information from institutional investors and its predictive power for several news variables, including the respective news media sentiment. In their hypothesis, a positive order flow (buy volume > sell volume) forecasts positive media sentiment. For their data, the authors refer to the Thomson Reuters News Analytics (TRNA) database covering the time period between 2003 and 2005. Their analysis includes more than 1 million observations for 1,667 stocks on 755 trading days. In their results, the institutional order flow increases (decreases) five days prior to announcements of positive (negative) news. The order flow information predicts the following sentiment of news announcements and stock returns on the announcement days. Hence, Hendershott et al. (2015) suggest that institutional investors are well-informed compared to news providers. Potential reasons that the authors mention could be related to the direct communication of institutional investors to publicly traded firms and brokerage firms, or the greater resources to process available information.

In an IPO-related study, Bajo and Raimondo (2017) research on the impact of news media sentiment on the level of IPO-underpricing, timing effects of news announcements and the associated reputation of the news provider. The authors build on US data that covers 2,814 IPOs and over 27,000 news articles recorded in the Factiva database in the time period between 1995 and 2013. Akin to other studies, news media sentiment was extracted with textual analysis tools based on the Loughran and McDonald (2011) word dictionary. Bajo and Raimondo (2017) document in their findings a positive relationship between positive news media sentiment and IPO underpricing. The effect is more pronounced for news announcements close to the IPO date and for news published by more reputable news providers.

All the literature introduced above have in common that news media sentiment is somehow related to stock returns. Even though many studies find confirmation for the behavioral sentiment hypothesis that investors react to non-informative news announcements, researchers cannot conclude with full evidence that news announcements do not convey fundamental or stale information. In reality, it is more likely that all three hypotheses (1. Behavioral sentiment hypothesis of non-informative news announcements, 2. Information hypothesis of underreaction to value relevant information that is not yet fully incorporated into prices, 3. Stale information hypothesis of old information that has no impact on financial markets) have proven their right to exist. The next section introduces the reader to the social media topic which significantly gained in importance in recent years, not only in the general society but also in financial markets.

2.7.3. Increasing Social Interaction with Digitalization of Information Channels

The rising of the internet has changed the way individuals process information, act, and communicate significantly in the past decades (Puppis et al., 2017). Different from the style of traditional media, the internet enables the society to combine different types of communication at once. Amongst others, it changes the directional components of communication, it increases the range of recipients and eases the access to information. All of these, on the other hand, define the type of communication (Neuberger, 2008). The reporting and announcement of news became continuous and permanent, globally accessible and more and more decentralized. Thus, the internet eliminates technical limitations heretofore traditional media press was exposed to. Traditional news media is commonly seen as a pure information intermediary with only limited opportunities for interactions between communities. The internet, however, enables a participative, interactive communication and reporting of information, which is self-organized and often linked globally (Neuberger, 2008). As a consequence, the borders of private and public communication disappear with the existence of the internet (Puppis et al., 2017).

Another term that was coined with this development is the “social media”. Kaplan and Haenlein (2010) define social media as a “group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content’’ (p. 61). Well-known internet-based applications in which individuals share their opinions, pictures, or video contents with the online community are, for example, Facebook, Twitter, or YouTube (Kaplan and Haenlein, 2010). Kaplan and Haenlein (2010) classify social media internet applications according to two dimensions based on theories in the field of social processes (self-presentation, self-disclosure) and media research (social presence, media richness). Self-presentation represents the idea that individuals seek to control the impression of others of themselves. The conscious or unconscious disclosure of personal information is associated with self-disclosure. In this connection, the revelation of personal information is critical for the establishment of a trustful relationship. The media dimension social presence describes how media differs in the way individuals communicate based on acoustic,

		Social presence/ Media richness		
		Low	Medium	High
Self-presentation/ Self-disclosure	High	Blogs	Social networking sites (e.g., Facebook)	Virtual social worlds (e.g., Second Life)
	Low	Collaborative projects (e.g., Wikipedia)	Content communities (e.g., YouTube)	Virtual game worlds (e.g., World of Warcraft)

Figure 2-7. Classification of Social Media

The figure depicts the categorization of internet-based applications into the dimensions of media research and social processes as proposed by Kaplan and Haenlein (2010), p. 62.

visual and physical contact. Lastly, media richness entails the concept that any type of communication aims to dissolve ambiguity and reduce uncertainty between the communication partners (Kaplan and Haenlein, 2010). The dimensions and related internet-based application types are shown in Figure 2-7.[13] The rise of the internet and social media also significantly changed the way information is produced, mediated, disseminated and consumed in financial markets (Vlastakis and Markellos, 2012). Barber and Odean (2002) describe in their study that new online investors increase their trading activity, are more speculative, but performed poorly compared to their non-online trades. Also, firms and regulators understand the gaining importance of social media and use this channel as a viable source for the disclosure of price-sensitive information. Consequently, the SEC announced in 2013 that companies are allowed to disclose sensitive information via social media channels (Lee et al., 2015).

One of the goals of this dissertation is to examine the interaction of social media activity and stock markets. More specifically, we aim to disentangle the information- and sentiment-related components of potential impacts of social media activity on financial markets. So far, the literature on social media shares ambivalent opinions and findings on its impact on financial markets. Social media applications that originally attracted the broader community instead of rather sophisticated individual investors are found to exhibit weak relationships with financial markets.[14] In this dissertation, we focus on user-generated content in social media applications that are highly dedicated to financial markets. Most of the financial studies on social media

[13] Please refer to Kaplan and Haenlein (2010), p. 60ff, for detailed descriptions on the internet-based applications and the differences in its characteristics.

[14] Sprenger et al. (2014) for example, only find weak evidence for return predictability of Twitter activity.

activity and its impact on financial market emerged in the early 2000s. These studies examine social media platforms such as internet message boards (e.g., YahooFinance, RagingBull or HotCopper), microblogging platforms (e.g., StockTwits), or more dedicated peer-based advising platforms (e.g., SeekingAlpha). More recently, another stream of literature focuses on internet search queries (e.g., GoogleTrend) to measure investor attention or sentiment from buzzwords or the frequency of words people were searching for on the internet. Before we introduce our empirical research design and results in chapter 3 to 5, we provide a brief literature review on selected and in our view most relevant studies related to the interaction of social media activity and financial markets in the next section.

2.7.4. Investor Sentiment and Social Media

The interaction of investors on social media takes place in a variety of platforms. In this section, we discuss the main characteristics and differences between internet message boards, microblogging platforms, and peer-based advising platforms before we subsequently provide a brief literature review on selected but in our view most relevant studies for this dissertation.

In one of the pioneering studies on the relation between social media and stock markets, Tumarkin and Whitelaw (2001) explored the characteristics of finance-related internet message boards. He divides internet messages boards into two categories, namely chat rooms and bulletin boards (e.g., Yahoo! Finance, RagingBull, TheLion.com, HotCopper). Chat rooms are characterized by live interactions of the chat room members, who discuss on individual stocks and the market as a whole. Furthermore, chat rooms lack historical documentation of discussions and do not allow offline users to participate in these discussions. Bulletin boards, on the other hand, allow the users to post messages at any time. Bulletin boards usually organize each topic separately in a thread. Bulletin board user do not interact “live” but can answer to other persons through the bulletin board at a later time. Internet message boards can be categorized into public and private sites. The former type, allows a broader base to participate in discussions. Private websites might be preferred by wild speculators (Tumarkin and Whitelaw, 2001). This dissertation and many other financial studies on social media mainly refer to internet message boards. The nature of postings in internet message boards is fairly simple and short. Most commonly, each message has a time stamp, contains a title and a text with up to 50 words (Antweiler and Frank, 2004).

More dedicated peer-advising platforms (e.g., SeekingAlpha) show opposite characteristics to internet message boards. Peer-advising platforms aim to provide opinions and extend analysis on financials rather than news. The articles and comments, however, are published from sophisticated investors rather than a professional analyst. A professionalized peer-advising platform ensures a higher quality of the content by reviewing submitted articles by a panel and articles are subject to editorial changes. Other users, yet, can comment and reply to articles published, sharing their opinion on the recommendation (Chen et al., 2014).

Another social media platform type used by financial investors are so-called microblogging platforms (e.g., StockTwits). These platforms are highly similar to internet message boards but most distinctive in three separate ways. First, bulletin boards categories topics in different threads. Older messages are, thus, shown prominently as long as no other user replies to the comment. Microblogging platforms are characterized by public timelines so that new messages are updated quickly and reflect the real-time opinion by topicality rather than a topic itself. Second and similar to chat-rooms, microblogging platforms have more a live-chat character based on a ticker that users are following. Microblogging users are, hence, confronted with new information in real-time and do not need to enter a bulletin board or chat-room actively. Third, microbloggers are more exposed to their reputation and have, therefore, a higher incentive to publish valuable information to attract more followers (Sprenger et al., 2014).

Most studies on social media and its relation to stock markets fell back to data extracted from internet message boards. We describe the most important studies in the following. The findings are summarized in Table 2-3.

In one of the first studies in this field, Tumarkin and Whitelaw (2001) explore the impact of internet message boards on stock markets. They extract more 181,633 messages from RagingBull on stocks in the US internet service sector with self-disclosed investor sentiment covering the time period between April 7th, 1999 to February 18th, 2000. Their results suggest that message board activity does not predict returns or abnormal trading activity but is somehow contemporaneously related to these market variables. In particular, stock returns and message board sentiment appear to be related on days with abnormal message board activity.

Table 2-3. Relation between Social Media and Financial Market Activity

This table depicts the findings of selected and most relevant studies on the relation between social media and financial market activity. The four financial, social media categories consist of IMB = internet message boards, FCA = financial commentary and analysis, MBP = microblogging platforms, and Others. The contemporaneous and predictive relationships between social media activity and RT = returns, VL = volatility, and TV = trade volume are summarized below. + (-) denotes a positive (negative) correlation, no = no significant correlation, 0 = reversal after positive contemp. Correlation, and missing values = no implications.

					Explanatory power of social media					
					Contemporaneous			*Predictive*		
Author (Year)	Category	Period	Region	Index-/ Firm-level	RT	VL	TV	RT	VL	TV
Tumarkin & Whitelaw (2001)	IMB	1999 - 2000	USA	firm	+	+		no	no	
Antweiler & Frank (2004)	IMB	2000	USA	firm	+	+	+	no	+	+
Das & Chen (2007)	IMB	2001	USA	index				(+)		
Sabherwal, Sarkar & Zhang (2011)	IMB	2005 – 2006	USA	firm	+	-	+	0	-	+
Kim & Kim (2014)	IMB	2005 - 2010	USA	both	no	no	no	no	no	no
Chen, De, Hu & Hwang (2014)	FCA	2005 - 2012	USA	firm				+		
Leung & Ton (2015)	IMB	2002 - 2008	AUS	firm	+	+	+	(no)		+
Heimer (2916)	Other	2009 - 2010	USA	firm						+
Renault (2017)	MBP	2012 - 2016	USA	index	+			0		

In the seminal work by Antweiler and Frank (2004), the authors examine how internet message board sentiment, disagreement amongst the users and the number of total board messages relate to stock returns, trading activity and return volatility. They downloaded more than 1.5 million board messages posted on Yahoo!Finance and RagingBull in the calendar year 2000 on 45 US firms listed on the Dow Jones Industrial Average (DJIA) and the Dow Jones Internet Commerce Index (XLK). Using a Naïve-Bayes algorithm, based on an individual training dataset of 1,000 messages, the authors classify the board sentiment according to buy, sell and hold categories. In their findings, the level of message posting activity is significantly related to the next day’s return, yet only of small economic impact when considering plausible transaction costs. The authors, thus, conclude that investor bullishness exhibit no predictive power for stock returns. In according to the no-trade theory, disagreement of internet message board users is found to be associated with increased trading on the same day. However, the high trading activity is followed by a reversal in the subsequent day. The authors also find that message postings predict return volatility, whereas bullishness and disagreement amongst users appear not to have a significant influence.

In another study, Das and Chen (2007) investigate the relationship between internet message board activity and stock returns on the aggregate and firm level. The authors apply five different machine learning algorithms to classify 145,110 messages according to different sentiment categories. The US data covers 24 stocks in the technology sector in the period between July 2001 and August 2001. Overall, the results of their study suggest that an aggregated sentiment index predicts the next day’s aggregate stock index returns. However, the authors cannot find a similar relationship on the individual firm-level. Since the stock index, on the other hand, does not predict board sentiment, the authors conclude that internet message board sentiment offers explanatory power for stock returns at least on the aggregate level. Furthermore, their results exhibit an inverse relationship between sentiment and disagreement. When disagreement increases amongst the internet message board users, sentiment appears to drop. Additionally, the authors find indications for a positive correlation between message board activity and sentiment. Finally, and consistent with the findings of Antweiler and Frank (2004), the authors find a strong relationship between message volume and return volatility.

Sabherwal et al. (2011) examine in another study the impact of internet stock message boards on stock returns on days without fundamental news but with high posting volume. The authors observe a pump-and-dump pattern which translates into a significant positive contemporaneous and a negative relationship in the subsequent two days between a credit-weighted sentiment index and the market return. In their study, the internet message board data was extracted from TheLion.com, consisting of a total of 12,000 messages between the period of July 2005 and July 2006. Similar to the previous studies, the authors fall back to self-disclosed sentiment and apply the Naive-Bayes algorithm to classify messages without self-disclosed sentiment. They also find that absolute sentiment scores are negatively related with contemporaneous and the next day's intraday volatility but positively associated with the number of small volume trades. Hence, the authors conclude that internet message board activity is a significant predictor of financial market activity. However, one can object that the results of their study are mostly only of small economic importance.

In a more recent study, Kim and Kim (2014) examined a broad dataset of 32 million messages on 91 firms posted on Yahoo!Finance between January 2005 and December 2010. The authors use sentiment information revealed by the investors but also apply machine learning algorithms to classify the messages according to sentiment categories. The authors find no contemporaneous and predictive relationship between investor sentiment and stock returns on the individual and aggregate level. However, the sentiment expressed on the internet message board are found to follow prior stock performances. Furthermore, the authors find no evidence for relations between message board activity and volatility or trading volume.

Deviating from the base ground of the previous studies which focused on internet message boards, Chen et al. (2014) research on the question whether social media predicts future stock returns and earnings surprises. The authors refer to a US platform for investors called SeekingAlpha. The platform contains sophisticated articles and commentaries written by more than 6,500 and 180,000 users, respectively. The coverage encompasses more than 7,000 firms in the period between 2005 and 2012. Similar to Tetlock (2007), Chen et al. (2014) conduct textual analysis of the articles and commentaries and focus on the fraction of negative words in the overall text. To disentangle the sentiment and information content of the articles and commentaries, the authors differentiate the analysis between the predictability of social media for stock performances and earnings surprises. In the case, that SeekingAlpha opinions do not contain fundamental information, one should expect no relationship between social media and earnings announcements. However, the authors find implications that SeekingAlpha opinions include value-relevant information. Consequently, the results endorse the return predictability of SeekingAlpha views for both, future stock performances but also earnings surprises. Additionally, the authors find no support for return reversals, which speaks against an overreaction of the market to content disseminated via social media.

Leung and Ton (2015) examine how internet message board activity in Australia relates to financial market activity. They extract more than 2.5 million messages from the leading Australian internet message board HotCopper for the period between 2003 and 2008. The authors exclude messages around public price-sensitive announcements which are reported to the Australian Securities Exchange (ASX) on a mandatory basis. The data covers more 2,142 stocks with more than 1.8 million observations. A positive contemporaneous relationship between investor sentiment and stock returns is only found for small capitalization stocks with high growth potential. Additionally, posting activity predicts trading volume and bid-ask spreads on the next day. The authors also document that bullish small stocks outperform bearish stocks with the absence of a return reversal. Overall, Leung and Ton (2015) find that large stocks are not affected by internet message board activity.

Heimer (2016) studies in a recent study, the role of social interaction in social media and its relation to the disposition effect, which states that investors tend to sell winners too quickly but hold on to losers. The authors refer to the myForexBook database which directly links brokerage accounts to a social network of retail traders. They rely on trading data between early 2009 and December 2010 with more than 2.2 million trades conducted by 5,693 traders. After the application of distinctive filters, the authors end up with a dataset consisting of trades by 2,598 traders. Heimer (2016) finds evidence that exposure to the social media platform myForexBook doubles the susceptibility to the disposition effect on trading activity. He explains this finding at some part with impression management by traders on social media. Efforts to maintain a positive self-image on the social media platform translate into trading strategies which enforce the disposition effect. Also, the increased social interaction between traders is found to be positively related to trading volume. Together, a positive self-image and network effects as a result of social interaction reinforce each other since the perception of success enables traders to interact more persuasively with others socially.

The previous studies examined the relationship between social media and stock market activity on the daily time horizon. Renault (2017) studies in his paper the relation between investor sentiment and stock returns on the intraday basis. In particular, he explores the sentiment information extracted from the microblogging platform StockTwits. His dataset consists of more than 60 million messages from more than 240,000 distinct users in the period between 2012 and 2016. The author compares in this study five different classifiers for sentiment classification, including an own investor lexicon (previous studies commonly referred to the word lexicon developed my Loughran and McDonald (2011)). He finds that the aggregated sentiment extracted from individual messages helps to predict intraday stock index returns of the S&P 500 ETF. Controlling for past market returns, Renault (2017) documents that the first half-hour change in aggregated investor sentiment forecasts the last-half hour S&P 500 index ETF return. His results are primarily driven by shifts in the sentiment of novice users with low experience levels in financial trading. Furthermore, the results imply a short-term price pressure with a return reversal on the following day. All in all, his findings support the sentiment and noise-trading hypothesis, that views shared in financial, social media do not convey value-relevant information.

The studies presented above show the ambiguity that researchers encounter when exploring the role of social media in financial markets. Additionally, most studies in this research field only shed light on the topic for a very short time horizon without any long-term implications for aggregated sentiment waves. This dissertation, therefore, aims to provide further insights and explore new settings in which social media induces noise or conveys value-relevant information to the market.

2.8. Summary

This chapter described the main concepts of traditional and behavioral finance theory. We summarized in the first step the core elements of the efficient market hypothesis. In this theory, financial markets react quickly and efficiently to new information so that prices immediately incorporate new information and do not reverse subsequently in the longer time horizon. However, researchers quickly began to question the basic assumptions of efficient markets. Empirical studies observed return patterns or market anomalies that cannot be explained by the actions of rational investors. Moreover, cognitive aspects became more important in explaining why financial markets might deviate from fundamentals. The new research stream is commonly subsumed and coined under the behavioral finance theory. In this connection, we introduced limitations to perfect markets which hinder rational investors to arbitrage away market imperfections. These limitations are in particular formalized within the noise-trader model which first gave a theoretical framework on risk-perception and behavioral explanations for non-rational trading patterns.

Three well-established behavioral models and theories evolved that attempt to provide a unified theory of behavioral biases and observed market anomalies. First, Daniel et al. (1998) provide in their model explanations for market over- and underreactions based on the cognitive concepts of overconfidence and biased self-attribution. Second, Barberis et al. (1998) developed a unified theory of investor sentiment and relate market over- and underreactions to cognitive biases resulting from representativeness and conservatism. Lastly, Hong and Stein (1999) created a comprehensive behavioral model to link the anomalies of market overreactions, underreactions, and momentum. Different from the previous two models, which can be classified as unified behavioral models, Hong and Stein (1999) refer to an interaction-based framework with heterogeneous agents. Hence, their model explains deviations from fundamental values with the interaction of different agents in the market rather than behavioral biases that might provoke investors to conduct uninformed trades.

After the explanation of market anomalies and possible causes for deviations in market prices from its fundamentals, this chapter briefly introduced the informational framework on how investors, firms, and analysts interact based on different media channels. We summarized the most important observations of behavioral related financial studies that deal with investor sentiment in association with traditional news media as well as social media. This dissertation aims to shed light on the elusive role of (social) media in financial markets. We, thus, empirically test in different settings how (social) media activity and investor sentiment relate to financial market activity.

Chapter 3 empirically tests the informativeness of investor sentiment in the general context and around event-specific situations to disentangle the sentiment and information content conveyed by investor sentiment in internet message boards. Hence, we evaluate the impact of the level of message board activity, sentiment, and disagreement on market variables such as stock returns, volatility, and earnings surprises.

Chapter 4 specifically deals with investor attention around M&A announcements and explores how activity in internet message boards contribute to market efficiency. Previous literature observed the phenomena of target run-ups, where target firms experience a rise in share prices before the actual M&A announcements. Some researchers explain this pattern in general with market efficiency, others with the leakage of insider information. We consequently examine how the crowd (individual investors on internet message boards) might discover such targets upfront or contribute to target run-ups.

In the last empirical setup, chapter 5 investigates the cross-sectional relation of media sentiment and stock returns in the short- and long-term perspective. In a classical asset pricing context, we conduct various empirical analysis to explore whether investors underreact or overreact to media sentiment in the short-term or whether media sentiment is perceived as a persistent risk factor in the market.

Chapter 6 finally concludes the findings of this dissertation and discusses the avenues for further research on the topic of investor sentiment and its relation to financial markets.

3. Agreeing to Disagree: Informativeness of Sentiments in Internet Message Boards[15]

ABSTRACT: We study the informativeness of the convergence of sentiments in posts on HotCopper, the largest Australian online stock message board. We find that positive sentiment is associated with noise-induced (uninformed) trading whereas negative sentiment contains value-relevant information about a firm's performance. Our empirical findings suggest that short selling activity reduces overreactions of abnormal returns in a noisy environment on the same day. Furthermore, we observe a sentiment convergence pattern around annual earnings announcements and low levels of sentiment homogeneity relate to significantly lower annual earnings surprise. This supports the view that disagreements amongst sentiments are a signal of bad news about firm fundamentals.

3.1. Introduction

In the past decade, the growth in the use of online technologies such as social media platforms to disseminate the interpretation of financial news meant that investors face a more disaggregated set of informational channels than ever before. We investigate how this form of financial innovation may add value-relevant information and how it relates to risks in stock price returns. We employ the sentiments of posts on HotCopper, the largest Australian online stock message board, to find that negative sentiment contains value-relevant information about a firm's performance and that disagreements amongst sentiments are a signal of bad news about firm fundamentals. Further, we reveal that volatilities of stock price returns induce higher levels of posting activities. Based on the seminal work of Antweiler and Frank (2004), studies on social media outlets (e.g., internet message boards, Twitter[16], Google[17]) examine how sentiment, message- and internet search volume are related to reactions in the equities markets. Studies on

[15] For helpful comments, we thank Jens Martin, Taylan Mavruk and further conference participants at the 2017 International Finance and Banking Society Asia Conference in Ningbo.

[16]The social media phenomena Twitter is rather found to be an echo of equities market activity (Sprenger et al., 2014) despite its indisputable US influence in political discussions. Recent studies relate emotions and moods on Twitter with equities market activity (Bollen et al., 2011; Zhang et al., 2011; Nofer and Hinz, 2015). However, results are ambiguous. Nofer and Hinz (2015) for example argue that follower-weighted social mood levels would predict market returns on the subsequent day. Bollen et al. (2011) only find significant relations for the mood "calm" with regards to market performance. Sprenger et al. (2014) on the other hand applied the method used by Antweiler and Frank (2004) on Twitter and found that individual stock market activity impacts on tweet activity rather than the other way around.

[17]A number of studies examine the relation between the Google Search Volume Index (SVI) and market activity to understand the role of sentiment and social media activity in terms of price discovery and investor attention. Google related studies on the other hand analyze the implications of Google search volume. They derive market sentiments on the aggregate level and

internet message boards have been contentious surrounding the return predictability of sentiment shared on social media (Tumarkin and Whitelaw, 2001; Antweiler and Frank, 2004; Das and Chen, 2007; Kim and Kim, 2014). Chen et al. (2014), however, find in their study that the fraction of negative words of articles and comments published on the peer-based advising platform Seeking Alpha predicts returns over different time horizons. The difference in results to other studies is mainly explained by the broader sample and the more sophisticated design of messages posted on Seeking Alpha. Nonetheless, the results only relate to negative sentiment and the relation between positive sentiment and equities market activity has received little attention, even though it is equally or even more so for internet message boards.[18]

Tumarkin and Whitelaw (2001) find a contemporaneous relation between message board activity and returns. Antweiler and Frank (2004) find significant, but the negative contemporaneous correlation between stock returns and the message volume the following day. Das and Chen (2007) find no significant relationship between internet message board sentiment and individual stock prices. However, at the aggregate level, results indicate a relation between sentiment and stock prices. In another study, Kim and Kim (2014) compare self-disclosed and machine classified sentiment based on the Naïve Bayes algorithm and find little evidence that sentiment would predict future stock returns at an individual or aggregate level (also for market volatility and trading volume). Chen et al. (2014) show in their study that opinions on Seeking Alpha strongly predict future returns and earnings surprises. Similar to other media related studies, they find no significant relation for positive word categories and therefore focused on the relation of the negativity of articles and comments on future stock performances. Leung and Ton (2015) find that message board activity strongly relates to small market capitalization activity. We argue that bullish stock portfolios outperform bearish stocks in the same month,

suggest that Google search volume predicts market developments (Da et al., 2011; Da et al., 2015). Da et al. (2011) find that search frequency in Google (SVI) is a direct measure of retail investor attention and that SVI predicts higher stock prices the subsequent two weeks with potential return reversals within one year. Drake et al. (2012) show that investor information demand increases market efficiency surrounding earnings announcements. Other studies relate SVI with market indices and volatility (Vlastakis and Markellos, 2012; Vozlyublennaia, 2014; Andrei and Hasler, 2015; Da et al., 2015). As suggested by Tetlock (2007), negative terms in English language are more reliable for identifying investors sentiment. Consequently, only applied negative terms to form their SVI based FEARS index used to measure the household sentiment. They find that the FEARS index predicts market returns, revealing contemporaneous low returns but higher returns the subsequent day. This might be consistent with the noise trading theory and the sentiment-induced divergence of asset pricing from the fundamental values.

[18] Antweiler and Frank (2004) and Leung and Ton (2015) show that sentiment expressed on internet message boards are strongly biased towards positive sentiment.

however with diminishing differences in subsequent months. Lead-lag-regressions show predictive power of message volume and sentiment for the next two days for small stocks however only with little economic significance. Renault (2017) provides evidence that the previous day last half-hour change in investor sentiment helps to forecast intraday stock index returns. In this study, we further attempt to analyze the role of stock message boards in the price discovery process. We examine whether positive and negative sentiment convey different levels of value-relevant market information and further elaborate on implications for financial regulators.

A common term used in relation with investor sentiment and noise trading is the term “Bullishness”. Brown and Cliff (2004) define “Bullishness” as investor sentiment attached to some degree of outperformance of stocks, generally measured by their positive abnormal returns. However, the classical finance theory does not support the role of investor sentiment. It argues that mispricing will be offset by rational investors who statistically optimize their portfolio, leading to a price equilibrium based on arbitrage. A deviation of market pricing and a firm’s fundamental would, therefore, result from an uninformed demand shock and limits on arbitrage (Baker and Wurgler, 2006). De Long et al. (1990) argue that if irrational noise traders trade based on their erroneous stochastic beliefs, they would affect prices and create risk in the asset pricing. As a result, excess market volatility, divergence from fundamental values and the reversion of stock returns are surrounded by market activity induced by noise traders. According to this theory, when sentiment rises, uninformed traders increase their capital allocation to assets with higher risk classes and will drive prices away from their fundamental. This is followed by returns reversal and convergence to the price equilibrium (Kim and Kim, 2014). If returns do not reverse hereafter, it implies that sentiment conveys value-relevant information for market participants.

Our study differentiates itself from former studies on internet message boards. We examine the relation between distinctive sentiment environments (positive and negative) and equities market activity. Former studies mainly focused on the influence of average bullishness scores or only negative sentiments on social media. However, we show that the segmentation of sentiment is essential in sentiment analysis with significantly distinctive implications for equities markets (uninformed vs. informed trading). Using the sentiment disclosed by posters on HotCopper, we do not rely on machine learning algorithms compared to former studies

(Tumarkin and Whitelaw, 2001; Antweiler and Frank, 2004; Das and Chen, 2007). For example, Das and Chen (2007) find that the popular Naïve-Bayes Algorithm revealed only a 50% accuracy on sentiment classification for their study. Our sample is, therefore, free from classification bias.

Second, from the best of our knowledge, this is the first study to shed light on the relationship between short selling activity and sentiment expressed on internet message boards. Former studies have argued that limits on arbitrage resulting from short time horizons and higher cost/risk profiles of especially low capitalization and growth stocks might prevent contrarian arbitrageurs to trade against noise traders (De Long et al., 1990; Shleifer and Vishny, 1997; Baker and Wurgler, 2007). Short selling, as one mean of arbitrage, is regulated by the Corporations Act 2001 and the Corporations Regulations 2001 in Australia.[19] Most short selling activity in Australia is based on covered short sales since naked short sales are generally restricted except given circumstances. A violation of reporting would result in an offense as defined by the Australian Securities & Investments Commission (ASIC). In our study, we use the short selling position data set from the ASIC to examine whether short selling activity contributes to the price stabilization process in a noisy or uninformed trading environment.

Third, we analyze a broad data sample of 3,050 stocks with 4,586,271 stock forum messages between January 2008 and May 2016. Previous studies usually focused on tech companies or on the most active firms on the internet message board of up to 100 stocks for an only short period of time (usually less than one year). Our broad sample, therefore, allows us to examine the distinctive relationship between social media and equities markets on the aggregate and individual stock level over a longer time horizon. Due to their focus at the aggregate index level, former studies were thus prone to cancelation errors of overly optimistic or pessimistic individual stock sentiments on the aggregate level (Kim and Kim, 2014) and they were also subject to time effects.

Fourth, we examine the relation between sentiment homogeneity and firm's fundamentals around annual earnings announcements as well as equities market performance. As previous studies only examined how agreement on sentiment (namely the standard deviation of posted sentiments) relates to future stock returns and volatility, we furthermore analyze how sentiment

[19] http://asic.gov.au/regulatory-resources/find-a-document/regulatory-guides/rg-196-short-selling/.

homogeneity cross-sectionally and contemporaneously relates to activities in the equities market.

We show that positive sentiment shared on internet message boards induce noisy (uninformed) market trading activities with significant contemporaneous abnormal returns but negative return reversals the following days. We find empirical evidence that short selling activity reduces overreactions on positive sentiment expressed on internet message boards on the same day. Due to costly short selling activities for especially low capitalization and growth stocks, we argue that only informed short-sellers would take the risk to bet against positive sentiment traders. Furthermore, we find that stocks with negative sentiment postings experience significantly lower abnormal returns. These effects are made visible by the segmentation of an average sentiment score into positive or negative sentiment scores. The results hold for small capitalization stocks. Additionally, we show that stock price volatility and internet message posting volume correlate with each other, however with stronger impact from volatility to posting volume. For the aforementioned implications on sentiment and stock price volatility, we find that significance and magnitudes in results also strongly depend on the differentiated analysis on the aggregate index or individual stock level.

Finally, we observe an agreement convergence pattern prior to annual earnings announcements, and we show that stocks with low levels of sentiment homogeneity (low sentiment and/or agreement) experience significantly lower annual earnings surprise. This supports the view that disagreement and/or low sentiment levels amongst investors are a signal of bad news about firm fundamentals. The overall findings suggest that positive and negative sentiment are drivers for noise- and value-prompted price movements, respectively. Also, we show that the level of sentiment homogeneity is an indicator of changes in firm's fundamentals before annual earnings announcements.

In section 3.2, we describe the message board data and financial data. Section 3.3 shows the event study results. Section 3.4 encompasses the main regression analysis and results from vector autoregressions as well as Granger causality tests. Section 3.5 describes the cross-sectional portfolio performance based on sentiment and agreement as well as results from regressions regarding earnings surprises. Section 3.6 concludes the overall findings.

3.2. Data and Research Design

The data for this study was downloaded from the HotCopper Message Board, Australia's largest message board with more than 250,000 registered members and more than 200,000 unique website visitors every month. Most members of this internet message board are Australian investors and share market traders generating more than 21 million monthly page views. In Australia, HotCopper has 18 times the traffic compared to its nearest competitors and comparable financial websites. HotCopper is a free access forum and enables investors to discuss financial topics such as the ASX and foreign stock markets, IPOs or Foreign Currency Trading.[20] Our dataset contains 4,586,271 forum messages posted in the period from January 2008 to May 2016. We include examples of opinions and messages extracted from HotCopper in Table 3-1 to provide a sense of information depth and content of board messages. Figure 3-1 compares the posting activity for small and large stocks of our current dataset with our previous study (Leung and Ton, 2015). Small stocks still account for most of the posting activity with similar pattern compared to the past study. Peaks of message board activity have moved to the opening (10 a.m.) trading hours of the ASX.

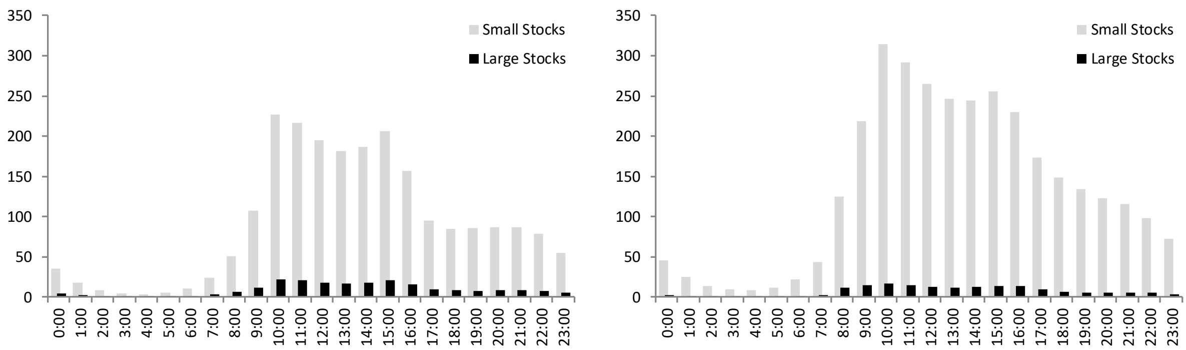

Figure 3-1. Message Postings (in thsd.): 2003-2008 vs. 2008-05/2016

3.2.1. Internet Stock Message Board Sentiment and Agreement

Previous studies were compelled to apply text classifier for sentiment classification of individual board messages since board users did not directly reveal their recommendations (Buy

[20] For more information please visit https://hotcopper.com.au/about/.

vs. Sell) on internet message boards. Outcomes, therefore, relied on the quality and the accuracy of the applied methods. Our study has the advantage to fall back on board messages with self-disclosed sentiment and therefore lowers the risk of false sentiment classification. HotCopper allows its users to classify their sentiment along seven categories: “Hold”, “Short-term Buy”, “Long-term Buy”, “Buy”, “Short-term Sell”, “Long-term Sell” and “Sell”. As time effects are difficult to measure (e.g., long-term sell vs. sell), we assign all short-term, long-term and sell/buy recommendations to “Sell/Buy”. Different findings on the relation of internet board message sentiment and market activity are existent and may be attributed to different measures of sentiment. In this connection, Baker and Wurgler (2007) conclude that one of the key issues for researchers to address is the matter of sentiment measurement and the quantification of its impact. Another question which needs to be answered is also the different nature of stock message boards and its degree of professionalism. But this must be elaborated in another study and is a not focus in this one. Thus, existing findings on social media sentiment and their impact on capital markets are ambiguous.

Some authors find contemporaneous correlations between sentiment and stock returns (Antweiler and Frank, 2004); others show that only negative sentiment predicts future stock returns (Chen et al., 2014). In turn, Kim and Kim (2014) argue that stock returns rather condition sentiment reaction than the other way around. All studies have in common, that analysis was either based on average sentiment scores or only contemplated the impact of negative sentiment on the capital market. To examine whether sentiment partitioning may improve the predictive power of message board sentiment scores, we employ the standardized Bullishness index from Antweiler and Frank (2004) for our sentiment analysis and disentangle the average sentiment index into a dedicated positive and negative sentiment score. Only buy and sell messages (forth on called financial relevant messages) are included into the bullishness index. The total number of relevant messages is, therefore, defined as $M_{i,t} = M_{i,t}^{BUY} + M_{i,t}^{SELL}$.

Following Antweiler and Frank (2004) the standardized bullish index $Bullishness_{i,t}$ for stock i at time t is defined as:

$$Bullishness_{i,t} \equiv \frac{M_{i,t}^{BUY} - M_{i,t}^{SELL}}{M_{i,t}} \cdot \ln(1 + M_{i,t}) \tag{10}$$

To measure the differentiated impact of positive and negative sentiment, we define the positive and negative sentiment for stock *i* on day *t* as:

$$PosSentiment_{i,t} \equiv \ln(1 + M_{i,t}^{BUY}) \tag{11}$$

and

$$NegSentiment_{i,t} \equiv \ln(1 + M_{i,t}^{SELL}) \tag{12}$$

We also include an agreement index $A_{i,t}$ (see Antweiler and Frank (2004)) to measure the degree of agreement between sentiments of messages. This score is also used in a later section to examine how sentiment and agreement jointly convey fundamental information around company events (especially earnings announcements).

The agreement index $A_{i,t}$ is defined by:

$$Agreement_{i,t} \equiv 1 - \sqrt{1 - \left(\frac{M_{i,t}^{BUY} - M_{i,t}^{SELL}}{M_{i,t}}\right)} \tag{13}$$

3.2.2. Financial Data

We obtain individual daily trading data from Compustat / Securities Industry Research Centre of Asia-Pacific (SIRCA) for our observation period of January 1st, 2008, to May 31st, 2016. The data contains exchange ticker codes for each transaction with a timestamp, price, price returns, highest and lowest daily price. We calculate $Volatility_{i,t-30,t-1}$ as the 30 trading-day standard deviations of returns prior to day *t*. Following Chakrabarty et al. (2012), we define daily volatility, $Volatility_{i,t}$, as the relative difference between the highest and the lowest price of the stock *i* on day *t* scaled by the daily closing price.

We use the market-weighted All Ordinaries Index to proxy the market performance since it includes 500 constituents and is, therefore, the broadest index in the Australian market. Using the market index, we calculate abnormal returns, $AbRet_{i,t-j,t-k}$, as the difference between the firm's compounded stock return and value-weighted market return over a defined holding period *j* to *k* (see Akbas (2016)).

Table 3-1. Example of HotCopper Messages

This table represents four examples of messages posted in the thread ‘Quarterly report due this week’ on the internet message board HotCopper (https://hotcopper.com.au).

Ticker	Thread time	Post ID	Posting time	Disclosure	User	Message	Sentiment
EDE	27/04/16 17:13	17615495	28/04/16 07:07	Held	Espinsight	My thoughts are that the Quarterly reports give a neat overview and can contain a clear vision of expectations, particularly for new investors, or those considering investing. Should likely be very positive and confirming.	Buy
EDE	26/04/16 10:41	17598450	26/04/16 12:38	Held	RULES	Plans in place to increase Colorado's capacity to 24,000,000 gals p.a. by late this year to early next year at approx 20% margin on $25.00/gal should pave the way for the cash you reckon is short. Time will Tell.	Buy
EDE	27/04/16 17:13	17618313	28/04/16 11:05	Held	brassmad	As a newbie to HC it's sometimes quite difficult to put together the structure of companies, so your post has helped me in that regard. I'm gradually getting my head around the acronyms but there's one in your post that I can't decipher..... could you let me know what R/I stands for? Really enjoy reading MOST of the comments posted.	Hold
EDE	26/04/16 10:41	17596748	26/04/16 10:41	Not Held	Colstone	The involvement of the state of Georgia as well giving tax breaks and no doubt future business will only benefit this company. But after having a look through their statements in the weekend they have basically no cash at the moment and plans to spend 68mil building a plant that will take years...	Sell

Table 3-2. Summary Statistics: On Firm/Trading Level

This table reports the summary statistics of the main internet message board and financial control variables. The observations are on a firm-day level. $LogMessages_{i,t}$ is the log transformation $(1+M_t)$, $Bullishness_{i,t}$ is the standardized bullishness index defined in formula (10), $PosSentiment_{i,t}$ and $NegSentiment_{i,t}$ describe the positive and negative sentiment denoted in formula (11) and (12), $Agreement_{i,t}$ is the agreement index described in formula (13), $AbRet_{i,t}$ describes the firm's abnormal return, calculated as the difference of compound raw returns and value-weighted market return, $Volatility_{i,t-30,t-1}$ is the 30-day-standard deviation of returns prior to day *t*, $Upgrade_{i,t}/Downgrade_{i,t}$ describe the number of analyst upgrade/downgrade recommendation on day *t*, $PosMeanES_{i,t}/NegMeanES_{i,t}$ denote dummy variables for positive/negative mean earnings surprise on day *t*.

	N	Mean	Median	Std. Dev.	10^{th} Pctl.	25^{th} Pctl.	75^{th} Pctl.	90^{th} Pctl.
Message board variables								
$LogMessages_{i,t}$	390,842	1.292	1.099	0.703	0.693	0.693	1.609	2.303
$Bullishness_{i,t}$	390,842	1.096	1.099	0.840	0.693	0.693	1.609	2.197
$PosSentiment_{i,t}$	390,842	1.218	1.099	0.733	0.693	0.693	1.609	2.197
$NegSentiment_{i,t}$	390,842	0.136	0.000	0.381	0.000	0.000	0.000	0.693
$Agreement_{i,t}$	390,842	0.925	1.000	0.246	1.000	1.000	1.000	1.000
Financial control variables								
$AbRet_{i,t}$	390,842	-0.001	-0.001	0.070	-0.058	-0.024	0.020	0.059
$AbRet_{i,t-1}$	390,842	-0.003	-0.002	0.067	-0.055	-0.023	0.016	0.049
$AbRet_{i,t-2}$	390,842	-0.002	-0.002	0.068	-0.057	-0.024	0.017	0.052
$Volatility_{i,t-30,t-1}$	390,842	0.047	0.039	0.038	0.017	0.026	0.057	0.083
$Upgrade_{i,t}$	390,842	0.043	0.000	0.412	0.000	0.000	0.000	0.000
$Downgrade_{i,t}$	390,842	0.074	0.000	0.630	0.000	0.000	0.000	0.000
$PosMeanES_{i,t}$	390,842	0.007	0.000	0.083	0.000	0.000	0.000	0.000
$NegMeanES_{i,t}$	390,842	0.002	0.000	0.046	0.000	0.000	0.000	0.000

In order to examine the value-relevant information content of board messages around financially relevant company events, in this case yearly earnings announcements, we obtain data on analyst recommendations and earnings forecast from the IBES summary, surprise and detail history file. The IBES summary file contains information about the number of recommendation upgrades/downgrades for firm *i* on day *t* ($Upgrade_{i,t}/Downgrade_{i,t}$). The IBES surprise history file tracks the mean consensus Earnings per Share-estimate for a particular fiscal period. We use this metric to assign positive and negative mean earnings surprise dummy variables to firm *i* on day *t*, ($PosMeanES_{i,t}/NegMeanES_{i,t}$). We also constructed median consensus analyst forecast to calculate annual earnings surprises for our analyses. The approach will be detailed in a later section.

3.2.3. Sample Characteristics and Summary Statistics

Table 3-2 presents the descriptive statistics of the main internet message board and financial variables used in our main analysis. It is apparent that positive sentiment dominates the underlying sentiment on HotCopper. Similar to previous studies, board message users rather express positive opinions and might want to avoid to speak against their own interest (e.g., Antweiler and Frank, 2004; Kim and Kim, 2014; Leung and Ton, 2015). The trend towards positive sentiment also comes along with a high agreement amongst users. Where the Agreement index might take values up to 1, the average Agreement score is 0.925, and in more than half of the firm-days, users agree on their sentiment (median of 1). Abnormal returns are slightly negative, which might result from larger firms outperforming smaller firms during the sample period and our use of a value-weighted market index similar to Kim and Kim (2014). We find a higher number of analyst downgrade recommendations but a higher number of firm-days with positive earnings surprises on day *t* during our sample period.

To examine the impact of sentiment and the value relevant information content of internet message board, we mainly conduct four different types of analysis with varying underlying datasets:

(a) **Event study dataset**: We segment our event study sample into events triggered by an abnormal level of positive or negative sentiment expressed on day *t*. Additionally, we analyze the impact of events with increasing minimum number of positive or negative messages [10, 20, 30, 40] on day *t*. Events triggered by Buy/Sell messages sum up to a range between 1,093/17 (min. 40 messages) and 13,126/493 (min. 10 messages). Again, the data set implies a high bias towards positive related board messages.

(b) **Regression dataset**: The total HotCopper message board data set covers 3,362 stocks (2,700 stocks with at least 100 messages) whereas the trading dataset contains 3,778 stocks between January 2008 and May 2016. We only deleted messages if no trade occurred on the day and the total number of stocks covered in our regression results in 3,050. The regression sample on firm-day level contains 283,585 to 390,842 observations depending on the holding period of the regression. The market capitalization of the stocks with available data has a mean of 691.1 million Australian dollars (AUD). Similar to our previous study, we find that the majority of stocks discussed on HotCopper can be classified as small stocks with a median stock capitalization of 19,2 million AUD.

(c) **(Panel) Vector autoregression (VAR) data set**: We analyze the causal relationship between (c.1) board sentiment and abnormal returns and (c.2) message volume and daily price volatility by using (panel) VAR models as well as the Granger causality test on the aggregate and individual stock level. We apply lag order selection tests to determine the optimal lag length for our (panel) VARs. On the aggregate (individual) level, the optimal leg length of 4 (3) for c.1 and 3 (3) for c.2 results in data sets with 380 (42,872) and 797 (42,872) observations, respectively.

(d) **Regressions around earnings announcements data set**: In order to examine the value content of internet message boards around company-specific events, in this paper annual earnings announcements, we construct earnings surprises using analyst forecasts and historical numbers. We obtain 479 observations and 560 observations for a cumulative period of [t-7, t-1], respectively.

3.3. Event Study

In section 3.2.1, we have argued that internet message board postings might have varying relations to the stock market depending on the sentiment expressed. Based on the findings of Tetlock (2007), we expect negative opinions on internet message boards to have a more pronounced influence on capital market features. In another study, Chen et al. (2014) find no correlation between positive sentiment and market features. De Long et al. (1990) argue in their study that noise traders act based on their erroneous beliefs which affects asset prices, moves them away from their fundamental values and reinforce market volatility. Due to the bullishness nature of message boards (e.g., Antweiler and Frank, 2004; Leung and Ton, 2015), one might argue that especially negative sentiment contains more value relevant information as internet message board users would like to discuss negative associated firm information in only very specific situations.

Hence, we test the hypothesis that positive and negative sentiment have a significantly different relation to stock market prices and that herding of bullish internet message board users quickly reflects in the market but only remain temporarily influential. Abnormal returns were

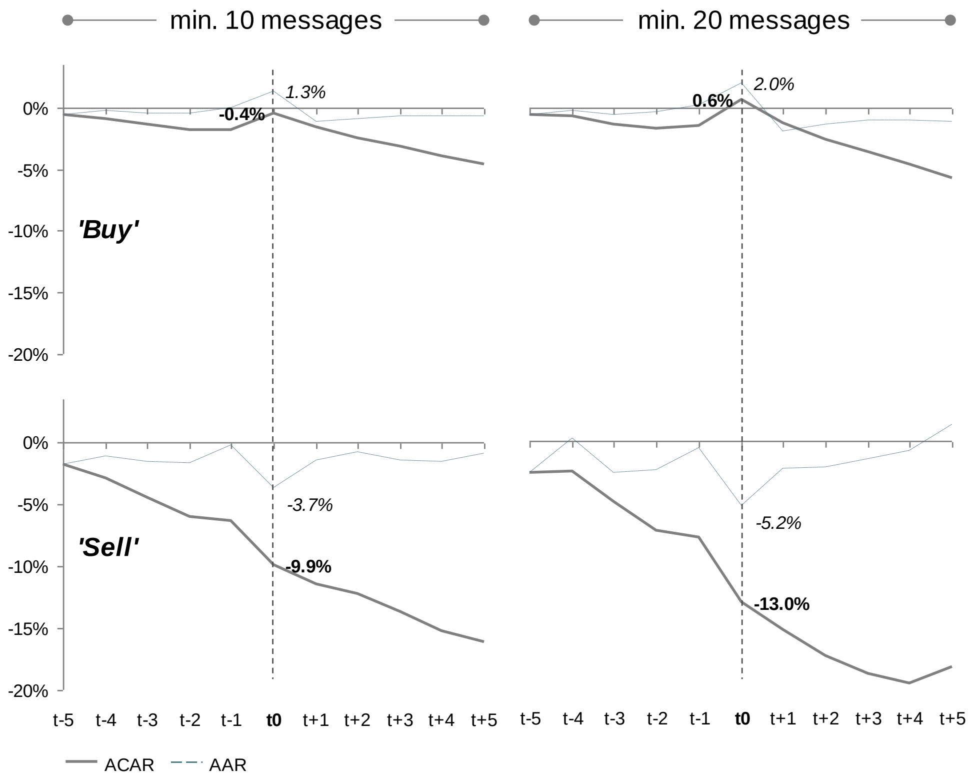

Figure 3-2. Abnormal Message Board Activity (min. 10/20 Buy/Sell Messages)

calculated based on the market excess model in order to examine the relationship between message board activity and abnormal markets features. We define an event as a day *t* with abnormal message posting volume (at least 10 buy/sell messages), where message volume on day *t* exceeds double the standard deviation of message posting volume in the previous five days. We, therefore, determine an event window of $t-5$ to $t+5$ and control for overlapping events and thus momentum-induced noise. Consequently, we only included the first event within a seven-day period. Results are shown in Figure 3-2 (min. of 10/20 messages) and Figure 3-3 (min. of 30/40 messages) and are tabulated in Table 3-3 ("Buy"-events) and Table 3-4 ("Sell"-events). On the event day, average cumulative abnormal returns (ACAR) are significant for positive and negative triggered events, however with lower impact for positively related events. Applying the parametric t-test and the non-parametric Wilcoxon-test, we find for event days with minimum of 20 messages highly significant ACARs of 2.03% (buy) and -5.23% (sell). Robustness tests on events with a minimum of 30 or 40 messages show similar and even stronger results in their magnitude.

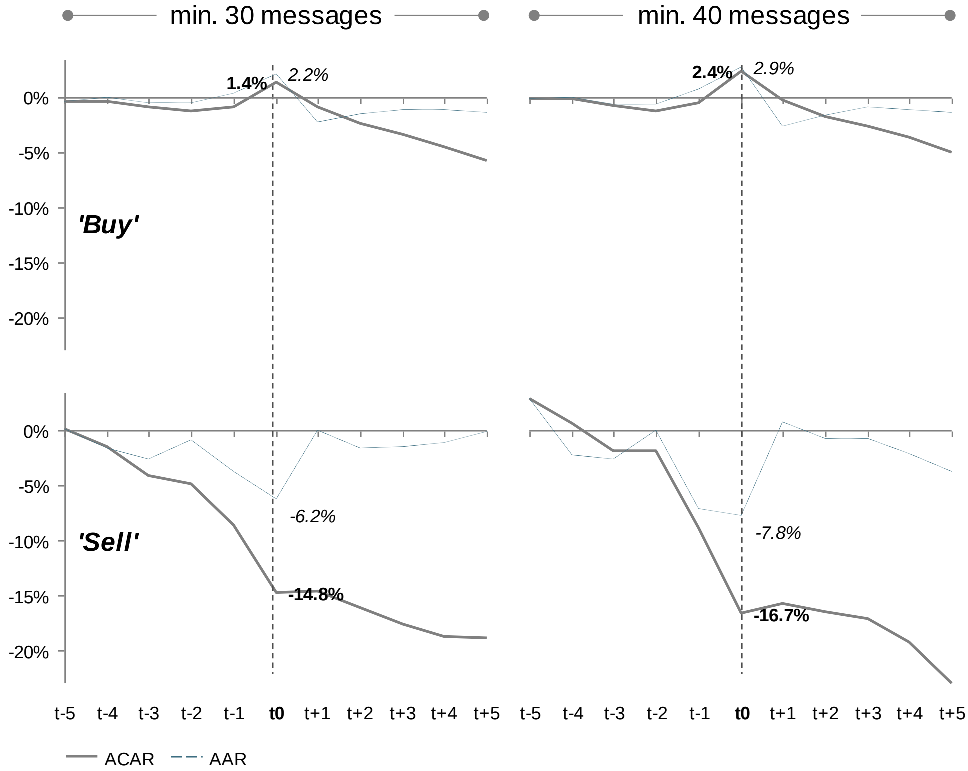

Figure 3-3. Abnormal Message Board Activity (min. 30/40 Buy/Sell Messages)

We find an increasing and significant trend of ACARs from 1.33%^{min10MSG} to 2.87%^{min40MSG} on the event day *t* with an increasing number of positive messages. Of even higher impact, we observe an increasing trend of ACARs from -3.65%^{min10MSG} to -7.75%^{min40MSG} on event day *t* for negative messages. In comparison, the median CARs on the event day tend to be significantly lower than the average CARs for both sentiment segments. This indicates that results are driven by particular stocks which are either hyped or negatively talked about in message boards. For events triggered by positive messages, CARs before the event [*t-5, t-1*] experience a drop in significance with increasing number of postings. However, we find significant negative CARs [*t+1, t+5*] following the event day *t* from -4.16% (min. 10 messages) to -7.40% (min. 40 messages). These findings support the noise trading theory by De Long et al. (1990). If sentiment rises, noise traders will invest in more risky assets, and the uninformed demand drives asset prices above the fundamental value.

Table 3-3. Event Study Results: "Buy"-Events

This table describes the average and median cumulative abnormal returns (CAR) for varying event windows surrounding abnormal positive posting volume. The significance is tested based on the parametric t-test and the non-parametric Wilcoxon test. ***, **, and * describe significance at 1%, 5% and 10%-levels, respectively.

Min. of 10 buy messages (n = 13,126)	**ACAR (%)**	**Median CAR (%)**	**t-test (t-value)**	**Wilcoxon (Z-Score)**
[-1,0]	1.33	0.49	12.33***	11.13***
0	1.33	0.38	15.49***	14.07***
[0,1]	0.22	-0.12	2.21**	0.50
[-5,-1]	-1.77	-1.55	-13.66***	-16.83***
[1,5]	-4.16	-3.27	-33.91***	-36.62***
[-5,5]	-4.59	-3.74	-22.92***	-24.76***
Min. of 20 buy messages (n = 4,247)				
[-1,0]	2.26	0.96	9.59***	9.30***
0	2.03	0.51	10.72***	9.54***
[0,1]	0.16	-0.18	0.74	-0.69
[-5,-1]	-1.40	-1.35	-5.30***	-7.18***
[1,5]	-6.25	-5.16	-26.17***	-27.24***
[-5,5]	-5.61	-4.89	-13.91***	-15.20***
Min. of 30 buy messages (n = 1,972)				
[-1,0]	2.67	1.06	6.96***	6.64***
0	2.23	0.52	6.89***	5.97***
[0,1]	-0.04	-0.39	-0.11	-1.14
[-5,-1]	-0.78	-0.95	-1.81*	-2.80***
[1,5]	-7.17	-6.14	-19.32***	-20.37***
[-5,5]	-5.72	-5.08	-8.74***	-9.60***
Min. of 40 buy messages (n = 1,093)				
[-1,0]	3.67	2.02	6.68***	6.43***
0	2.87	0.72	6.10***	5.29***
[0,1]	0.30	-0.50	0.62	-0.63
[-5,-1]	-0.44	-0.67	-0.71	-1.32
[1,5]	-7.40	-6.51	-14.82***	-15.52***
[-5,5]	-4.95	-4.71	-5.44***	-6.19***

Subsequently, prices then revert to their fundamental values with associated lower returns which also comes along with excess market volatility (e.g., Kim and Kim, 2014). We do not observe a "pump and dump" behavior (e.g., Sabherwal et al. (2011)), where long positions are built before the event day *t* (positive hype) and then sold subsequently to generate profits.

Contrarily to the results surrounding an abnormal volume of positive messages, the analysis of negative sentiment indicates another finding. For negative sentiment, CARs before the event

Table 3-4. Event Study Results: "Sell"-Events

This table describes the average and median cumulative abnormal returns (CAR) for varying event windows surrounding abnormal negative posting volume. The significance is tested based on the parametric t-test and the non-parametric Wilcoxon test. ***, **, and * describe significance at 1%, 5% and 10%-levels, respectively.

	ACAR (%)	Median CAR (%)	t-test (t-value)	Wilcoxon (Z-Score)
Min. of 10 sell messages (n = 493)				
[-1,0]	-3.89	-3.16	-4.33***	-7.89***
0	-3.65	-1.87	-6.00***	-8.52***
[0,1]	-5.14	-3.27	-7.14***	-8.43***
[-5,-1]	-6.28	-3.86	-5.86***	-7.46***
[1,5]	-6.17	-3.42	-7.02***	-7.68***
[-5,5]	-16.12	-10.89	-10.18***	-10.66***
Min. of 20 sell messages (n = 100)				
[-1,0]	-5.75	-5.78	-2.11**	-3.97***
0	-5.23	-3.13	-2.92***	-4.23***
[0,1]	-7.48	-4.08	-3.89***	-4.15***
[-5,-1]	-7.75	-6.32	-3.02***	-3.98***
[1,5]	-5.28	-3.57	-2.55**	-2.63***
[-5,5]	-18.26	-14.51	-4.59***	-4.84***
Min. of 30 sell messages (n = 38)				
[-1,0]	-9.89	-8.86	-2.46**	-2.75***
0	-6.20	-3.33	-2.18**	-2.52**
[0,1]	-6.10	-1.89	-1.85*	-1.37
[-5,-1]	-8.58	-10.78	-2.70**	-2.43**
[1,5]	-4.20	-2.38	-1.35	-1.95*
[-5,5]	-18.98	-13.74	-3.66***	-3.14***
Min. of 40 sell messages (n = 17)				
[-1,0]	-14.86	-14.82	-2.28**	-2.15**
0	-7.75	-6.27	-1.74	-1.92*
[0,1]	-6.89	-3.29	-1.23	-1.16
[-5,-1]	-8.90	-11.26	-1.83*	-1.63
[1,5]	-6.42	-5.31	-1.89*	-1.92*
[-5,5]	-23.07	-20.33	-2.78**	-2.30**

[*t-5, t-1*] are significantly and economically meaningful negative ranging from -6.28%^{min10MSG} to -8.90%^{min40MSG}. Peak average abnormal returns then follow on the event day *t* ranging from -3.65%^{min10MSG} to -7.75%^{min40MSG}. Furthermore, we generally find significant negative CARs in the event window [*t+1, t+5*]. The development of the CARs for negative events suggests that message board users discuss and interpret the negative development of firms. One might argue that message board users especially anticipate the negative momentum of underperforming stocks. The peak of negative abnormal returns on the event day *t* and the absence of return reversals within five days, however, imply that message board users may contribute to price discovery by interpreting and analyzing the firm's situation and allow other users to understand the downward slope of stock price performance further.

In summary, the event study findings support our hypothesis that negative message board sentiment has a substantially detrimental relation to abnormal returns and that bullish board users might act as noise traders in the market, reinforcing stock price volatility.

3.4. Predictability of Investor Sentiment for Abnormal Returns

To examine the intertemporal relationship between message board activity, in particular, the average investor sentiment score or segmented sentiments (positive vs. negative), and abnormal stock performance, we organize our principal analysis around the following regression specifications:

$$AbRet_{i,t0,t+j} = \alpha + \beta_1\ LogMes_{i,t} + \beta_2\ Bullishness_{i,t} + \beta_3\ Agreement_{i,t} + \gamma X + \varepsilon_{i,t} \tag{14}$$

$$\begin{aligned} AbRet_{i,t0,t+j} = {} & \alpha + \beta_1\ LogMes_{i,t} + \beta_2\ PosSentiment_{i,t} + \beta_3\ NegSentiment_{i,t} \\ & + \beta_4\ Agreement_{i,t} + \gamma X + \varepsilon_{i,t} \end{aligned} \tag{15}$$

where $AbRet_{i,t0,t+j}$ denotes the difference of compound raw returns and value-weighted market return from day t to $t+j$ (where j = 0,5,10, and 30 respectively) for firm i. We showed in our previous study, that message board sentiment would be incorporated into stock prices within one month. Thus, we expect a maximum time window of $t+30$ to be sufficient (Leung and Ton, 2015). The general regression specification is based on Chen et al. (2014) but adapted for our research goals. The regression dataset contains 283,585 until 390,842 observations on firm-day level depending on the time window. Our main message board variables are defined as follows or already described in section 3.2: $LogMes_{i,t}$ is the log transformation ($1+M_t$), $Bullishness_{i,t}$ is the standardized bullishness index defined in formula (1), $PosSentiment_{i,t}$ and $NegSentiment_{i,t}$ describe the positive and negative sentiment denoted in formula (2) and (3), $Agreement_{i,t}$ is the agreement index described in formula (4).

The vector X includes the following control variables: $Volatility_{i,t-30,t-1}$ is the 30-day-standard deviation of returns prior to day *t*, $Upgrade_{i,t}$/$Downgrade_{i,t}$ describe the number of analyst upgrade/downgrade recommendation on day *t*, $PosMeanES_{i,t}$/$NegMeanES_{i,t}$ denote dummy variables for positive/negative mean earnings surprise on day *t*. We further include $AbRet_{i,t-1}$, $AbRet_{i,t-2}$, and $AbRet_{i,t-j,t-1}$ to control for possible autocorrelation. Lastly, we include the interaction terms *Bull x* $Agree_{i,t}$ ($Bullishness_{i,t}$ *x* $Agreement_{i,t}$) and *LogMes x* $Vola_{i,t}$ ($LogMes_{i,t}$ *x* $Volatility_{i,t-30,t-1}$). Due to the broad variety of observed firms in our data set, we assume significant cross-sectional differences in message posting volumes as well as firm-characteristics. We, therefore, use firm/year-fixed effects for each stock in our regressions[21]. Additionally, we use clustered standard errors by firm and year to account for the lack of independence in firms' abnormal returns (heteroscedasticity), as well as serial- and cross-correlation. This approach is consistent with the method used by Petersen (2009).

Results of the regression are tabulated in Table 3-5. The analyst-based coefficient estimates of the control variables are generally in line with our expectations. For $AbRet_{i,t0,t+30}$ (column 10) we find positive estimates for $Upgrade_{i,t}$ and $PosMeanES_{i,t.}$ and negative (significant) values for $Downgrade_{i,t}$ and $NegMeanES_{i,t}$. In general, we find that message volume is significantly negative, and the average bullishness/agreement index is significantly positively associated with abnormal returns throughout our observed holding periods. Applying segmented sentiment, the significance of positive sentiment diminishes after 30 trading days, but for negative sentiment, the coefficient estimates remain significant with increasing impact (from $\beta_{t0}^{NegSentiment}$= -0.010 to $\beta_{t0,t+30}^{NegSentiment}$ = -0.033). Therefore, the significance of the average bullishness index is mainly driven by negative sentiment. Similar to Tetlock (2007) and Chen et al. (2014), we find predictive power of negative sentiment shared on internet message boards. We also do find a significant positive correlation between positive sentiment and abnormal returns until the holding period of 10 trading days with $\beta_{t0,t+j}^{PosSentiment}$ ranging from +0.006 to +0.012. The effect, however, diminishes after 30 trading days and speaks for a return reversal and the theory of noise trading at some part.

[21] To test the robustness of the fixed-effect vs. the random-effect model we have conducted the Hausman-specification test on the panel data. Results confirmed the validity of the fixed-effect regression model specification. Results are not tabulated here.

Other implications from our regressions are that $Bullishness_{i,t}$ and $Agreement_{i,t}$ are dependent on each other: the significant coefficient of the interaction term *Bull x Agree* implies the higher the $Agreement_{i,t}$ the higher the impact of $Bullishness_{i,t}$ on abnormal returns. Furthermore, volatility significantly predicts future abnormal returns. Negative realized volatility is significantly negatively related to $AbRet_{t0}$. The significance reverts to a positive relationship for subsequent holding periods. To also examine the connection between realized volatility and the message volume, we include the interaction term $LogMes\ x\ Vola_{i,t}$. At first, the interaction term $LogMes\ x\ Vola_{i,t}$ is slightly positively significant for the contemporaneous regressions, with a coefficient of about -0.062 for all specifications. The relation then reverts into negative and becomes highly significant for a holding period of 30 trading days with coefficients of around 0.320 for all specifications. This suggests that the higher the number of posted messages on day *t* the lower the impact of volatility of the past 30 trading days, $Vola_{i,t-30,t-1}$, on future abnormal returns.

We also repeat regressions for our sample divided in large (Table 3-6) and small capitalization stocks (Table 3-7), since we found previously that small stocks are rather impacted by internet message board activity than large stocks (Leung and Ton, 2015). Again, the results and the coefficient estimates of the internet message board variables especially hold for small capitalization stocks. For large stocks, we only find weak evidence on the contemporaneous correlation between positive or negative sentiment abnormal returns. For robustness and to test the extent to which our results might have been affected by sparseness of message postings by different firms, we conduct the same regressions on a data set with firm-days with at least ten relevant buy or sell messages a day. The overall structure and pattern remained stable and results are tabulated in Table 3-8 to Table 3-10.

In summary, our regression findings that average and segmented sentiment scores must be treated individually suggest that opinions expressed via finance related social media outlets contribute to price discovery for firms experiencing negative abnormal return momentum but also induce positive shocks which may be attributed to the outcome of noise trading. However, the causal explanation if sentiment leads stock performance or vice versa cannot be clearly answered.

Table 3-5. Message Board Variables and Abnormal Returns – All Stocks

Firm- and year-fixed regressions were conducted. $LogMessages_{i,t}$ is the log transformation (1+Mt), $Bullishness_{i,t}$ is the standardized bullishness index defined in formula (1), $PosSentiment_{i,t}$ / $NegSentiment_{i,t}$ is the log transformation (1+MtBuy / MtSell), $Agreement_{i,t}$ is the agreement index described in formula (4), $AbRet_{i,t}$ describes the firm's abnormal return, calculated as the difference of compound raw returns and value-weighted market return, $Vola_{i,t-30,t-1}$ is the 30-day-standard deviation of returns prior to day *t*, $Upgrade_{i,t}$/$Downgrade_{i,t}$ describe the number of analyst upgrade/downgrade recommendation on day *t*, $PosMeanES_{i,t}$/$NegMeanES_{i,t}$ denote dummy variables for positive/negative mean earnings surprise on day *t*. T-statistics computed are based on standard errors clustered by firm and year and are denoted in parentheses. ***, **, and * describe significance at 1%, 5% and 10% level, respectively. All stocks are included in these panel regressions. The constant is not reported.

	$AbRet_{t0}$	**$AbRet_{t0}$**	**$AbRet_{t0}$**	**$AbRet_{t0,t+5}$**	**$AbRet_{t0,t+5}$**	**$AbRet_{t0,t+5}$**	**$AbRet_{t0,t+10}$**	**$AbRet_{t0,t+10}$**	**$AbRet_{t0,t+10}$**	**$AbRet_{t0,t+30}$**	**$AbRet_{t0,t+30}$**	**$AbRet_{t0,t+30}$**
$LogMes_{i,t}$	-0.000		-0.002**	-0.008***		-0.011***	-0.009***		-0.011***	-0.018***		-0.021***
	(0.001)		(0.001)	(0.002)		(0.003)	(0.002)		(0.003)	(0.006)		(0.006)
$Bullishness_{i,t}$	0.008***		0.006***	0.018***		0.015***	0.014***		0.012***	0.017***		0.015***
	(0.001)		(0.001)	(0.002)		(0.002)	(0.002)		(0.002)	(0.005)		(0.005)
$PosSentiment_{i,t}$		0.008***			0.012***			0.006***			0.001	
		(0.000)			(0.001)			(0.001)			(0.003)	
$NegSentiment_{i,t}$		-0.010***			-0.027***			-0.025***			-0.033***	
		(0.001)			(0.004)			(0.004)			(0.010)	
$Agreement_{i,t}$	0.008***	0.004***	0.006***	0.009**	0.000	0.003	0.015***	0.005	0.010**	0.014**	0.002	0.009
	(0.001)	(0.001)	(0.001)	(0.005)	(0.006)	(0.005)	(0.004)	(0.005)	(0.005)	(0.007)	(0.009)	(0.007)
Bull x $Agree_{i,t}$			0.003***			0.008***			0.006***			0.006**
			(0.001)			(0.002)			(0.002)			(0.003)
$Vola_{i,t-30,t-1}$	-0.089***	-0.088***	-0.089***	0.278***	0.279***	0.278***	0.513***	0.514***	0.513***	1.577***	1.577***	1.577***
	(0.011)	(0.011)	(0.011)	(0.072)	(0.072)	(0.072)	(0.096)	(0.096)	(0.096)	(0.174)	(0.174)	(0.174)
LogMes x $Vola_{i,t}$	0.062*	0.063*	0.062*	-0.095	-0.088	-0.092	-0.094	-0.089	-0.093	-0.322***	-0.314**	-0.320***
	(0.033)	(0.033)	(0.033)	(0.077)	(0.078)	(0.078)	(0.090)	(0.090)	(0.090)	(0.123)	(0.123)	(0.123)
$Upgrade_{i,t}$	-0.000	-0.000	-0.000	0.001	0.001	0.001	-0.000	-0.000	-0.000	0.003	0.003	0.003
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
$Downgrade_{i,t}$	0.000	0.000	0.000	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.003**	-0.003**	-0.003**
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$PosMeanES_{i,t}$	-0.003**	-0.003**	-0.003**	-0.000	-0.000	-0.000	-0.002	-0.002	-0.002	0.008	0.008	0.008
	(0.002)	(0.002)	(0.002)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)	(0.007)	(0.007)	(0.007)
$NegMeanES_{i,t}$	-0.002	-0.002	-0.002	-0.009	-0.009	-0.009	-0.006	-0.005	-0.006	-0.003	-0.002	-0.003
	(0.003)	(0.003)	(0.003)	(0.007)	(0.007)	(0.007)	(0.009)	(0.009)	(0.009)	(0.017)	(0.017)	(0.017)
$AbRet_{i,t-1}$	0.083***	0.083***	0.084***	-0.070*	-0.070*	-0.070*	-0.111***	-0.111***	-0.111***	-0.080**	-0.080**	-0.080**
	(0.012)	(0.012)	(0.012)	(0.038)	(0.038)	(0.038)	(0.020)	(0.020)	(0.020)	(0.034)	(0.034)	(0.034)
$AbRet_{i,t-2}$	-0.000	-0.000	0.000	0.087***	0.087***	0.087***	0.023	0.023	0.023	0.041	0.041	0.041
	(0.005)	(0.005)	(0.005)	(0.031)	(0.031)	(0.031)	(0.022)	(0.022)	(0.022)	(0.034)	(0.034)	(0.034)
$AbRet_{i,t-5,t-1}$				-0.105***	-0.105***	-0.105***						
				(0.017)	(0.017)	(0.017)						
$AbRet_{i,t-10,t-1}$							-0.087***	-0.087***	-0.087***			
							(0.011)	(0.011)	(0.011)			
$AbRet_{i,t-30,t-1}$										-0.185***	-0.185***	-0.185***
										(0.019)	(0.019)	(0.019)
Observations	390,842	390,842	390,842	344,523	344,523	344,523	331,655	331,655	331,655	283,585	283,585	283,585
Adjusted R^2	1.7%	1.7%	1.7%	0.7%	0.7%	0.7%	1.3%	1.3%	1.3%	4.6%	4.6%	4.6%

Table 3-6. Message Board Variables and Abnormal Returns – Only Large Stocks

Firm- and year-fixed regressions were conducted. $LogMessages_{i,t}$ is the log transformation (1+Mt), $Bullishness_{i,t}$ is the standardized bullishness index defined in formula (1), $PosSentiment_{i,t}$ / $NegSentiment_{i,t}$ is the log transformation (1+MtBuy / MtSell), $Agreement_{i,t}$ is the agreement index described in formula (4), $AbRet_{i,t}$ describes the firm's abnormal return, calculated as the difference of compound raw returns and value-weighted market return, $Vola_{i,t-30,t-1}$ is the 30-day-standard deviation of returns prior to day *t*, $Upgrade_{i,t}$/$Downgrade_{i,t}$ describe the number of analyst upgrade/downgrade recommendation on day *t*, $PosMeanES_{i,t}$/$NegMeanES_{i,t}$ denote dummy variables for positive/negative mean earnings surprise on day *t*. T-statistics computed are based on standard errors clustered by firm and year and are denoted in parentheses. ***, **, and * describe significance at 1%, 5% and 10% level, respectively. Only large capitalization stocks are included in these panel regressions. The constant is not reported.

	$AbRet_{t0}$	$AbRet_{t0}$	$AbRet_{t0}$	$AbRet_{t0,t+5}$	$AbRet_{t0,t+5}$	$AbRet_{t0,t+5}$	$AbRet_{t0,t+10}$	$AbRet_{t0,t+10}$	$AbRet_{t0,t+10}$	$AbRet_{t0,t+30}$	$AbRet_{t0,t+30}$	$AbRet_{t0,t+30}$
$LogMes_{i,t}$	-0.001**		-0.001**	-0.005		-0.009	-0.009		-0.009*	0.019		0.021
	(0.001)		(0.001)	(0.007)		(0.008)	(0.006)		(0.005)	(0.018)		(0.018)
$Bullishness_{i,t}$	0.002***		0.002***	0.002		-0.002	0.004		0.005	-0.010		-0.007
	(0.000)		(0.000)	(0.003)		(0.004)	(0.004)		(0.005)	(0.011)		(0.012)
$PosSentiment_{i,t}$		0.001			-0.003			-0.003			0.011	
		(0.001)			(0.005)			(0.005)			(0.010)	
$NegSentiment_{i,t}$		-0.003***			-0.011			-0.011			0.036	
		(0.001)			(0.009)			(0.007)			(0.030)	
$Agreement_{i,t}$	0.001	-0.001	0.000	-0.007	-0.014	-0.012	-0.013	-0.016	-0.012	-0.019	0.002	-0.016
	(0.001)	(0.001)	(0.001)	(0.016)	(0.020)	(0.016)	(0.016)	(0.017)	(0.015)	(0.027)	(0.033)	(0.025)
Bull x $Agree_{i,t}$			0.000			0.010*			-0.001			-0.006
			(0.001)			(0.005)			(0.003)			(0.008)
$Vola_{i,t-30,t-1}$	-0.025	-0.025	-0.026	2.020***	2.029***	2.017***	1.797**	1.799**	1.798**	3.054***	3.029***	3.056***
	(0.043)	(0.043)	(0.043)	(0.718)	(0.720)	(0.716)	(0.834)	(0.834)	(0.834)	(0.780)	(0.762)	(0.779)
$Upgrade_{i,t}$	-0.000	-0.000	-0.000	0.001	0.001	0.001	-0.000	-0.000	-0.000	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(0.003)	(0.003)
$Downgrade_{i,t}$	0.001***	0.001***	0.001***	0.000	0.000	0.000	0.001	0.001	0.001	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
$PosMeanES_{i,t}$	-0.003	-0.003	-0.003	-0.006	-0.005	-0.006	-0.020	-0.020	-0.020	-0.023*	-0.026*	-0.023*
	(0.004)	(0.004)	(0.004)	(0.006)	(0.006)	(0.006)	(0.015)	(0.015)	(0.015)	(0.013)	(0.014)	(0.013)
$NegMeanES_{i,t}$	-0.008**	-0.008**	-0.008**	-0.013	-0.012	-0.015	-0.010	-0.011	-0.010	-0.065***	-0.066***	-0.064***
	(0.004)	(0.004)	(0.004)	(0.010)	(0.011)	(0.010)	(0.013)	(0.012)	(0.013)	(0.023)	(0.023)	(0.022)
$AbRet_{i,t-1}$	0.273***	0.273***	0.273***	-0.129	-0.130	-0.131	-0.131	-0.131	-0.131	0.129	0.133	0.129
	(0.073)	(0.073)	(0.073)	(0.177)	(0.177)	(0.177)	(0.193)	(0.192)	(0.193)	(0.089)	(0.089)	(0.089)
$AbRet_{i,t-2}$	-0.010	-0.010	-0.010	0.133	0.134	0.132	-0.206	-0.206	-0.206	-0.035	-0.037	-0.035
	(0.040)	(0.040)	(0.040)	(0.142)	(0.142)	(0.142)	(0.140)	(0.140)	(0.140)	(0.121)	(0.120)	(0.121)
$AbRet_{i,t-5,t-1}$				-0.019	-0.021	-0.020						
				(0.052)	(0.051)	(0.052)						
$AbRet_{i,t-10,t-1}$							-0.063***	-0.064***	-0.063***			
							(0.023)	(0.023)	(0.023)			
$AbRet_{i,t-30,t-1}$										-0.179**	-0.178**	-0.179**
										(0.079)	(0.078)	(0.079)
Observations	28,122	28,122	28,122	26,808	26,808	26,808	26,193	26,193	26,193	24,379	24,379	24,379
Adjusted R^2	7.9%	7.9%	7.9%	0.8%	0.8%	0.8%	2.2%	2.2%	2.2%	6.8%	6.9%	6.8%

Table 3-7. Message Board Variables and Abnormal Returns – Only Small Stocks

Firm- and year-fixed regressions were conducted. $LogMessages_{i,t}$ is the log transformation (1+Mt), $Bullishness_{i,t}$ is the standardized bullishness index defined in formula (1), $PosSentiment_{i,t}$ / $NegSentiment_{i,t}$ is the log transformation (1+MtBuy / MtSell), $Agreement_{i,t}$ is the agreement index described in formula (4), $AbRet_{i,t}$ describes the firm's abnormal return, calculated as the difference of compound raw returns and value-weighted market return, $Vola_{i,t-30,t-1}$ is the 30-day-standard deviation of returns prior to day *t*, $Upgrade_{i,t}$/$Downgrade_{i,t}$ describe the number of analyst upgrade/downgrade recommendation on day *t*, $PosMeanES_{i,t}$/$NegMeanES_{i,t}$ denote dummy variables for positive/negative mean earnings surprise on day *t*. T-statistics computed are based on standard errors clustered by firm and year and are denoted in parentheses. ***, **, and * describe significance at 1%, 5% and 10% level, respectively. Only small-capitalization stocks are included in these panel regressions. The constant is not reported.

	$AbRet_{t0}$	**$AbRet_{t0}$**	**$AbRet_{t0}$**	**$AbRet_{t0,t+5}$**	**$AbRet_{t0,t+5}$**	**$AbRet_{t0,t+5}$**	**$AbRet_{t0,t+10}$**	**$AbRet_{t0,t+10}$**	**$AbRet_{t0,t+10}$**	**$AbRet_{t0,t+30}$**	**$AbRet_{t0,t+30}$**	**$AbRet_{t0,t+30}$**
$LogMes_{i,t}$	-0.000		-0.002*	-0.009***		-0.012***	-0.009***		-0.012***	-0.025***		-0.028***
	(0.001)		(0.001)	(0.003)		(0.003)	(0.003)		(0.004)	(0.006)		(0.006)
$Bullishness_{i,t}$	0.009***		0.008***	0.021***		0.019***	0.015***		0.013***	0.022***		0.019***
	(0.001)		(0.001)	(0.003)		(0.003)	(0.003)		(0.003)	(0.006)		(0.006)
$PosSentiment_{i,t}$		0.009***			0.013***			0.006***			-0.001	
		(0.000)			(0.001)			(0.002)			(0.003)	
$NegSentiment_{i,t}$		-0.012***			-0.030***			-0.026***			-0.042***	
		(0.002)			(0.005)			(0.005)			(0.009)	
$Agreement_{i,t}$	0.009***	0.006***	0.007***	0.012***	0.003	0.007	0.020***	0.010	0.016***	0.021***	0.006	0.016*
	(0.001)	(0.002)	(0.002)	(0.005)	(0.006)	(0.005)	(0.005)	(0.006)	(0.006)	(0.008)	(0.010)	(0.009)
Bull x $Agree_{i,t}$			0.003***			0.007***			0.005**			0.007**
			(0.001)			(0.002)			(0.002)			(0.003)
$Vola_{i,t-30,t-1}$	-0.078***	-0.077***	-0.078***	0.292***	0.292***	0.292***	0.467***	0.468***	0.467***	1.483***	1.482***	1.483***
	(0.012)	(0.012)	(0.012)	(0.078)	(0.078)	(0.078)	(0.108)	(0.108)	(0.108)	(0.200)	(0.200)	(0.200)
LogMes x $Vola_{i,t}$	0.054	0.055	0.054	-0.114	-0.106	-0.111	-0.069	-0.063	-0.068	-0.276**	-0.265**	-0.274**
	(0.037)	(0.037)	(0.037)	(0.085)	(0.087)	(0.086)	(0.090)	(0.089)	(0.090)	(0.121)	(0.122)	(0.121)
$Upgrade_{i,t}$	-0.000	-0.000	-0.000	-0.002	-0.002	-0.002	-0.005	-0.005	-0.005	0.019	0.019	0.019
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.022)	(0.022)	(0.022)
$Downgrade_{i,t}$	-0.000	-0.000	-0.000	-0.002	-0.002	-0.002	-0.004	-0.004	-0.004	-0.007	-0.007	-0.007
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.006)	(0.006)	(0.006)
$PosMeanES_{i,t}$	-0.004	-0.004	-0.004	0.001	0.001	0.001	0.001	0.001	0.001	0.011	0.011	0.011
	(0.002)	(0.002)	(0.002)	(0.006)	(0.006)	(0.006)	(0.008)	(0.008)	(0.008)	(0.012)	(0.012)	(0.012)
$NegMeanES_{i,t}$	-0.000	0.000	0.000	-0.008	-0.007	-0.008	-0.004	-0.003	-0.003	-0.016	-0.015	-0.015
	(0.003)	(0.003)	(0.003)	(0.009)	(0.009)	(0.009)	(0.012)	(0.013)	(0.012)	(0.022)	(0.022)	(0.022)
$AbRet_{i,t-1}$	0.070***	0.070***	0.070***	-0.090**	-0.090**	-0.090**	-0.111***	-0.111***	-0.111***	-0.096***	-0.096***	-0.096***
	(0.012)	(0.013)	(0.012)	(0.037)	(0.037)	(0.037)	(0.022)	(0.022)	(0.022)	(0.036)	(0.036)	(0.036)
$AbRet_{i,t-2}$	0.001	0.001	0.001	0.070**	0.070**	0.070**	0.026	0.026	0.026	0.042	0.042	0.042
	(0.006)	(0.006)	(0.006)	(0.030)	(0.030)	(0.030)	(0.023)	(0.023)	(0.023)	(0.035)	(0.035)	(0.035)
$AbRet_{i,t-5,t-1}$				-0.128***	-0.128***	-0.128***						
				(0.019)	(0.019)	(0.019)						
$AbRet_{i,t-10,t-1}$							-0.084***	-0.084***	-0.084***			
							(0.011)	(0.011)	(0.011)			
$AbRet_{i,t-30,t-1}$										-0.189***	-0.189***	-0.189***
										(0.022)	(0.022)	(0.022)
Observations	284,452	284,452	284,452	246,781	246,781	246,781	238,759	238,759	238,759	208,803	208,803	208,803
Adjusted R^2	1.6%	1.6%	1.6%	1.0%	1.0%	1.0%	1.2%	1.2%	1.2%	4.6%	4.6%	4.6%

Table 3-8. Message Board Variables and Abnormal Returns – All (Min. 10 Mes.)

Firm- and year-fixed regressions were conducted. *$LogMessages_{i,t}$* is the log transformation (1+Mt), *$Bullishness_{i,t}$* is the standardized bullishness index defined in formula (1), *$PosSentiment_{i,t}$* / *$NegSentiment_{i,t}$* is the log transformation (1+MtBuy / MtSell), *$Agreement_{i,t}$* is the agreement index described in formula (4), *$AbRet_{i,t}$* describes the firm's abnormal return, calculated as the difference of compound raw returns and value-weighted market return, *$Vola_{i,t-30,t-1}$* is the 30-day-standard deviation of returns prior to day *t*, *$Upgrade_{i,t}$/$Downgrade_{i,t}$* describe the number of analyst upgrade/downgrade recommendation on day *t*, *$PosMeanES_{i,t}$/$NegMeanES_{i,t}$* denote dummy variables for positive/negative mean earnings surprise on day *t*. T-statistics computed are based on standard errors clustered by firm and year and are denoted in parentheses. ***, **, and * describe significance at 1%, 5% and 10% level, respectively. Only firm-days with a minimum of 10 sell and buy messages were included. The constant is not reported.

	$AbRet_{t0}$	**$AbRet_{t0}$**	**$AbRet_{t0}$**	**$AbRet_{t0,t+5}$**	**$AbRet_{t0,t+5}$**	**$AbRet_{t0,t+5}$**	**$AbRet_{t0,t+10}$**	**$AbRet_{t0,t+10}$**	**$AbRet_{t0,t+10}$**	**$AbRet_{t0,t+30}$**	**$AbRet_{t0,t+30}$**	**$AbRet_{t0,t+30}$**
$LogMes_{i,t}$	0.008***		0.001	-0.009		-0.019***	-0.018***		-0.024***	-0.046*		-0.061**
	(0.003)		(0.003)	(0.006)		(0.007)	(0.007)		(0.008)	(0.025)		(0.024)
$Bullishness_{i,t}$	0.010***		0.009***	0.019***		0.018***	0.016***		0.015***	0.033*		0.031
	(0.002)		(0.002)	(0.004)		(0.004)	(0.005)		(0.005)	(0.019)		(0.020)
$PosSentiment_{i,t}$		0.019***			0.017***			0.006			0.003	
		(0.002)			(0.004)			(0.004)			(0.008)	
$NegSentiment_{i,t}$		-0.019***			-0.037***			-0.034***			-0.064**	
		(0.003)			(0.008)			(0.008)			(0.033)	
$Agreement_{i,t}$	0.027***	0.001	0.001	0.027***	-0.017	-0.011	0.035***	-0.004	0.013	0.015	-0.050	-0.041
	(0.005)	(0.007)	(0.007)	(0.009)	(0.015)	(0.013)	(0.011)	(0.018)	(0.017)	(0.028)	(0.056)	(0.036)
Bull x $Agree_{i,t}$			0.010***			0.015***			0.009*			0.022*
			(0.002)			(0.005)			(0.005)			(0.011)
$Vola_{i,t-30,t-1}$	-0.238***	-0.238***	-0.239***	-0.164	-0.165	-0.164	0.144	0.141	0.143	1.255***	1.249***	1.254***
	(0.058)	(0.058)	(0.058)	(0.114)	(0.114)	(0.113)	(0.176)	(0.176)	(0.176)	(0.368)	(0.368)	(0.368)
LogMes x $Vola_{i,t}$	0.090**	0.093**	0.090**	0.072	0.076	0.074	-0.001	0.000	-0.001	-0.035	-0.032	-0.035
	(0.038)	(0.038)	(0.038)	(0.051)	(0.052)	(0.052)	(0.073)	(0.073)	(0.073)	(0.125)	(0.126)	(0.126)
Upgradei,t	-0.001	-0.002	-0.002*	0.000	0.000	-0.000	-0.001	-0.001	-0.001	-0.014	-0.014	-0.014
	(0.001)	(0.001)	(0.001)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.011)	(0.011)	(0.011)
$Downgrade_{i,t}$	0.001	0.001	0.000	-0.002	-0.002	-0.002	-0.007*	-0.007*	-0.007*	0.006	0.005	0.006
	(0.001)	(0.001)	(0.001)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.008)	(0.008)	(0.008)
$PosMeanES_{i,t}$	-0.013**	-0.013**	-0.013**	-0.004	-0.004	-0.005	0.012	0.012	0.011	0.024	0.024	0.023
	(0.005)	(0.005)	(0.005)	(0.013)	(0.013)	(0.013)	(0.015)	(0.015)	(0.015)	(0.028)	(0.028)	(0.028)
$NegMeanES_{i,t}$	0.002	0.003	0.003	0.039	0.040	0.039	0.011	0.012	0.012	-0.037	-0.033	-0.035
	(0.011)	(0.011)	(0.011)	(0.032)	(0.032)	(0.032)	(0.030)	(0.030)	(0.030)	(0.072)	(0.071)	(0.071)
$AbRet_{i,t-1}$	0.056*	0.056*	0.055*	-0.141***	-0.141***	-0.141***	-0.098**	-0.098**	-0.099**	0.067	0.067	0.066
	(0.031)	(0.031)	(0.031)	(0.047)	(0.047)	(0.047)	(0.044)	(0.044)	(0.044)	(0.085)	(0.085)	(0.085)
$AbRet_{i,t-2}$	-0.017	-0.017	-0.017	0.030	0.030	0.030	-0.020	-0.020	-0.020	0.220*	0.219*	0.219*
	(0.019)	(0.019)	(0.019)	(0.041)	(0.041)	(0.041)	(0.046)	(0.047)	(0.046)	(0.119)	(0.119)	(0.119)
$AbRet_{i,t-5,t-1}$				-0.099***	-0.099***	-0.099***						
				(0.028)	(0.028)	(0.028)						
$AbRet_{i,t-10,t-1}$							-0.108***	-0.108***	-0.108***			
							(0.023)	(0.023)	(0.022)			
$AbRet_{i,t-30,t-1}$										-0.238***	-0.238***	-0.238***
										(0.032)	(0.032)	(0.032)
Observations	35,722	35,722	35,722	31,879	31,879	31,879	30,366	30,366	30,366	24,474	24,474	24,474
Adjusted R^2	2.9%	3.1%	3.1%	1.2%	1.3%	1.3%	2.0%	2.1%	2.0%	5.7%	5.7%	5.7%

Table 3-9. Message Board Variables and Abnormal Returns – Large (Min. 10 Mes.)

Firm- and year-fixed regressions were conducted. $LogMessages_{i,t}$ is the log transformation (1+Mt), $Bullishness_{i,t}$ is the standardized bullishness index defined in formula (1), $PosSentiment_{i,t}$ / $NegSentiment_{i,t}$ is the log transformation (1+MtBuy / MtSell), $Agreement_{i,t}$ is the agreement index described in formula (4), $AbRet_{i,t}$ describes the firm's abnormal return, calculated as the difference of compound raw returns and value-weighted market return, $Vola_{i,t-30,t-1}$ is the 30-day-standard deviation of returns prior to day *t*, $Upgrade_{i,t}$/$Downgrade_{i,t}$ describe the number of analyst upgrade/downgrade recommendation on day *t*, $PosMeanES_{i,t}$/$NegMeanES_{i,t}$ denote dummy variables for positive/negative mean earnings surprise on day *t*. T-statistics computed are based on standard errors clustered by firm and year and are denoted in parentheses. ***, **, and * describe significance at 1%, 5% and 10% level, respectively. Only firm-days with a minimum of 10 sell and buy messages for large capitalization stocks were included. The constant is not reported.

	$AbRet_{t0}$	$AbRet_{t0}$	$AbRet_{t0}$	$AbRet_{t0,t+5}$	$AbRet_{t0,t+5}$	$AbRet_{t0,t+5}$	$AbRet_{t0,t+10}$	$AbRet_{t0,t+10}$	$AbRet_{t0,t+10}$	$AbRet_{t0,t+30}$	$AbRet_{t0,t+30}$	$AbRet_{t0,t+30}$
$LogMes_{i,t}$	0.001		-0.002	-0.001		-0.010	-0.018		-0.037	0.142		0.104
	(0.003)		(0.003)	(0.027)		(0.027)	(0.021)		(0.028)	(0.130)		(0.110)
$Bullishness_{i,t}$	0.002		0.002	-0.008		-0.008	0.013**		0.010	-0.050		-0.053
	(0.002)		(0.002)	(0.019)		(0.019)	(0.006)		(0.007)	(0.039)		(0.042)
$PosSentiment_{i,t}$		0.004			-0.012			0.007			0.049	
		(0.004)			(0.013)			(0.014)			(0.048)	
$NegSentiment_{i,t}$		-0.006**			-0.007			-0.047**			0.178	
		(0.003)			(0.036)			(0.019)			(0.137)	
$Agreement_{i,t}$	0.007	-0.002	-0.009	0.029	0.003	-0.020	-0.026	-0.092*	-0.134	0.059	0.268	-0.128
	(0.007)	(0.007)	(0.008)	(0.043)	(0.072)	(0.050)	(0.031)	(0.051)	(0.090)	(0.049)	(0.182)	(0.136)
Bull x $Agree_{i,t}$			0.007**			0.021			0.048			0.083
			(0.003)			(0.014)			(0.031)			(0.067)
$Vola_{i,t-30,t-1}$	-0.088	-0.089	-0.093	2.153***	2.152***	2.135***	4.682***	4.672***	4.657***	3.991***	3.986***	3.963***
	(0.062)	(0.062)	(0.061)	(0.235)	(0.240)	(0.236)	(0.831)	(0.828)	(0.806)	(0.253)	(0.250)	(0.256)
$Upgrade_{i,t}$	-0.002*	-0.002*	-0.003**	0.001	0.001	-0.000	-0.003	-0.003	-0.004	-0.021	-0.021	-0.023
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.005)	(0.004)	(0.005)	(0.013)	(0.013)	(0.015)
$Downgrade_{i,t}$	0.001	0.001	0.001	-0.000	-0.000	-0.001	-0.010*	-0.010*	-0.010*	0.022***	0.022***	0.022***
	(0.001)	(0.001)	(0.001)	(0.003)	(0.003)	(0.003)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)
$PosMeanES_{i,t}$	-0.013	-0.013	-0.012	-0.056**	-0.057**	-0.055**	-0.057**	-0.058**	-0.057**	-0.020	-0.031	-0.011
	(0.010)	(0.010)	(0.009)	(0.026)	(0.025)	(0.026)	(0.028)	(0.027)	(0.029)	(0.040)	(0.039)	(0.044)
$NegMeanES_{i,t}$	-0.039***	-0.040***	-0.033***	-0.077***	-0.086***	-0.061**	-0.159***	-0.165***	-0.123***	-0.101*	-0.093*	-0.039
	(0.004)	(0.004)	(0.005)	(0.019)	(0.016)	(0.024)	(0.030)	(0.028)	(0.019)	(0.058)	(0.053)	(0.111)
$AbRet_{i,t-1}$	0.567***	0.567***	0.566***	-0.729***	-0.729***	-0.734***	-0.589***	-0.593***	-0.597***	0.188**	0.189**	0.182**
	(0.087)	(0.087)	(0.088)	(0.226)	(0.227)	(0.227)	(0.132)	(0.132)	(0.128)	(0.091)	(0.094)	(0.086)
$AbRet_{i,t-2}$	-0.221***	-0.222***	-0.222***	0.322**	0.325**	0.317**	-0.656***	-0.658***	-0.665***	0.330***	0.326**	0.321***
	(0.065)	(0.065)	(0.064)	(0.135)	(0.134)	(0.132)	(0.102)	(0.100)	(0.102)	(0.117)	(0.125)	(0.112)
$AbRet_{i,t-5,t-1}$				0.073*	0.071*	0.072*						
				(0.040)	(0.039)	(0.040)						
$AbRet_{i,t-10,t-1}$							0.039	0.037	0.038			
							(0.048)	(0.047)	(0.047)			
$AbRet_{i,t-30,t-1}$										-0.189***	-0.186***	-0.189***
										(0.059)	(0.058)	(0.060)
Observations	1,452	1,452	1,452	1,335	1,335	1,335	1,289	1,289	1,289	1,211	1,211	1,211
Adjusted R^2	23.4%	23.5%	23.6%	0.1%	0.1%	0.0%	9.2%	9.3%	9.3%	7.9%	8.2%	8.0%

Table 3-10. Message Board Variables and Abnormal Returns – Small (Min. 10 Mes.)

Firm- and year-fixed regressions were conducted. $LogMessages_{i,t}$ is the log transformation (1+Mt), $Bullishness_{i,t}$ is the standardized bullishness index defined in formula (1), $PosSentiment_{i,t}$ / $NegSentiment_{i,t}$ is the log transformation (1+MtBuy / MtSell), $Agreement_{i,t}$ is the agreement index described in formula (4), $AbRet_{i,t}$ describes the firm's abnormal return, calculated as the difference of compound raw returns and value-weighted market return, $Vola_{i,t-30,t-1}$ is the 30-day-standard deviation of returns prior to day *t*, $Upgrade_{i,t}$/$Downgrade_{i,t}$ describe the number of analyst upgrade/downgrade recommendation on day *t*, $PosMeanES_{i,t}$/$NegMeanES_{i,t}$ denote dummy variables for positive/negative mean earnings surprise on day *t*. T-statistics computed are based on standard errors clustered by firm and year and are denoted in parentheses. ***, **, and * describe significance at 1%, 5% and 10% level, respectively. Only firm-days with a minimum of 10 sell and buy messages for small capitalization stocks were included. The constant is not reported.

	$AbRet_{t0}$	$AbRet_{t0}$	$AbRet_{t0}$	$AbRet_{t0,t+5}$	$AbRet_{t0,t+5}$	$AbRet_{t0,t+5}$	$AbRet_{t0,t+10}$	$AbRet_{t0,t+10}$	$AbRet_{t0,t+10}$	$AbRet_{t0,t+30}$	$AbRet_{t0,t+30}$	$AbRet_{t0,t+30}$
$LogMes_{i,t}$	0.007*		0.001	-0.016**		-0.024***	-0.016**		-0.018**	-0.044**		-0.060***
	(0.004)		(0.004)	(0.007)		(0.008)	(0.008)		(0.008)	(0.018)		(0.022)
$Bullishness_{i,t}$	0.013***		0.011***	0.026***		0.024***	0.015***		0.014***	0.035**		0.031**
	(0.003)		(0.003)	(0.005)		(0.005)	(0.005)		(0.005)	(0.014)		(0.013)
$PosSentiment_{i,t}$		0.021***			0.018***			0.005			0.006	
		(0.002)			(0.005)			(0.005)			(0.010)	
$NegSentiment_{i,t}$		-0.019***			-0.046***			-0.027***			-0.068***	
		(0.005)			(0.010)			(0.009)			(0.024)	
$Agreement_{i,t}$	0.028***	0.006	0.006	0.023**	-0.026	-0.006	0.046***	0.019	0.041**	0.015	-0.056	-0.051
	(0.007)	(0.010)	(0.009)	(0.011)	(0.019)	(0.017)	(0.012)	(0.018)	(0.017)	(0.024)	(0.045)	(0.045)
Bull x $Agree_{i,t}$			0.009***			0.012**			0.002			0.027**
			(0.003)			(0.006)			(0.005)			(0.012)
$Vola_{i,t-30,t-1}$	-0.237***	-0.238***	-0.238***	-0.074	-0.076	-0.074	0.167	0.164	0.167	1.331***	1.327***	1.330***
	(0.075)	(0.074)	(0.075)	(0.114)	(0.114)	(0.114)	(0.165)	(0.165)	(0.165)	(0.419)	(0.418)	(0.418)
LogMes x $Vola_{i,t}$	0.071*	0.074*	0.072*	0.044	0.050	0.046	0.034	0.035	0.034	-0.042	-0.038	-0.041
	(0.042)	(0.042)	(0.042)	(0.056)	(0.057)	(0.056)	(0.052)	(0.053)	(0.052)	(0.126)	(0.127)	(0.127)
$Upgrade_{i,t}$	-0.001	-0.001	-0.001	-0.004	-0.004	-0.004	0.001	0.001	0.001	0.003	0.003	0.003
	(0.003)	(0.003)	(0.003)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.015)	(0.015)	(0.015)
$Downgrade_{i,t}$	0.001	0.001	0.001	0.009	0.009	0.009	0.004	0.004	0.004	0.004	0.004	0.004
	(0.003)	(0.003)	(0.003)	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)	(0.010)	(0.010)	(0.010)
$PosMeanES_{i,t}$	-0.011	-0.011	-0.011	-0.003	-0.004	-0.003	0.030	0.030	0.030	0.035	0.035	0.035
	(0.009)	(0.009)	(0.009)	(0.020)	(0.020)	(0.020)	(0.027)	(0.027)	(0.027)	(0.035)	(0.035)	(0.035)
$NegMeanES_{i,t}$	0.006	0.007	0.006	0.045	0.047	0.045	0.012	0.013	0.012	-0.060	-0.057	-0.058
	(0.017)	(0.017)	(0.017)	(0.045)	(0.046)	(0.046)	(0.040)	(0.041)	(0.041)	(0.097)	(0.096)	(0.096)
$AbRet_{i,t-1}$	0.061*	0.061*	0.061*	-0.093*	-0.092*	-0.093*	-0.051	-0.051	-0.051	0.086	0.087	0.086
	(0.036)	(0.036)	(0.036)	(0.052)	(0.052)	(0.052)	(0.033)	(0.033)	(0.033)	(0.096)	(0.096)	(0.096)
$AbRet_{i,t-2}$	-0.018	-0.018	-0.018	0.066	0.067	0.066	0.015	0.016	0.015	0.189	0.188	0.188
	(0.023)	(0.023)	(0.023)	(0.046)	(0.046)	(0.046)	(0.039)	(0.039)	(0.039)	(0.131)	(0.131)	(0.131)
$AbRet_{i,t-5,t-1}$				-0.154***	-0.154***	-0.154***						
				(0.025)	(0.025)	(0.025)						
$AbRet_{i,t-10,t-1}$							-0.137***	-0.137***	-0.137***			
							(0.018)	(0.018)	(0.018)			
$AbRet_{i,t-30,t-1}$										-0.259***	-0.259***	-0.258***
										(0.037)	(0.038)	(0.037)
Observations	25,135	25,135	25,135	22,486	22,486	22,486	21,719	21,719	21,719	18,884	18,884	18,884
Adjusted R^2	3.0%	3.1%	3.1%	2.0%	2.0%	2.0%	2.5%	2.5%	2.5%	6.6%	6.6%	6.6%

3.4.1. Predictability of Message Board Sentiment and Abnormal Returns

To further investigate the causal relationship between message board sentiment and abnormal returns and to address the endogeneity issues in the data, we apply a Vector Autoregression model (VAR) on the aggregate and the individual firm level (panel VAR[22]).

3.4.1.1. Sentiment and Abnormal Returns at the Aggregate Level

We first conduct a test for the optimal lag length to apply the most adequate lead-lag regression specification. Results are reported in Table 3-11. Three out of five tests indicate that a lag structure of four fits best for our model. Only the Hannan-Quinn information criteria and the Schwarz information criteria imply an optimal lag structure of 3 and 2, respectively. Hence, we construct our VAR model based on four endogenous lags to closer examine the causal relationship between the segmented sentiment and abnormal returns. We consider the following three equations to test the intertemporal interaction of sentiment and abnormal returns:

$$AbRet_t = \alpha_1 + \sum_{j=1}^{L} \beta_{1,j}, AbRet_{t-j} + \sum_{j=1}^{L} \gamma_{1,j}, PosSent_{t-j} + \sum_{j=1}^{L} \delta_{1,j}, NegSent_{t-j} + \varepsilon_{1t} \tag{16}$$

$$PosSent_t = \alpha_2 + \sum_{j=1}^{L} \beta_{2,j}, AbRet_{t-j} + \sum_{j=1}^{L} \gamma_{2,j}, PosSent_{t-j} + \sum_{j=1}^{L} \delta_{2,j}, NegSent_{t-j} + \varepsilon_{2t} \tag{17}$$

$$NegSent_t = \alpha_3 + \sum_{j=1}^{L} \beta_{3,j}, AbRet_{t-j} + \sum_{j=1}^{L} \gamma_{3,j}, PosSent_{t-j} + \sum_{j=1}^{L} \delta_{3,j}, NegSent_{t-j} + \varepsilon_{3t} \tag{18}$$

where $AbRet_t$ is the equally-weighted abnormal return and $PosSent_t/NegSent_t$ the aggregated sentiment level of the 3,050 sample stocks at time t between January 11th, 2008 and May 27th, 2016. We apply the lag exclusion χ^2 Wald-tests on each lag in the VAR to test whether aggregated investor sentiment Granger-cause aggregated abnormal stock returns or vice versa. The first two null hypothesis are therefore $H_{1/2}$: $\gamma/\delta_{1,1} = \gamma/\delta_{1,2} = \ldots = \gamma/\delta_{1,L} = 0$, implying that aggregated positive/negative sentiment does not Granger-cause aggregated future abnormal stock returns. The third and fourth null hypothesis of interest are $H_{3/4}$: $\beta_{2/3,1} = \beta_{2/3,2} = \ldots = \beta_{2/3,L} = 0$, indicating that aggregated abnormal stock returns do not Granger-cause aggregated

[22] Based on the model by Abrigo and Love (2015).

Table 3-11. Lag-Order Selection Statistics for VAR – Aggregate Level

* indicates the lag order selected by each criterion, where LH-Ratio = Likelihood-Ratio, FPE = Final prediction error, AIC = Akaike information criterion, SBIC = Schwarz information criterion, and HQIC = Hannan-Quinn information criterion.

Lag	LH-Ratio	DoF	p-Value	FPE	AIC	HQIC	SBIC
0				0.000	-11.719	-11.707	-11.688
1	791.334	9	0.000	0.000	-13.755	-13.705	-13.630
2	82.153	9	0.000	0.000	-13.923	-13.837	-13.706*
3	49.629	9	0.000	0.000	-14.007	-13.883*	-13.696
4	20.092*	9	0.017	0.000*	-14.012*	-13.852	-13.608

positive or negative sentiment, respectively. Table 3-12 shows the results for our lag 4 VAR specification. For comparison, we additionally show results for the lag 2 VAR. The coefficient estimates for equation (16) on the aggregated positive and negative sentiment variables are only highly significant for the negative sentiment on the previous day ($\gamma_{1,1}$ = -0.063). The p-value of the χ^2-test statistics for H_2 is 0.000, and the hypothesis that aggregated negative sentiment does not Granger-cause aggregated abnormal stock returns must, therefore, be rejected. In line with previous results, we also find a negative relationship in equation (16) for positive sentiment on the previous day and abnormal returns which is line with the return reversal observed in the event study, even though not found significant here. On the aggregate level, we thus find indications that negative sentiment predicts abnormal returns and that aggregated positive sentiment has no Granger-relation to aggregated abnormal returns.

3.4.1.2. Sentiment and Abnormal Returns at the Individual Level

To further examine the individual Granger-relationship between investor sentiment and abnormal returns on the individual level, we perform a panel vector autoregression. Hence, we also test the hypothesis $H_{1\text{-}4}$ on the individual level. Based on the test for the optimal lag length for the panel data, we use the lag of 3 for the panel VAR. For comparison, we also show the results for lag 2 and 4 of the panel VAR in Table 3-13.

Table 3-12. Vector Autoregressions at Aggregate Level – Abnormal Returns

The observations are on an aggregate level. $Abret_t$ is the average difference of value-weighted market and stock return, $PosSentiment_t$ / $NegSentiment_t$ is the log transformation of (1 + M_t^{Buy} / M_t^{Sell}). Z-statistics are reported in parenthesis. χ^2-test statistics are shown for the exclusion of the individual variable for the Granger-causality Wald-test. ***, **, and * describe significance at 1%, 5% and 10%-levels, respectively.

Lags	*2*	*4 (opt.)*	*2*	*4 (opt.)*	*2*	*4 (opt.)*
	$Abret_t$	**$Abret_t$**	**$PosSent_t$**	**$PosSent_t$**	**$NegSent_t$**	**$NegSent_t$**
Intercept	0.009**	0.013*	0.090***	0.040	0.008	0.006
	(2.477)	(1.931)	(5.225)	(1.394)	(0.829)	(0.373)
$Abret_{t-1}$	0.031	-0.067	0.181	0.247	-0.172**	-0.115
	(1.108)	(-1.297)	(1.436)	(1.148)	(-2.483)	(-0.900)
$Abret_{t-2}$	0.022	-0.020	0.028	0.064	0.050	-0.102
	(0.852)	(-0.392)	(0.234)	(0.310)	(0.761)	(-0.828)
$Abret_{t-3}$		0.012		0.062		-0.147
		(0.261)		(0.318)		(-1.256)
$Abret_{t-4}$		-0.014		0.195		-0.104
		(-0.308)		(1.064)		(-0.953)
$PosSent_{t-1}$	-0.005	-0.007	0.602***	0.407***	0.015	-0.019
	(-0.918)	(-0.575)	(22.458)	(7.868)	(0.992)	(-0.627)
$PosSent_{t-2}$	0.006	0.006	0.309***	0.240***	0.009	-0.026
	(1.076)	(0.454)	(11.565)	(4.488)	(0.621)	(-0.818)
$PosSent_{t-3}$		0.009		0.115**		0.048
		(0.716)		(2.262)		(1.572)
$PosSent_{t-4}$		-0.005		0.186***		0.022
		(-0.397)		(3.889)		(0.771)
$NegSent_{t-1}$	-0.063***	-0.065***	0.101**	-0.081	0.518***	0.435***
	(-5.763)	(-3.117)	(2.029)	(-0.943)	(18.897)	(8.433)
$NegSent_{t-2}$	-0.001	-0.008	0.002	0.222**	0.236***	0.052
	(-0.083)	(-0.344)	(0.046)	(2.377)	(8.597)	(0.925)
$NegSent_{t-3}$		-0.013		-0.075		0.209***
		(-0.584)		(-0.804)		(3.719)
$NegSent_{t-4}$		-0.008		0.015		0.059
		(-0.392)		(0.183)		(1.206)
Observations	1,230	380	1,230	380	1,230	380
χ^2-stat *AbRet*			2.224	3.201	6.393	5.177
p-Value *AbRet*			0.329	0.525	0.041**	0.270
χ^2-stat *PSent.*	1.159	1.091				
p-Value *PSent.*	0.560	0.896				
χ^2-stat *NSent.*	60.352	33.191				
p-Value *NSent.*	0.000***	0.000***				

Results on the individual level yield different implications on the Granger-relationships compared to the aggregate level: Positive sentiment significantly predicts abnormal returns at the significance level of 5% with coefficient estimates of -0.002 and +.0.001 for *t-1* and *t-3*, respectively (see Table 3-13 (column 2)).

Table 3-13. Panel Vector Autoregression at Individual Level – Abnormal Returns

The observations are on the individual stock level. *Abret*$_t$ is the difference of value-weighted market and stock return, *PosSentiment*$_t$ / *NegSentiment*$_t$ is the log transformation of (1 + M_t^{Buy} / M_t^{Sell}). Z-statistics are reported in parenthesis. χ^2-test statistics are shown for the exclusion of the individual variable for the Granger-causality Wald-test. ***, **, and * describe significance at 1%, 5% and 10%-levels, respectively.

	AbRet$_t$	**AbRet$_{t,\ opt.}$**	**AbRet$_t$**	**PosSent$_t$**	**PosSent$_{t,opt}$**	**PosSent$_t$**	**NegSent$_t$**	**NegSent$_{t,opt}$**	**NegSent$_t$**
AbRet$_{t-1}$	-0.034***	-0.036***	-0.025	0.172***	0.301***	0.352***	0.020	0.024	-0.035
	(-3.988)	(-2.754)	(-1.526)	(4.846)	(5.630)	(3.595)	(0.804)	(0.622)	(-0.477)
AbRet$_{t-2}$	-0.019***	-0.018*	-0.025	-0.215***	-0.048	-0.066	0.044**	0.068**	-0.036
	(-2.840)	(-1.753)	(-1.186)	(-6.637)	(-0.891)	(-0.659)	(2.086)	(2.027)	(-0.554)
AbRet$_{t-3}$		0.009	0.046***		-0.212***	-0.156		0.009	0.110*
		(0.932)	(3.364)		(-4.199)	(-1.611)		(0.307)	(1.773)
AbRet$_{t-4}$			-0.020*			-0.340***			0.028
			(-1.667)			(-3.669)			(0.541)
PosSent$_{t-1}$	-0.001	-0.002**	-0.003***	0.442***	0.404***	0.367***	-0.007**	-0.001	0.004
	(-1.207)	(-2.294)	(-2.998)	(86.134)	(58.193)	(32.002)	(-2.251)	(-0.295)	(0.528)
PosSent$_{t-2}$	0.001***	0.000	-0.000	0.249***	0.196***	0.167***	-0.019***	-0.010***	-0.003
	(2.912)	(0.788)	(-0.345)	(52.833)	(30.531)	(15.095)	(-7.201)	(-2.720)	(-0.493)
PosSent$_{t-3}$		0.001**	0.000		0.163***	0.125***		-0.008**	-0.011*
		(1.998)	(0.107)		(25.949)	(11.697)		(-2.109)	(-1.954)
PosSent$_{t-4}$			0.001			0.136***			-0.007
			(1.184)			(13.100)			(-1.241)
NegSent$_{t-1}$	-0.003***	-0.001	-0.001	0.031***	0.041***	0.044**	0.340***	0.318***	0.301***
	(-3.414)	(-1.147)	(-0.598)	(4.260)	(4.036)	(2.541)	(49.809)	(33.665)	(19.418)
NegSent$_{t-2}$	-0.001	-0.000	-0.000	-0.018**	-0.013	-0.016	0.170***	0.129***	0.110***
	(-1.042)	(-0.084)	(-0.030)	(-2.553)	(-1.309)	(-0.912)	(26.552)	(14.811)	(7.508)
NegSent$_{t-3}$		-0.001	0.000		-0.014	-0.001		0.107***	0.096***
		(-1.180)	(0.041)		(-1.448)	(-0.047)		(11.977)	(6.301)
NegSent$_{t-4}$			-0.003			-0.028			0.059***
			(-1.566)			(-1.628)			(4.059)
Observations	90,503	42,872	15,039	90,503	42,872	15,039	90,503	42,872	15,039
χ^2-stat *AbRet*				73.106	52.271	29.567	4.735	4.286	3.744
p-Val. *AbRet*				0.000***	0.000***	0.000***	0.094*	0.232	0.442
χ^2-stat *PosSent*	11.991	12.919	12.582						
p-Val. *PosSent*	0.002***	0.005***	0.014**						
χ^2-stat *NegSent*	13.974	2.800	2.736						
p-Val. *NegSent*	0.001***	0.423	0.603						

This result is in line with our event study which suggests that abnormal return reversals occur the day after an event day of abnormal positive sentiment. Additionally, results in Table 3-13 (column 5) show that coefficient estimates ($\beta_{2,t-1}$ = +0.301 and $\beta_{2,t-3}$ = -0.212) for abnormal returns are highly significant at 1%-level, which indicates that abnormal returns predict positive sentiment. The χ^2-test statistics for H_1 of 52.271 (p-value = 0.000) are higher than for H_3 with 12.910 (p-value = 0.005), yet both hypotheses can be rejected at the significant level of 1%. These results imply that positive sentiment and abnormal returns both Granger-cause each other, however with larger impact from abnormal returns on positive sentiment. In other words, message board users rather react to abnormal return shocks, but also provide (noisy) information which is then incorporated into abnormal returns, albeit of smaller economic impact.

For negative sentiment and abnormal returns, however, we do not find a Granger-relationship based on the optimal lag length 3. The Granger-causality Wald-test cannot reject the hypothesis H_2 and H_4. As we look at the results for a lag of 2 (Table 3-13 (column 1)), the coefficient estimate for $\delta_{1,t-1}$ of -0.003 is highly significant, and the χ^2 Wald-test rejects hypothesis H_2 which means that negative sentiment Granger-causes abnormal returns and not vice versa. We expect that this difference results from the data structure and the dominance of bullishness in the data set. Since negative related messages are less present on the HotCopper internet message board, we believe that the lag order of 3 and the smaller data set results in insignificance.

Summarizing the (panel) VAR findings, we first find on an aggregate level that negative sentiment Granger-causes aggregated abnormal returns. This suggests that the aggregated sentiment level of message board users can predict market movements. Secondly, we find that positive sentiment and abnormal returns Granger-cause each other on an individual level, yet with significantly more significant impact from abnormal returns to positive sentiment. Therefore, message board users rather react to market activity but also disseminate information that moves stock prices. The predicted abnormal return reversal after positive messages on a subsequent day also speaks for the noise trading theory by De Long et al. (1990), where stock prices are moved away from fundamentals but then return to the real fundamental value. For negative sentiment, we find on an individual level and based on the optimal lag length of 3 for the whole panel data set that negative sentiment is not Granger-related to abnormal returns. However, as we reduce the lag length to 2, we observe that negative sentiment predicts abnormal returns on the individual stock level. This effect might be induced by the structure of the strongly positively biased data set.

To examine the dynamic interaction between the endogenous variables (*$Abret_t$, $PosSent_t$, and $NegSent_t$*) of the panel VAR process, we apply the impulse response analysis. For the validity of the panel VAR application, we first test on the stability of the panel VAR process. Please refer to Sims (1980), Hamilton and Susmel (1994) and Lütkepohl (2005) for an econometric explanation of the model. As stability implies stationarity of the VAR model, we can find an infinite-order vector moving-average (VMA) representation, which is needed for the interpretation of impulse-response functions. Consider that equations (16) - (18) can be formulated as:

$$Y_{i,t} = C_0 + Y_{i,t-1}A_1 + Y_{i,t-2}A_2 + \cdots + Y_{i,t-L}A_L + \varepsilon_{i,t} \tag{19}$$

where $Y_{i,t}$ is a (*1 x n*) vector of the endogenous variables, A_1, A_2, …, A_L are (*m x n*) coefficient matrices and $\varepsilon_{i,t}$ is a (*1 x n*) vector of error terms. The panel VAR process is stable when the moduli of all eigenvalues of the companion matrix $\bar{A}$ are less than 1.

The companion matrix is defined as:

$$\bar{A} = \begin{bmatrix} A_1 & A_2 & \dots & A_L & A_{L-1} \\ I_n & 0_n & \dots & 0_n & 0_n \\ 0_n & I_n & \dots & 0_n & 0_n \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_n & 0_n & \dots & I_n & 0_n \end{bmatrix} \tag{20}$$

where I_n is the identity (*n x n*) matrix. Our robustness tests show that all moduli of the eigenvalues of $\bar{A}$ are strictly less than 1 and thus account for the stability of our panel VAR process (results are not tabulated here).

Based on the work of Abrigo and Love (2015), we apply the Cholesky impulse-response function, to address the issue that the error terms e_{it} might be contemporaneously correlated. The Cholesky adaption is based on the simple impulse-response function Φ_i, which can be expressed as an infinite vector-moving average with the following VMA specifications[23]:

$$\boldsymbol{\Phi_i} = \begin{cases} \boldsymbol{I_n} \quad , & \boldsymbol{L = 0} \\ \displaystyle\sum_{j=1}^{L} \boldsymbol{\Phi_{t-j} A_j}, & \boldsymbol{L = 1, 2, ..} \end{cases} \tag{21}$$

Figure 3-4 depicts the results of the impulse response function based on equation (19). We focus on the dynamic interaction of positive/negative sentiment and abnormal returns. Abnormal returns show no contemporaneous reaction to negative sentiment shocks, but negative peaks occur after 4 days and successively disappear. Positive sentiment shocks lead to a negative peak of abnormal returns on the following day, also in accordance to previous event study and regressions results in sections 3.3 and 3.4.1. This again indicates a negative market reaction on a subsequent day. However, we do not observe a contemporaneous market reaction. A reason could be that a high number of board messages are posted after the closing hours of the ASX as shown in Figure 3-1.

Abnormal return shocks come with different impact. We observe contemporaneous responses of negative and positive sentiment to abnormal return shocks with gradually decreasing impact, yet with larger response magnitudes for positive sentiment. In line with our expectations, negative abnormal return shocks come with positive responses for negative sentiment whereas positive abnormal return shocks come along with positive sentiment responses. We interpret these findings as follows: Message board users tend to react to abnormal returns shocks. For negative abnormal returns shocks, message board users intensify their research on recent developments and future expectations, contribute and may add valuable information to the price discovery process. When experiencing positive abnormal return shocks, message board users first react with positive postings and then trade regardless of their informational situation. Bloomfield et al. (2009) clearly distinguishes noise traders between “liquidity” traders, who trade due to unexplained liquidity reasons, or “uninformed trades”, who might trade despite

[23] We run our statistical analysis with the panel VAR STATA package by Abrigo and Love (2015)

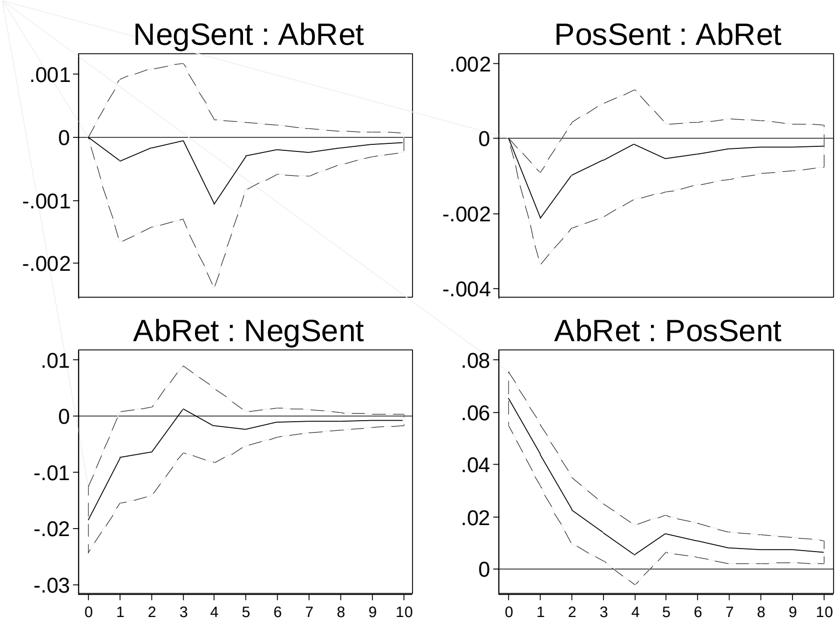

Figure 3-4. Impulse-Response-Functions – Abnormal Returns (Impulse : Response)

having no advantages in information or other exogenous motivational reasons for trade. Liquidity based trading would be the nearest explanation for the negative abnormal return on the day following positive sentiment shocks and thus induce volatility in the market.

In summary, our impulse response function results confirm our prior findings that negative and positive sentiment have differentiated relations to stock market performances. Negative abnormal return responses to negative sentiment shocks show a 4-day delay while positive sentiment shocks lead to a negative abnormal return response on a subsequent day. On the other hand, message board sentiment contemporaneously reacts to abnormal return shocks. Negative abnormal returns follow positive sentiment shocks, implying trading activities of liquidity traders who might induce additional volatility in the market.

3.4.2. Informed Short Selling Against Positive Noise Traders

Previous results in this paper show a contemporaneous positive relationship between positive sentiment and a firm's abnormal returns with subsequent return reversals. This indicates that trades were dominated by sentimental traders who show a propensity to either speculation or over-optimism (Baker and Wurgler, 2007). Prior literature argued that misevaluation of asset prices could only be partially offset by contrarian arbitrageurs or in specific cases (un-)informed short sellers. The high costs and risks associated with betting against sentimental investors (Shleifer and Vishny, 1997) could lead to the conclusion that, for example, only well informed short sellers would bet against overpriced stock movements, which are driven by sentimental investors (Diamond and Verrecchia, 1987). Therefore, we believe that stocks which are hyped on internet message boards and are also targeted by informed short sellers are less prone to experience a positive abnormal return shock with following return reversals. Hence, we conduct the same regressions as in section 3.4, based on equation (15) and furthermore include the variable $PercShort_{i,t}$ and the interaction terms $PercShort\ x\ PosSentiment_{i,t}$ and $PercShort\ x\ NegSentiment_{i,t}$. $PercShort_{i,t}$ describes the ratio between the number of reported short positions and the number of shares outstanding on stock *i* and day *t*. The results are tabulated in Table 3-14.

In line with the prior literature, we find that the share of short selling positions negatively and significantly predicts abnormal returns (e.g., Figlewski and Webb, 1993; Aitken et al., 1998). The contemporaneous relationship between $PercShort_{i,t}$ and a firm's abnormal return is at first slightly positive but then reverts into negative for the period of 30 days. One reason could be that informed sellers preferentially target overvalued stocks. Since the ASIC publishes the total short positions for financial products only four days after reporting[24], one should expect a time-lag of the negative impact of short selling positions on a stock's excess returns. Due to the concern that our results might be affected of trading days with only little message board activity, we conduct the same regressions only including observations with a minimum of 10 and 20 messages on day *t*. The direction of our results remains robust even though the results are less or not significant anymore.

[24] Please see http://asic.gov.au/regulatory-resources/markets/short-selling/short-selling-reporting-short-position-reporting/ as of September 17th, 2017.

Table 3-14. Short Selling and Sentiment Regressions

PosSentiment$_{i,t}$ / *NegSentiment*$_{i,t}$ is the log transformation (1+MtBuy / MtSell), *Agreement*$_{i,t}$ is the agreement index described in formula (4), *AbRet*$_{i,t}$ describes the firm's abnormal return, calculated as the difference of compound raw returns and value-weighted market return, *PercShort*$_{i,t}$ denotes the share of reported short positions of total shares outstanding, *Other Controls* include all other control variables and interaction terms of former regressions. T-statistics computed are based on standard errors clustered by firm and year and are denoted in parentheses. ***, **, and * describe significance at 1%, 5% and 10% level, respectively.

	Panel A: All observations				*Panel B: Min. 20 messages on day t*			
	AbRet$_{t0}$	**AbRet$_{t0,t+5}$**	**AbRet$_{t0,t+10}$**	**AbRet$_{t0,t+30}$**	**AbRet$_{t0}$**	**AbRet$_{t0,t+5}$**	**AbRet$_{t0,t+10}$**	**AbRet$_{t0,t+30}$**
PosSentiment$_{i,t}$	0.008***	0.011***	0.004**	-0.002	0.037***	0.042***	0.014	0.027*
	(0.000)	(0.001)	(0.002)	(0.004)	(0.005)	(0.012)	(0.011)	(0.015)
NegSentiment$_{i,t}$	-0.012***	-0.029***	-0.024***	-0.034***	-0.023***	-0.041**	-0.046**	-0.060**
	(0.001)	(0.005)	(0.005)	(0.011)	(0.008)	(0.020)	(0.020)	(0.030)
AgreeInd$_{i,t}$	0.003**	-0.006	0.004	0.003	0.010	-0.020	-0.036	-0.095
	(0.001)	(0.007)	(0.007)	(0.010)	(0.020)	(0.043)	(0.053)	(0.074)
PercShort$_{i,t}$	0.001***	-0.000	-0.003***	-0.008***	0.023***	0.028**	-0.019	-0.010
	(0.000)	(0.001)	(0.001)	(0.002)	(0.005)	(0.012)	(0.028)	(0.022)
PosSent x PercShort$_{i,t}$	-0.001***	-0.000	0.001***	0.003***	-0.007***	-0.010***	0.003	0.001
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.003)	(0.007)	(0.005)
NegSent x PercShort$_{i,t}$	0.001***	-0.000	0.000	0.001	0.002***	0.000	-0.001	0.000
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
Other Controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Firm FE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Firm-Year-clustered SE	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Observations	329,308	305,317	299,276	258,223	9,894	9,089	8,748	6,722
Adjusted R-squared	1.7%	0.6%	1.3%	4.5%	5.4%	2.4%	3.9%	3.0%
F-value	73.07	20.09	15.36	15.76	19.21	10.02	6.724	4.272

Additionally, we find a significantly negative relationship between positive sentiment expressed on internet message boards and short selling positions on stock *i* on day *t* (negative interaction term *PercShort x PosSentiment*$_{i,t}$). The magnitude of the coefficient increases as we conduct our regressions with a minimum level of message board activity on stock i on day t. This finding implies that a higher ratio of a firm's short position reduces a possible overreaction of a stock's abnormal return on positive sentiment expressed on internet message boards. Therefore, we find empirical evidence that short selling reduces the impact of (positive) sentimental investors on the same day. From the economic point of view, it seems unlikely that our dependent variable, the abnormal return, causes short selling activity. However, we finally cannot eliminate the possibility of confounding events which motivate a short seller to build up that position.

3.4.3. Volatility and Message Board Activity

We have argued in former sections that the activity of noise traders, be it due to liquidity or other exogenous reasons, induce volatility in the market. To closer understand the drivers behind volatility, we first regress volatility against the message board variables including the market return as a control variable for different time periods following Antweiler and Frank (2004). Results are tabulated in Table 3-15. We find that all three message board variables are significantly related to volatility. The message volume reveals significant coefficient estimates of +0.033 and +0.022 for the period *t* and *t+1* at the significance level of 1%.

The bullishness index, in general, has a negative impact on volatility with an also highly significant coefficient estimate of about -0.005 on day *t*. *Agreement* seems to be important in the time window of *t+1* to *t+30* with a coefficient of +0.005. It appears that the message volume has the most substantial impact on market volatility. To further examine the causal relationship between message board activity and market volatility, we conduct a VAR analysis in the next section.

Table 3-15. Regressions on Volatility

The observations are on a firm-day level. $LogMessages_{i,t}$ is the log transformation ($1+M_t$), $Bullishness_{i,t}$ is the standardized bullishness index defined in formula (10), Agreement is the agreement index described in formula (13), $MarketRet_{i,t}$ describes the All Ordinaries market return, $Vola_{i,t}$ and $Vola_{i,t+1}$ are the intraday price volatility, $Vola_{i,t+1,t+5/10/30}$ is the standard deviation of return in the respective time window. Robust standard errors are denoted in parenthesis. ***, **, and * describe significance at 1%, 5% and 10%-levels, respectively.

	$Vola_{i,t}$	$Vola_{i,t+1}$	$Vola_{i,t+1,t+5}$	$Vola_{i,t+1,t+10}$	$Vola_{i,t+1,t+30}$
$LogMessages_{i,t}$	0.033***	0.022***	0.009***	0.008***	0.006***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
$Bullishness_{i,t}$	-0.005***	-0.005***	-0.004***	-0.004***	-0.005***
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
$Agreement_{i,t}$	-0.000	0.003	0.004*	0.004*	0.005***
	(0.004)	(0.003)	(0.002)	(0.002)	(0.002)
$MarketRet_t$	-0.070**	-0.141***	-0.063	-0.087**	-0.047*
	(0.034)	(0.043)	(0.044)	(0.041)	(0.028)
Constant	0.032***	0.040***	0.037***	0.043***	0.046***
	(0.004)	(0.003)	(0.003)	(0.003)	(0.002)
Observations	671,029	670,304	822,288	853,854	860,192
Adjusted R^2	3.91%	1.73%	0.41%	0.34%	0.35%

3.4.4. Predictability of Message Board Activity for Market Volatility

3.4.4.1. Message Volume and Volatility at the Aggregate Level

To examine whether message board activity forecast next-periods stock price volatility and to assess how these two variables interact intertemporally (short-term), we consider the following two equations:

$$Vola_t = \alpha_1 + \sum_{j=1}^{L} \beta_{1,j}, Vola_{t-j} + \sum_{j=1}^{L} \gamma_{1,j}, LogMes_{t-j} + \varepsilon_{1t} \quad (22)$$

$$LogMes_t = \alpha_1 + \sum_{j=1}^{L} \beta_{2,j}, Vola_{t-j} + \sum_{j=1}^{L} \gamma_{2,j}, LogMes_{t-j} + \varepsilon_{1t} \quad (23)$$

where $LogMes_t$ is the equally-weighted message board activity at time *t* and $Vola_t$ is the equally-weighted stock price volatility of the 3,050 sample stocks at time *t*. Based on the test on the optimal lag length, we apply a lag of 3 (results reported in Table 3-16).

Results for the VAR model on the aggregated level are shown in Table 3-17. The null-hypothesis (H_5) that: $\gamma_{1,1} = \gamma_{1,2} = \gamma_{1,3} = 0$ from equation (22) cannot be fully rejected with a p-value of the χ^2-test statistic for H_5 of 0.395. However, the p-value of the χ^2-test statistic for H_6 is 0.093. The null-hypothesis (H_6) that: $\beta_{2,1} = \beta_{2,2} = \beta_{2,3} = 0$ from equation (23) can, therefore, be rejected at the 10%-signifiance level. In another words, the Granger-causality tests indicate that message board activity on the aggregate level may be positively Granger-caused by prior stock price volatility.

Table 3-16. Lag-Order Selection Statistics for VAR – Aggregate Level

* indicates the lag order selected by each criterion, where FPE = Final prediction error, AIC = Akaike information criterion, SBIC = Schwarz information criterion, and HQIC = Hannan-Quinn information criterion.

Lag	Likelihood-Ratio	DoF	p-Value	FPE	AIC	HQIC	SBIC
0	0.0000			0.0000	-6.5294	-6.5211	-6.5086
1	924.8699	4	0.0000	0.0000	-8.9422	-8.9175	-8.8800
2	68.6352	4	0.0000	0.0000	-9.1017	-9.0606	-8.9981
3	26.4214*	4	0.0000	0.0000	-9.1502	-9.0926*	-9.0051*
4	8.4347	4	0.0769	0.0000*	-9.1514*	-9.0773	-8.9647

Table 3-17. Vector Autoregression at Aggregate Level – Volatility

The observations are on an aggregate level. *Vola$_{tt}$* is the scaled difference between the lowest and highest stock price and *LogMes$_t$* is the log transformation (1+M$_t$). Z-statistics are reported in parenthesis. χ^2-test statistics are shown for the exclusion of the individual variable for the Granger-causality Wald-test. ***, **, and * describe significance at 1%, 5% and 10%-levels, respectively.

Explanatory variable	$\mathrm{Vola}_t = \alpha_1 + \sum_{j=1}^{L} \beta_{1,j}, \mathrm{Vola}_{t-j} + \sum_{j=1}^{L} \gamma_{1,j}, \mathrm{LogMes}_{t-j} + \varepsilon_{1t}$			$\mathrm{LogMes}_t = \alpha_1 + \sum_{j=1}^{L} \beta_{2,j}, \mathrm{Vola}_{t-j} + \sum_{j=1}^{L} \gamma_{2,j}, \mathrm{LogMes}_{t-j} + \varepsilon_{1t}$		
Lags	*2*	*3 (opt)*	*4*	*2*	*3 (opt.)*	*4*
	Vola$_t$	**Vola$_t$**	**Vola$_t$**	**LogMes$_t$**	**LogMes$_t$**	**LogMes$_t$**
Intercept	0.014***	0.011***	0.008*	0.133***	0.118***	0.082*
	(6.308)	(3.995)	(1.812)	(5.869)	(4.240)	(1.957)
LogMes$_{t-1}$	0.002	0.002	-0.005	0.661***	0.590***	0.521***
	(0.844)	(0.599)	(-0.801)	(24.693)	(16.500)	(8.952)
LogMes$_{t-2}$	-0.003	-0.003	-0.000	0.267***	0.180***	0.119**
	(-1.119)	(-0.867)	(-0.022)	(10.035)	(4.732)	(2.003)
LogMes$_{t-3}$		-0.001	-0.002		0.168***	0.208***
		(-0.348)	(-0.352)		(5.187)	(3.798)
LogMes$_{t-4}$			0.004			0.097**
			(0.717)			(2.082)
Vola$_{t-1}$	0.482***	0.454***	0.561***	-0.809***	-0.855**	-0.006
	(19.136)	(11.704)	(9.076)	(-3.131)	(-2.265)	(-0.010)
Vola$_{t-2}$	0.312***	0.287***	0.335***	0.524**	0.555	0.386
	(12.642)	(8.224)	(5.144)	(2.072)	(1.636)	(0.641)
Vola$_{t-3}$		0.135***	0.027		-0.053	0.403
		(4.293)	(0.532)		(-0.174)	(0.858)
Vola$_{t-4}$			0.050			-0.876*
			(0.953)			(-1.817)
Observations	1,230	797	380	1,230	797	380
χ^2-stat	1.361	2.978	3.271	9.823	6.416	3.863
p-Value	0.506	0.395	0.513	0.007	0.093	0.425

3.4.4.2. Message Volume and Volatility at the Individual Level

To further examine the individual Granger-relationship between message board activity (volume) and stock price volatility on the individual level, we perform a panel VAR following an impulse response analysis similar to the previous section. We first test the hypothesis H_5 and H_6 on the individual level. Based on the test for optimal lag length for the panel data, we use the lag of 3 for the panel VAR. For comparison, we also show the results for lag 2 and 4 of the panel VAR in Table 3-18.

We find for the optimal lag length of 3, that the previous day message board volume significantly predicts volatility, however, with an economically small impact (coefficient estimate of +0.004). On the other hand, we also observe that past days volatility strongly predicts message board activity even though with changing signs ($\beta_{2,3} = -0.124, \beta_{2,2} = -0.249\ and\ \beta_{2,1} = +0.483$). Both χ^2-test statistics for H_5 of 51.632 (p-value = 0.000) and for H_6 of 66.607 (p-value = 0.000) are highly significant so that both hypotheses can be rejected. In other words, message board volume and stock price volatility Granger-cause each other. Nevertheless, we can conclude the reaction of message board volume to stock market volatility is significantly higher than vice versa.

To also examine the dynamic interaction of message board volume and stock price volatility, we again apply the Cholesky based impulse function. Figure 3-5 shows the corresponding results. Stock price volatility reacts to message board volume shocks on day *t+1* with decreasing but remaining impact after ten days. Setting a one-standard-deviation volatility shock, we observe a considerably high contemporaneous message board activity response compared to the other direction. Our results suggest, that message board activity follows market volatility even though message board activity might induce stock price volatility, albeit of small economic impact.

Table 3-18. Panel Vector Autoregressions at Individual Level – Volatility

The observations are on the individual stock level. *$Vola_t$* is the scaled difference between the lowest and highest stock price and *$LogMes_t$* is the log transformation ($1+M_t$). Z-statistics are reported in parenthesis. χ^2-test statistics are shown for the exclusion of the individual variable for the Granger-causality Wald-test. ***, **, and * describe significance at 1%, 5% and 10%-levels, respectively.

Explanatory variable	$\text{Vola}_t = \alpha_{i,1} + \sum_{j=1}^{L} \beta_{1,j}, \text{Vola}_{i,t-j} + \sum_{j=1}^{L} \gamma_{1,j}, \text{LogMes}_{i,t-j} + \varepsilon_{1t}$			$\text{LogMes}_t = \alpha_{i,1} + \sum_{j=1}^{L} \beta_{2,j}, \text{Vola}_{i,t-j} + \sum_{j=1}^{L} \gamma_{2,j}, \text{LogMes}_{o,t-j} + \varepsilon_{1t}$		
Lags	*2*	*3 (opt.)*	*4*	*2*	*3 (opt.)*	*4*
	$Vola_t$	**$Vola_t$**	**$Vola_t$**	**$LogMes_t$**	**$LogMes_t$**	**$LogMes_t$**
$LogMes_{t-1}$	0.004***	0.004***	0.004***	0.430***	0.398***	0.369***
	(9.176)	(6.809)	(4.373)	(80.483)	(56.013)	(32.134)
$LogMes_{t-2}$	-0.000	0.001	0.001	0.244***	0.192***	0.163***
	(-0.134)	(1.175)	(1.557)	(48.998)	(28.975)	(14.461)
$LogMes_{t-3}$		-0.001	0.000		0.163***	0.128***
		(-1.455)	(0.528)		(24.717)	(11.671)
$LogMes_{t-4}$			0.001			0.124***
			(0.895)			(11.389)
$Vola_{t-1}$	0.237***	0.239***	0.232***	0.426***	0.483***	0.361***
	(25.306)	(16.604)	(11.132)	(9.479)	(6.851)	(2.807)
$Vola_{t-2}$	0.097***	0.075***	0.058***	-0.216***	-0.249***	-0.251*
	(12.020)	(6.131)	(2.798)	(-5.617)	(-3.858)	(-1.958)
$Vola_{t-3}$		0.088***	0.081***		-0.124**	-0.223*
		(7.923)	(4.690)		(-2.078)	(-1.956)
$Vola_{t-4}$			0.070***			0.004
			(3.591)			(0.040)
Observations	90,503	42,872	15,039	90,503	42,872	15,039
χ^2-stat	86.244	51.632	22.645	115.499	66.607	15.218
p-Value	0.000	0.000	0.000	0.000	0.000	0.004

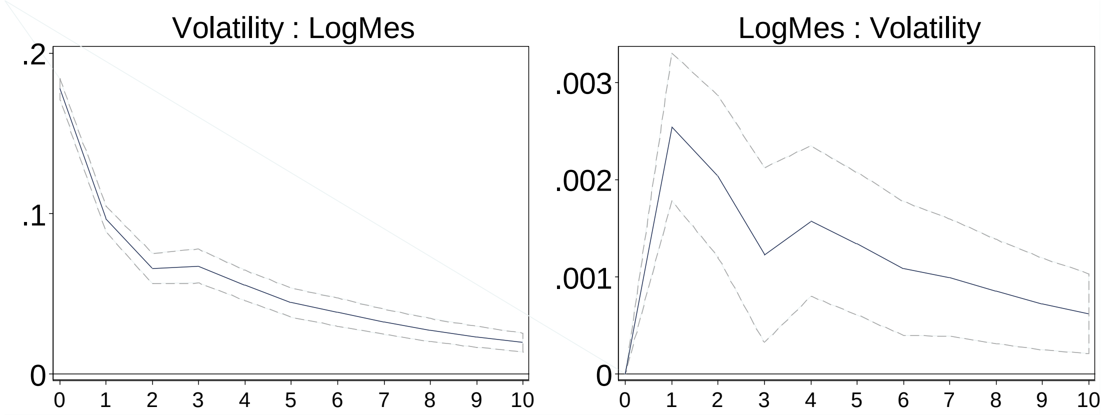

Figure 3-5. Impulse-Response-Functions – Volatility (Impulse : Response)

3.5. Fundamental Information in Message Boards around Company Events

The differentiated impact of social media activity found in our event study (Figure 3-2 and Figure 3-3) and in our multivariate analysis underlines its importance in capital markets. Despite our different steps taken (event study, panel VAR, multivariate regressions), we cannot clearly argue in general whether social media users act as noise traders, who move prices away from their fundamentals, or convey financially relevant information and thus contribute in price discovery. Consequently, researchers must distinguish between the impact of social media in non-event and event specific environments. Thus, we now examine the cross-sectional relationships between the message board variables and fundamental values around annual earnings announcements.

Financial analysts act as essential intermediaries in financial markets and are subject to a broad body of research streams. Two main reasons of existence come along with their role: the discovery of private information and furthermore the interpretation of publicly available information (e.g., Ivković and Jegadeesh, 2004; Asquith et al., 2005; Chen et al., 2010).

Chen et al. (2014) argued that annual earnings reported by firms are probably not affected by social media activity. Since it would also be unlikely that financial analysts revise their recommendations based on negative sentiment (therefore negative sentiment would predict negative earnings surprise), social media would represent an information channel with predictive power. They find that negative opinions revealed on the investment-related platform SeekingAlpha predict future negative earnings surprises. One of the main disadvantages of looking at analyst forecasts is the sole reflection of analysts opinions, rather than the consideration of market information, which could be available to other well-informed market participants (Akbas, 2016). Attributable to the area of Behavioral Finance, opinions might be subject to a positive bias as financial analysts encounter the desire to conform, or in other words "herd" (Olsen, 1996). Herding characteristically moves the mean Earnings per Share (EPS) forecast towards a specific direction and lowers the forecast dispersion. Former studies showed that analysts forecasts have been overly optimistic compared to the actual reported EPS (e.g., Olsen, 1996). A reinforcing factor is also that financial analysts are judged by their degree of conformity with other analyst forecasts since the quality of predictions is exposed to uncontrollable exogenous factors. A consensus forecast is, therefore, in the interest of all analysts to protect their right for existence and thus their human capital (e.g., Scharfstein and Stein, 1990; Trueman, 1994; Froot et al., 1992; Olsen, 1996). Hence, we argue that if financial analysts release optimistic consensus recommendations and social media users agree in the optimistic outlook of the firm's performance, then earnings surprises might be positive.

Following the work of Chen et al. (2014) and Akbas (2016), we conduct a firm/year-fixed regression of annual earnings surprises on message board variables and various control variables to examine the value-content of internet message boards. Our model extends the approach of Chen et al. (2014) by additional consideration of positive sentiment and the degree of agreement in message board discussions. If message board activity would not contain value-relevant information, then no relationship should exist between earnings surprises and our message board variables. However, our results suggest that social media does provide financially relevant information in event-specific environments. For comparison, we construct two different types of earnings surprises as our dependent variable. The standardized unexpected earnings surprise based on analyst forecasts (*SUEAF*) and the standardized unexpected earnings based on the historical time series information (*SUEHIST*).

The standardized unexpected earnings (*SUEAF*) based analyst forecasts is defined as:

$$SUEAF_{i,t} = \frac{(X_{i,t} - Xmed_{i,t-90d})}{P_{i,t}} \tag{24}$$

where X_{it} is the primary Earnings per Share (EPS) before significant items for firm *i* in financial year *t* and $Xmed_{i,t\text{-}90d}$ is the EPS-median of most recent analyst forecasts over 90 days prior to the annual earnings announcement, and $P_{i,t}$ is the price per share for firm *i* at the end of the financial year *t* from I/B/E/S. To eliminate the impact of outliers, we winsorized the top and bottom 1% of the observations.

The standardized unexpected earnings based on the random walk model (*SUEHIST*) is defined as follows:

$$SUE_{it} = \frac{(X_{i,t} - X_{i,t-1})}{P_{i,t}} \qquad (25)$$

where $X_{i,t-1}$ is the primary Earnings per Share before significant items for firm *i* in the previous financial year. For our control variables and following Akbas (2016), we first include $Ret_{50,}$ the compounded return over the period of [-61, -12] days prior to the earnings announcement date and Ret_5 for the five-day return period [-6, -2] prior to the earnings announcement date. We also include $Volatility_{i,10}$ which is the standard deviation of daily returns in the time window [-11, -2] prior to the earnings announcement. Next, we include the log-transformed average turnover $LogTurnover_{i,50}$ over the time window [-61, -12] to account for potential average volume effects as stated by Berkman et al. (2009). Additionally, we add the log-transformed market capitalization $LogSize_{i,t}$, which is the log-transformation of shares outstanding times the share price at the end of the financial year and also accounts for skewness in the data set (small capitalization stocks are predominant in the data set as described in Section 3.2). Lastly, we include cumulated message board variables $LogMessages_{i,[RP]}$, $Bullishness_{i,[RP]}$, *PosSentiment/NegSentiment* $_{i,[RP]}$ and *Agreement* $_{i,[RP]}$ with the reference periods (RP) of [-2, -1], [-7, -1], and [-30 ,-1] to measure the information content over a sufficient period of time.

Table 3-19 reports the summary statistics for the message board period of [-7, -1] days before the earnings announcement based on analyst forecasts. We find a mean of -0.014 for scaled earnings surprise (SUEAF) which supports the argument that analysts tend to herd and are too optimistic in their consensus forecast.

Table 3-19. Summary Statistics

This table reports the summary statistics for the main regression surrounding annual earnings announcements. $SUEAF_{i,t}$ is the difference in actual EPS and forecasted EPS using analyst forecasts 90 days prior to the earnings announcement date scaled by the stock price of the end of the year, $LogMessages_{i,7}$ is the log transformation of (1 + M_t) for the event window [-7,-1], $Bullishness_{i,7}$ is the cumulated bullishness index using formula (12), $Agreement_{i,7}$ is the cumulated agreement index using formula (13), $Return_{i,50}$ is the compounded return over the period of [-61,-12] and $Return_{i,5}$ for the five-day return period [-6,-2] prior to the earnings announcement date. $LogTurnover_{i,50}$ is the log-transformed average turnover over the time window [-61,-12] prior to the earnings announcement date, $Volatility_{i,10}$ is the standard deviation of daily returns in the time window [-11,-2] and $LogSize_{i,t}$ is the log-transformation of the market capitalization at the end of the financial year.

VARIABLES	N	Mean	Median	Std. Dev.	10th Pctl.	25th Pctl.	75th Pctl.	90th Pctl.
$SUEAF_{i,t}$	479	-0.014	-0.001	0.186	-0.090	-0.013	0.009	0.052
$LogMessages_{i,7}$	479	2.147	2.079	1.038	0.693	1.386	2.944	3.555
$Bullishness_{i,7}$	479	1.784	1.791	1.196	0.649	1.075	2.565	3.359
$Agreement_{i,7}$	479	0.817	1.000	0.338	0.169	1.000	1.000	1.000
$Return_{i,50}$	479	-0.027	-0.020	0.352	-0.362	-0.174	0.120	0.304
$Return_{i,5}$	479	-0.003	0.000	0.083	-0.099	-0.042	0.040	0.092
$LogTurnover_{i,50}$	479	13.840	13.510	2.081	11.330	12.260	15.360	16.810
$Volatility_{i,10}$	479	0.029	0.025	0.017	0.012	0.017	0.036	0.051
$LogSize_{i,t}$	479	24.430	24.110	1.721	22.530	23.310	25.430	26.970

In past literature, researchers link (excess) trading volume to divergence in investor opinion (e.g., Beaver, 1968; Bamber, 1987; Kandel and Pearson, 1995; Garfinkel and Sokobin, 2006). As we hypothesis that opinion convergence would, on the other hand, contribute to price discovery, we can directly refer to the sentiment expressed in the internet message board instead of using trading volume as a proxy. Figure 3-6 shows the development of the cumulated agreement index before the earnings announcement date *t*. For example, for the event window of 7 days prior the announcement date [-7, -1], we cumulated all financially relevant board messages (sell and buy recommendations) and constructed the agreement index based on formula (13). We find a convergence pattern as we approach the earnings announcement date with event windows of [-60, -1], [-30, -1], [-7, -1] and [-2, -1]. As DeMarzo et al. (2003) pointed out, the main prerequisites for the convergence of beliefs are that investors may not be isolated from each other and that their beliefs are not fixed in a sense that discussions would stop. Social media platforms enable retail investors to participate in discussions rather than isolating its users in distinctive discussions. Hence, social media generally meet the first requirement for

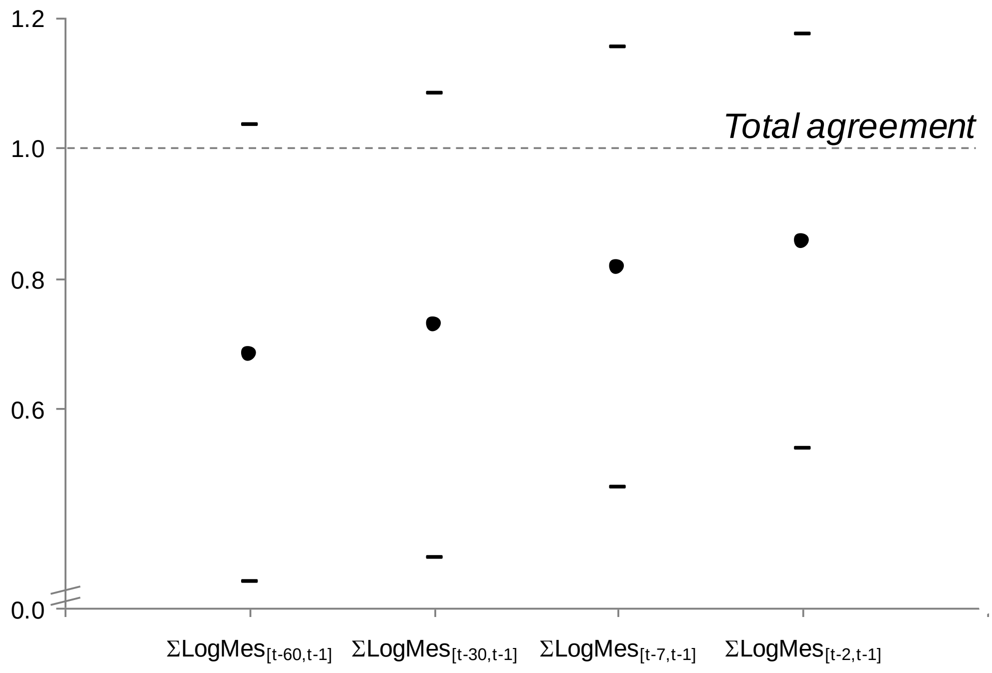

Figure 3-6. Average Cumulated Agreement Index Relative to the Event Date t

belief convergence. However, it is not clear how message board users with fixed beliefs interact in their discussions. We take a closer look at the cross-sectional impact of sentiment on agreement on earnings surprises in the next section.

3.5.1. Portfolio Analysis

Akbas (2016) argues that extraordinary low trading volume contains unfavorable information about a firm's fundamentals since informed investors would not trade – given short selling constraints – based on the bad information they have. We believe that the direct measure of agreement combined with the underlying sentiment would also act as a signal of bad news of a change in firm's fundamentals, equivalent to the abnormal low trading volume found by Akbas (2016). Hence, we construct a portfolio for each year end and assigned the stocks to quartiles based on the combined sentiment and agreement score (*Agreement index x Bullishness index*). Figure 3-7 depicts the average earnings surprise based on analyst forecasts (SUEAF) for each quartile. The mean *SUEAF* for quartile 4 is negative at -2.6% and significant at the

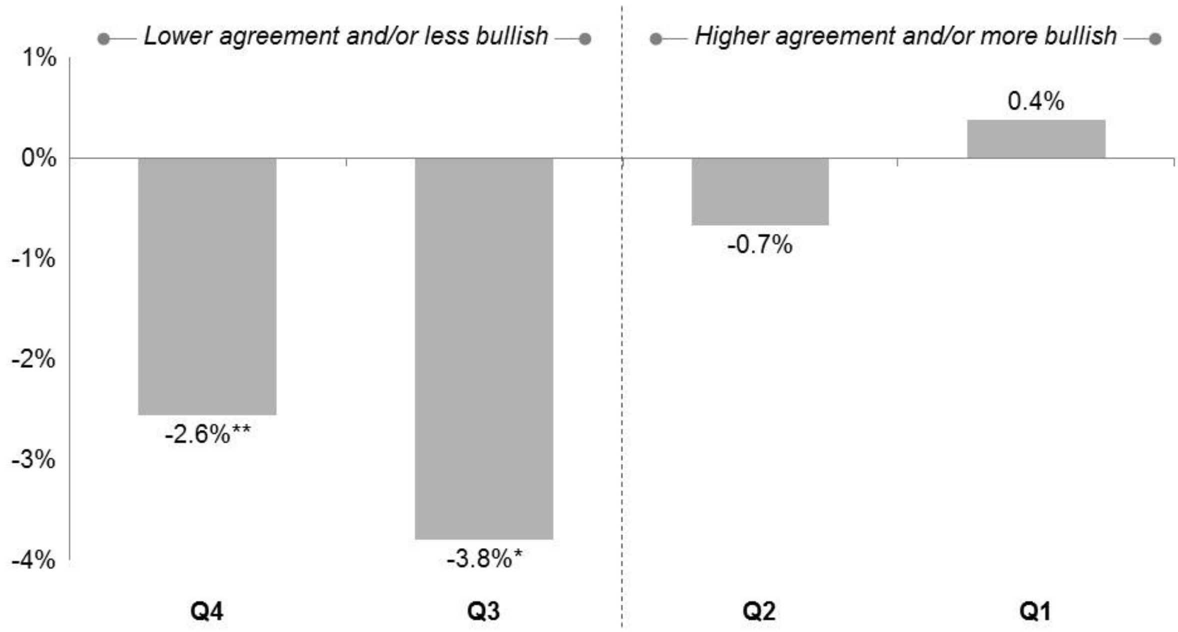

Figure 3-7. Average Unexpected Earnings (SUEAF) by Agreement/Sentiment Quartiles
The figure presents time-series averages of annual mean values of unexpected earnings based on analyst forecasts, within *Agreement x Bullishness* quartiles. The weights are based on the number of messages posted a week before the actual earnings announcement. SUEAF is the difference between the median analyst forecast over the 90-day-period before the announcement and actual earnings divided by the year-end price. ***, ** and * indicate significance at the 1%, 5%, and 10% levels, respectively.

5% level. For quartile 3, we find a negative mean *SUEAF* of -3.8%, and it is significant at the 10% level. The mean *SUEAF* turns into positive for quartile 1, however, not found significant anymore. The difference of -2.2% between quartile 4 (lowest sentiment and agreement level) and 1 (highest sentiment and agreement level) is significant at the 5% level, based on the Satterthwaite method. The trend depicted in Figure 3-7 thus suggests that low levels of combined sentiment and agreement convey negative information about earnings surprises. This finding is as expected; however one cannot undoubtfully argue that either negative sentiment or high levels of disagreement convey negative information about future earnings surprises. Both variables must be treated jointly in this discussion.

3.5.2. Regressions on Earnings Surprises

In this section, we conduct a cross-sectional fixed-effects regression with firm-year clustered standard errors on SUEAF and SUEHIST to analyze the relation between message board variables and earnings surprises while controlling for factors that may affect this relation. The starting point of the regressions (see results in Table 3-20, column (1)) is as follows:

$$\begin{aligned} SUEAF_{i,t} \;/\; SUEHIST_{i,t} \\ &= \alpha_t + \beta_{1,t} LogMessage_{i,7} + \beta_{2,t} Bullishness_{i,7} \\ &+ \beta_{3,t} Agreement_{i,7} + \gamma_1 LogSize_{i,t} + \gamma_2 Volatility_{i,10} \\ &+ \gamma_3 Return_{i,50} + \gamma_4 Return_{i,5} + \gamma_5 LogTurnover_{i,50} + + \varepsilon_{i,t} \end{aligned} \tag{26}$$

The main message board and control variables are described in the previous section. We perform the regression for both types of earnings surprises, $SUEAF_{i,t}$ and $SUEHIST_{i,t}$ to examine whether retail investors on social media relate to specific events or information in their discussions. If retail investors developed to sophisticated, well-informed investors, then we would expect them to acquire the most relevant financial analyst reports before the earnings announcement. The primary variables of interest are the message board related variables (incl. interaction terms). We extend the regression with the dummy variables *High-/LowAgreeD* indicating whether the messages posted can be assigned to the top or bottom 20%-quintile at the end of the fiscal year, analogous to Akbas (2016), to test the impact of abnormal agreement or disagreement on earnings announcement. We then add the interaction terms $BullInd \; x \; AgreeInd_{i,7}$ to test the mutual relation of this score with earnings announcements. Results are tabulated for *SUEAF* (*SUEHIST*) in Table 3-20 (Table 3-21).

For $SUEAF_{i,t}$, the results show that the coefficient estimates for bullishness (agreement) is negative (positive) and significant at the level of 10% (5%) in the basis regression (Table 3-20 (column 1)). As we segregate the high and low agreement quintiles (Table 3-20 (column 2)), we find that it is *HighAgreeD* which is positively related to $SUEAF_{i,t}$. By adding the interaction term *BullInd x AgreeInd*, we find a positive relation of bullishness and agreement with $SUEAF_{i,t}$ with a positive coefficient estimate of +0.082 which is significant at the 1% level (Table 3-20 (column 3)). In other words, the higher the bullishness, the higher the impact of the agreement

Table 3-20. Message Board Activity as Predictor of Earnings Surprise (SUEAF)

Firm- and year-fixed regressions were conducted. T-statistics computed are based on standard errors clustered by firm and year and are denoted in parentheses. $SUEAF_{i,t}$ is the difference in actual EPS and forecasted EPS using analyst forecasts 90 days prior to the earnings announcement date scaled by the stock price of the end of the year, $LogMessages_{i,7}$ is the log transformation of (1 + M_t) for the event window [-7, -1], $Bullishness_{i,7}$ is the cumulated bullishness index using formula (12), $Agreement_{i,7}$ is the cumulated agreement index using formula (13), High/LowAgree$D_{i,7}$ is a dummy variable indicating the cumulated agreement index to be in the top/bottom 20-percentile, $Return_{50}$ is the compounded return over the period of [-61, -12] and $Return_5$ for the five-day return period [-6, -2] prior to the earnings announcement date. $LogTurnover_{50}$ is the log-transformed average turnover over the time window [-61, -12] prior to the earnings announcement date, $Volatility_{i,10}$ is the standard deviation of daily returns in the time window [-11, -2] and $LogSize_{i,t}$ is the log-transformation of the market capitalization at the end of the financial year. ***, **, and * describe significance at 1%, 5% and 10% level, respectively.

	(1)	(2)	(3)	(4)	(5)
	$SUEAF_{t0}$	$SUEAF_{t0}$	$SUEAF_{t0}$	$SUEAF_{t0}$	$SUEAF_{t0}$
$LogMessages_{i,7}$	0.019	0.032**	-0.012	-0.007	
	(0.013)	(0.013)	(0.014)	(0.019)	
$Bullishness_{i,7}$	-0.022*	-0.028**	-0.061***	-0.037***	
	(0.012)	(0.012)	(0.018)	(0.014)	
$PosSentiment_{i,7}$					0.000
					(0.011)
$NegSentiment_{i,7}$					0.029
					(0.020)
$Agreement_{i,7}$	0.060**		-0.045		
	(0.026)		(0.029)		
$HighAgreeD_{i,7}$		0.147*		0.016	0.144*
		(0.078)		(0.082)	(0.083)
$LowAgreeD_{i,7}$		0.084		0.049	0.083
		(0.068)		(0.073)	(0.068)
BullInd x AgreeInd $_{i,7}$			0.082***		
			(0.020)		
BullInd x HighAgreeD $_{i,7}$				0.053***	
				(0.018)	
BullInd x Low_AgreeD $_{i,7}$				-0.015	
				(0.023)	
$Return_{50}$	-0.057**	-0.066**	-0.070***	-0.070***	-0.062**
	(0.024)	(0.026)	(0.024)	(0.025)	(0.026)
$Return_5$	0.539**	0.467**	0.522**	0.464***	0.466**
	(0.228)	(0.181)	(0.206)	(0.178)	(0.183)
$LogTurnover_{50}$	0.013	0.017	0.024	0.025	0.013
	(0.020)	(0.020)	(0.020)	(0.020)	(0.019)
$Volatility_{i,10}$	-2.158**	-2.230**	-1.978**	-1.876**	-2.203**
	(0.945)	(0.930)	(0.884)	(0.893)	(0.955)
$LogSize_{i,t}$	0.002	-0.008	-0.008	-0.013	-0.002
	(0.021)	(0.023)	(0.020)	(0.022)	(0.022)
Constant	-0.227	-0.129	-0.016	0.018	-0.220
	(0.377)	(0.373)	(0.357)	(0.357)	(0.377)
Observations	479	479	479	479	479
Adjusted R^2	14.6%	17.5%	22.0%	20.7%	16.5%

Table 3-21. Message Board Activity as Predictor of Earnings Surprise (SUEHIST)

Firm- and year-fixed regressions were conducted. T-statistics computed are based on standard errors clustered by firm and year and are denoted in parentheses. $SUEHIST_{i,t}$ is the difference in actual EPS in year *t* and the previous year actual EPS scaled by the stock price of the end of the year, $LogMessages_{i,7}$ is the log transformation of (1 + M_t) for the event window [-7, -1], $Bullishness_{i,7}$ is the cumulated bullishness index using formula (3), $Agreement_{i,7}$ is the cumulated agreement index using formula (4), $High/LowAgreeD_{i,7}$ is a dummy variable indicating the cumulated agreement index to be in the top/bottom 20-percentile, $Return_{50}$ is the compounded return over the period of [-61,-12] and $Return_5$ for the five-day return period [-6, -2] prior to the earnings announcement date. $LogTurnover_{50}$ is the log-transformed average turnover over the time window [-61, -12] prior to the earnings announcement date, $Volatility_{i,10}$ is the standard deviation of daily returns in the time window [-11, -2] and $LogSize_{i,t}$ is the log-transformation of the market capitalization at the end of the financial year. ***, **, and * describe significance at 1%, 5% and 10% level, respectively.

	(1) $SUEHIST_{t0}$	(2) $SUEHIST_{t0}$	(3) $SUEHIST_{t0}$	(4) $SUEHIST_{t0}$	(5) $SUEHIST_{t0}$
$LogMes_{i,7}$	-0.597	-0.587	-0.551*	-0.672*	
	(0.368)	(0.369)	(0.329)	(0.383)	
$BullInd_{i,7}$	0.402	0.414	0.445	0.127	
	(0.344)	(0.336)	(0.396)	(0.166)	
$PosSentiment_{i,7}$					-0.127
					(0.110)
$NegSentiment_{i,7}$					-0.792
					(0.505)
$AgreeInd_{i,7}$	-0.543		-0.405		
	(0.651)		(0.540)		
$High\_Agree\_D_{i,7}$		0.337		-0.643	0.015
		(0.449)		(0.972)	(0.582)
$Low\_Agree\_D_{i,7}$		0.888**		-0.365	0.957***
		(0.348)		(0.729)	(0.319)
BullInd x AgreeInd $_{i,7}$			-0.105		
			(0.204)		
BullInd x HighAgreeD $_{i,7}$				0.374	
				(0.252)	
BullInd x LowAgreeD $_{i,7}$				0.828	
				(0.558)	
$Return_{50}$	-0.374	-0.357	-0.362	-0.335	-0.362
	(0.286)	(0.297)	(0.302)	(0.316)	(0.279)
$Return_5$	1.867**	1.435**	1.897**	1.040	1.401**
	(0.744)	(0.714)	(0.773)	(0.649)	(0.693)
$LogTurnover_{50}$	-0.194	-0.203	-0.209	-0.232	-0.156
	(0.205)	(0.208)	(0.217)	(0.212)	(0.189)
$Volatility_{i,10}$	12.003	13.132	11.808	12.423	15.509
	(9.286)	(9.376)	(9.311)	(9.597)	(10.019)
$LogSize_{i,t}$	0.335*	0.337*	0.351*	0.367*	0.307*
	(0.178)	(0.184)	(0.190)	(0.194)	(0.171)
Constant	-4.974***	-5.852***	-5.312***	-5.021***	-5.586***
	(1.743)	(1.851)	(1.904)	(1.875)	(1.723)
Observations	560	560	560	560	560
Adjusted R^2	10.0%	11.4%	10.0%	14.5%	11.4%

on $SUEAF_{i,t}$, and vice versa. Investors using social media relate information from analyst reports with their newest findings and analysis. Information about changes of a firm's fundamental values are discussed, and results suggest that situations in which investors are bullish and agreed on lead to higher earnings surprises. One must consider in our results, that social media users in our analysis could have a timing advantage against financial analysts since we consider analyst reports of the past 90 days for our earnings surprise calculation. Since our results only hold for the time window of [-7, -1], we can assume that retail investors on social media have sufficient time to access older reports and invest the effort to interpret and extend the information content of the report. The overall regression results in this section are in line with our previous finding in the portfolio section that increasing score of bullishness and agreement are a positive signal for earnings surprises.

The results for $SUEHIST_{i,t}$, on the other hand, did not show any relevant significance for message board variables. Our results, therefore, suggest that retail investors on social media are important market participants who disseminate value-relevant information and thus contribute to the improvement of market efficiency. They discuss, interpret and disseminate information depending on the type of event, sentiment, and agreement among the users.

3.6. Conclusion

We investigate the differential information content of internet message boards in non-specific event setups and surrounding annual earnings announcements. We first find that positive sentiment is positively related to abnormal returns, but the effect diminishes after a month. In the short-term, the relation holds in both directions but with implications that positive sentiment follows the previous day excelling stock performance. Furthermore, we observe a pattern of noise trading activity surrounding events with abnormal positive sentiment postings. More specifically, abnormal returns are positively contemporaneously associated with abnormal positive sentiment postings, however with negative return reversals on the subsequent days. Short selling activities reduce this presumably observed contemporaneous overreaction in firm's abnormal returns. We argue that only informed sellers initiate short selling activities when they believe that sentiment diverges far beyond a firm's fundamentals. Hence, short sellers arbitrage against noisy sentiment traders. However, due to limits of arbitrage and hyping of rather small stocks we do see a remaining contemporaneous relation between positive sentiment and a firm's abnormal returns.

Secondly, we find that negative sentiment incorporates information about stock underperformances with a negative correlation of up to one month as analyzed in this paper. Contrary to the characteristics of positive sentiment postings, we find indications that negative sentiment predicts the underperformance of stocks compared to the market in the short-term. Abnormal return reversals into positive remain absent after days of abnormal high postings with negative sentiment. The impact of negative sentiment is thereby much more economically meaningful compared to messages with positive sentiment. As the questions arise if social media might induce market volatility, we thirdly find that increased internet message board postings are rather caused by previous stock price swings than vice versa. Even though our findings imply a bilateral-direction in causality, the impact of message board activity on volatility reveals only modest economic significance. Lastly, our results provide evidence that message board sentiment and agreement – or sentiment homogeneity - amongst the users predict earnings surprises using analyst forecasts. This is in line with former studies (Chen et al., 2014; Leung and Ton, 2015) which propagate the dissemination of value-relevant information through internet message boards or social media outlets.

We summarize our findings that internet message boards as an outlet of social media have a substantial impact on equities markets however with significant differential effects depending on the sentiment and the surrounding events. Additionally, regulators should succumb to the discussion whether arbitrageurs contribute to the price stabilization process, especially in noisy market environments.

4. Investor Sophistication and Attention in Target Price Run-Ups

ABSTRACT: We analyze direct and indirect measures of investor attention before M&A announcements and their relation to the well-documented phenomena of target run-ups. Controlling for attention of (un-)sophisticated individual and institutional investors, we show that dedicated internet investment platforms contribute to the run-ups of small enterprise M&A targets. The fundamental characteristics of these firms covered only in internet investment platforms do not economically differ from other small firms that receive no (social) media attention. Contrarily, institutional investors covering large stocks are influenced by analyst recommendations. Altogether, the results are consistent with the market expectations hypothesis around M&A announcements.

4.1. Introduction

"*But whether a contrarian or a trend follower, an investor is less likely to purchase a stock that is out of the limelight*." (Barber and Odean, 2008, p. 813).

M&A (mergers and acquisitions) literature shows that the target run-up phenomenon occurs around two to three months before the first official bid announcement (Keown and Pinkerton, 1981; Asquith, 1983; Schwert, 1996; King, 2009; Brigida and Madura, 2012). One strand of research argues that run-ups result from efficient markets, which quickly process valuable information from news articles or rumors (Gupta and Misra, 1989; Murray, 1994; Zivney et al., 1996; Clarkson et al., 2006; Gao and Oler, 2012; Siganos and Papa, 2012) or pre-announcement relationships based on toeholds (Mikkelson and Ruback, 1985; Choi, 1991). Alternatively, Meulbroek (1992) explains that target run-ups are susceptible to insider information leakage because bid announcements are valuation-relevant events. This idea is extended by Tang and Xu (2016) who show that unreported insider trading causes target run-ups.

The two channels put forth to explain target run-ups, however, neglect the issues that 1) institutional and individual investors would not only have access to different sources of information but differ in the resources they have available to process this information and also in the quality of their investment decisions; 2) unsophisticated and sophisticated individual investors have varying impact on financial markets; and 3) new information is only valuable if the receiver pays attention to it (Da et al., 2011). For example, Huberman and Regev (2001) demonstrate that stock prices only significantly react to information when investors pay attention to it.

We study the contribution of investor sophistication and attention amongst institutional and individual investors to target run-ups prior M&A announcements. People purchase stocks with the belief that they will increase in value. In traditional efficient market theories, those gains are based on valuable information which is processed efficiently by market participants and is thus incorporated into stock prices. However, how do institutional and individual investors catch the attention of potential target firms who seem to gain in value? Barber and Odean (2008) suggest that it is the extreme returns that catch an investor's attention and that these returns are related to news or attention-grabbing information. Thus, extreme returns act as news vis-à-vis events without other details. Investors face a fundamental search problem in a world of thousands of stocks (Barber and Odean, 2008). This leads to the question of how investors search for winner stocks in the M&A context. There are several theories and models existing which try to identify potential M&A targets based on firm characteristics and performances (e.g., Palepu, 1986; Comment and Schwert, 1995).

We measure sophisticated individual investor attention using posting activity on HotCopper, which is the leading online Australian equities discussion platform. Individual investors receive attention surrounding a stock when they actively post messages and discuss them on social media platforms. Da et al. (2011) show that internet search volume is likely to capture the attention of unsophisticated individual investors. Hence, we further introduce merger attention signals based on internet search queries on Google. In Australia, Google's search engine is the market leader with a total market share of 94% of total search inquiries. We can, therefore, assume that financial Google search queries cover a broad audience of unsophisticated individual investors. In another study, Drake et al. (2012) find that abnormal internet search queries preempt information on upcoming firm events such as earnings announcements.

Furthermore, the authors argue that individual investor attention is positively related to media coverage but negatively associated with investor distraction.[25] In addition to the previous studies, there have been several further studies added to the investor attention literature which make

[25] Drake et al. (2012) describe investor distraction as the level of competing earnings news.

use of internet search queries as a new direct measure of individual investor attention (Vlastakis and Markellos, 2012; Vozlyublennaia, 2014; Andrei and Hasler, 2015; Da et al., 2015; Welagedara et al., 2017). However, these studies do not disentangle the relation between investor attention and financial market activities (e.g., index returns for stocks, bonds, and commodities, return volatility, analyst recommendation revisions, earnings announcements) for unsophisticated and sophisticated individual investors.

Prior studies on run-ups, such as from Jarrell and Poulsen (1989) and Schwert (1996), test the relation between press speculation and run-ups. However, these studies do not distinguish the behavioral characteristics between individual and institutional investors. In a pre-IPO-setting, Liu et al. (2009) claim that media coverage reflects institutional investor attention. On the contrary, Fang and Peress (2009) find that analyst coverage is more likely to be the predominant information source for institutional investors and that (traditional) news media coverage serves individual investors. Womack (1996) points out that institutional investors obtain their information from costly databases and brokerage firms which provide sophisticated analyst reports. These reports are by nature evaluative and predictive. Analyst recommendations are future predictions of stock valuations and thus include all available sources of costly industry and firm-related information. Furthermore, buy and sell recommendations directly test the intrinsic value of an analyst's information. It is well known that target firm valuations are particularly subject to bid announcements. We employ changes in analyst recommendations from Institutional Brokers' Estimate System (I/B/E/S) and traditional news media from Thomson Reuters News Analytics (TRNA) to examine the relationship between institutional investor attention and target run-ups.

Further, evidence reveals that target run-ups still occur in the Australian equity market despite the introduction of tighter guidelines surrounding the leakage of price sensitive information prior to bid announcements. The ASIC (Australian Securities and Investment Commission) examined the extent of media leakage two weeks before 40 takeovers (46% of total value) and 40 equity raisings (22% of total secondary equity) between July 2006 and March 2013. During this period, the industry-led Governance Institute of Australasian Investor Relations Association, with the support of ASIC, introduced guidelines to enforce the release of price-sensitive information on July 1st, 2010. It was shown that 45% of takeovers and 35% of equity raisings were leaked prior the installment of the guidelines and the level of leakage dropped to 20% for takeovers and moderately increased to 40% for equity raisings after the installment. The Australian Financial Review and The Australian newspapers were the most active reporters of transaction leakage.[26] The occurrence of information leakage despite the strict regulatory guidelines in the Australian equity market provides the ideal setup for us to test the impact of traditional news media and social media attention around bid announcements.

We find that smaller and underperforming stocks that only catch the attention of sophisticated individual investors on HotCopper experience a significantly higher target run-up before bid announcements. Firms with similar fundamental characteristics but no (social) media attention do not experience a significant run-up, but only a short-term announcement effect. Large firms, on the other hand, are found to be sensitive to analyst opinions. Positive and negative analyst upgrades have significant influences on target run-ups in their respective directions. Merger signals in traditional news media only appear to play a minor role for institutional investors. Also, Google search activity before bid announcements of unsophisticated investors does not significantly or economically meaningfully explain target run-ups. In summary, we find distinctive drivers of institutional and individual investor attention which explain target run-ups and they are consistent with the market expectations hypothesis.

The rest of the paper is organized as follows. Section 4.2 describes the data and research design of this study. Section 4.3 analyzes the timing effects and magnitudes of target run-ups for different types of traditional and social media attention. Section 4.4 provides evidence that institutional and sophisticated individual investor measures capture investor attention before bid announcements. Whether investors are net buyers or net sellers around attention-grabbing events is tested in section 4.5. In section 4.6, we test the relevance of unsophisticated investor attention via internet search queries. Section 4.7 examines the relation between target run-up characteristics and markup pricing around bid announcements. Section 4.8 and 4.9 discuss potential endogeneity issues and conclude the study.

[26] http://download.asic.gov.au/media/1344584/rep393-published-27-May-2014.pdf.

4.2. Data and Research Design

M&A announcement data was downloaded from the Thomson Reuters Securities Data Company's (SDC) database. The sample consists of acquirer firms which can either be a private or public firm, and takeover targets which are listed on the Australian Securities Exchange (ASX). The highly regulated Australian market requires persons or firms to report a potential acquisition to the relevant market operator (typically the ASX) within two days as defined in the Corporations Act 2001, when ownership thresholds of 5% (so-called "substantial holdings") are exceeded or a 1% change in ownership is pursued. Exceeding ownership thresholds of 20% are only allowed under specified circumstances.[27] As we strive to examine M&A announcements effects collectively, we also include acquisition announcements which potentially result in minority stakes (<50% after acquisition). We do not restrict our study to full ownership persuasions since we want to examine the overall relation of announcements for (un-)sophisticated individual and institutional investors with such public announcements.

The sample consists of 3,165 transaction announcements between January 2008 and August 2015 with sufficient returns data available for our analysis. We obtained our trading data from Securities Industry Research Centre of Asia-Pacific (SIRCA). To circumvent the impact of confounding events and especially multiple announcements for the same target company within a short-term period, we only included the first takeover announcement within a two-month period. This filter yields 2,765 takeover events.

For an approximation of institutional investor attention, we include 15,135 news articles and alerts about 513 firms from the TRNA database which occurred in the two-month period before the actual takeover announcement. In order to differentiate between a random press coverage (press article or alert regardless of content) or dedicated merger signals (specific takeover content), we analyzed the news and alert headlines for M&A related keywords. To add another indirect proxy of institutional investor attention, we also included analyst up- and downgrades data which was extracted from the I/B/E/S database.

[27] http://www.clearstream.com/clearstream-en/products-and-services/market-coverage/asia- pacific/australia/disclosure- requirements ---australia/7380, November 15th, 2017.

To analyze the relationship between individual investor attention and M&A announcement effects, we use data from HotCopper, the most popular internet investment platform in Australia. We find 83,988 messages posted on 1,045 distinct firms in this forum in the two-month period before the takeover announcement. Since price sensitive ASX announcements are directly linked to HotCopper, we use the HotCopper database to control for public firm announcements which are posted with directly related company tickers.

For each event, we examine whether it was covered by traditional and social media in a two-month period before the M&A announcement date. If we find a takeover signal and posts on the social media investment platform HotCopper, we classify the event as a deal with "full media" coverage. In the case of single takeover signals or HotCopper coverage, we define these events as "only news" or "only HotCopper". Among the 2,765 events in our sample, there is full media coverage for 407 target firms. We find for the majority of 1,779 announcements that firm talks solely took place on the investment platform HotCopper.

Panel A in Table 4-1 shows the yearly distribution of our M&A sample. The number of M&A announcements decreased continuously after the period of the financial crisis in 2008/2009. The share of M&A announcements only experiencing news signals remains relatively low. M&A announcements just catching the attention of HotCopper users account for the largest group across all years. Panel B in Table 4-1 presents the industry distribution of our sample according to the SDC industry classification. The Materials industry sector constitutes about half of our sample, followed by Financials and Energy & Power with each accounting for nearly 10% of the total sample. This industry distribution remains consistent amongst all subsamples. We, therefore, find no bias or tendencies in our sample that, for example, Materials companies experience full media attention compared to other sectors.

Table 4-2 reports the financial summary statistics for target deal characteristics. Details about the variable construction are described in Table A-11 in the Appendix. The table presents the means, medians and standard deviations for the four subsamples based on their (social) media coverage followed by the total sample. To better interpret the differences of characteristics for distinctive subsamples, we report the differences in means and medians for each target characteristic and subsample and test the mean differences with the Welsh-test to account for differences in variances of the different samples.

Table 4-1. M&A Announcement Sample Distribution

This table shows the sample distribution of 2,765 Australian M&A announcements between Jan. 2008 and Aug. 2015. The target is public and the acquirer is either private or public. Panels A and B present the number of M&A announcements by year and industry, respectively. The full sample consists of all four types of media coverage.

Panel A: By announcement year

			Media coverage							
	Full sample		Full media		Only news		Only HC		No media	
Year	#	%	#	%	#	%	#	%	#	%
2008	522	18.9	46	11.3	24	31.6	263	14.8	189	37.6
2009	511	18.5	43	10.6	10	13.2	368	20.7	90	17.9
2010	462	16.7	41	10.1	8	10.5	353	19.8	60	11.9
2011	305	11.0	47	11.5	9	11.8	215	12.1	34	6.8
2012	366	13.2	65	16.0	10	13.2	251	14.1	40	8.0
2013	269	9.7	51	12.5	3	3.9	186	10.5	29	5.8
2014	216	7.8	72	17.7	6	7.9	103	5.8	35	7.0
2015*	114	4.1	42	10.3	6	7.9	40	2.2	26	5.2
Total	**2,765**	**100.0**	**407**	**100.0**	**76**	**100.0**	**1,779**	**100.0**	**503**	**100.0**

Panel B: By target industry

			Media coverage							
	Full sample		Full media		Only news		Only HC		No media	
SDC Industry Classification	#	%	#	%	#	%	#	%	#	%
Consumer Prod. & Serv.	107	3.9	15	3.7	1	1.3	69	3.9	22	4.4
Consumer Staples	110	4.0	26	6.4	1	1.3	65	3.7	18	3.6
Energy and Power	261	9.4	47	11.5	8	10.5	144	8.1	62	12.3
Financials	298	10.8	32	7.9	6	7.9	198	11.1	62	12.3
Healthcare	124	4.5	11	2.7	6	7.9	84	4.7	23	4.6
High Technology	156	5.6	14	3.4	0	0.0	119	6.7	23	4.6
Industrials	148	5.4	31	7.6	4	5.3	89	5.0	24	4.8
Materials	1,235	44.7	175	43.0	39	51.3	802	45.1	219	43.5
Media and Entertainment	87	3.1	19	4.7	2	2.6	55	3.1	11	2.2
Real Estate	146	5.3	22	5.4	5	6.6	100	5.6	19	3.8
Retail	52	1.9	11	2.7	2	2.6	31	1.7	8	1.6
Telecommunications	41	1.5	4	1.0	2	2.6	23	1.3	12	2.4
Total	**2,765**	**100.0**	**407**	**100.0**	**76**	**100.0**	**1,779**	**100.0**	**503**	**100.0**

Table 4-2. Financial Descriptives of Sample

This table presents the mean, median and standard deviation summary statistics of M&A announcements between Jan. 2008 and Aug. 2015 for events with data available (at least 1,998 events). Observations are winsorized within a 2%-percentile window to control for possible outliers.

	Full media	*Only news*	*Only HC*	*No media*	*Total*
Market capitalization [mn. AUD]					
Mean	2,071.13	8,638.85	191.19	150.43	696.38
Median	312.81	1,020.69	22.99	25.03	33.60
Std. Dev.	6,039.37	20,460.49	1,121.36	701.01	4,471.00
Equity ratio					
Mean	0.60	0.56	0.68	0.71	0.67
Median	0.61	0.57	0.77	0.83	0.75
Std. Dev.	0.28	0.30	0.32	0.31	0.31
EPS ltm [AUD]					
Mean	-15.49	0.40	-2.33	-4.55	-4.62
Median	0.00	0.09	-0.02	-0.02	-0.01
Std. Dev.	254.26	4.52	38.53	64.47	106.50
Market-to-Book					
Mean	2.34	2.84	2.28	2.37	2.32
Median	1.45	1.44	1.14	1.38	1.24
Std. Dev.	3.14	3.89	3.43	3.22	3.36
Scaled EBITDA ltm					
Mean	0.00	-0.02	-0.13	-0.18	-0.11
Median	0.03	0.02	-0.05	-0.08	-0.04
Std. Dev.	0.21	0.33	0.40	0.41	0.38
Net analyst up-/downgrades					
Mean	-1.61	-2.53	-0.38	-0.27	-0.60
Median	0.00	0.00	0.00	0.00	0.00
Std. Dev.	8.42	13.12	3.43	2.51	4.91

To test whether differences may result from differences in distributions, namely location and shape of the distribution, we apply the Mann-Whitney-U-test for the test in differences of medians. [28] The results in differences are reported in Table 4-3. It is interesting to observe that target firms which solely receive social media attention are not significantly different to firms with non-media attention regarding firm size (market capitalization), earnings (EPS), valuations (market-to-book) and change in analyst recommendations. Furthermore, they show lower

[28] Tests in differences of medians only apply, if shapes of distributions do not significantly differ from each other.

equity ratios and higher operating performances (EBITDA), which is significant but not of economic importance compared to the other subsamples. Compared to firms that additionally or only experience acquisition signals on news media, firms only talked about on social media are significantly smaller (market capitalization), possess higher equity ratios, have earned less (EPS), lower valuations (market-to-book), lower operative performances (EBITDA), and fewer analyst downgrades. These results suggest that small target firms which only catch attention on social media are similar to those firms which experience no media attention at all. Furthermore, these firms appear to be undervalued and underperforming compared to firms also covered by traditional media.

4.3. Event Study with M&A Announcements and Merger Signals

We follow Schwert (1996) and define target *run-ups* as $\sum_{-t}^{-1} AR_{it}$, target *markups* as $\sum_{-1}^{+t} AR_{it}$, and target *premium* as the summation of *run-up* and *markup*. We examine the pre-bid and markup abnormal returns using the market model regression approach. For each target firm, we calculate the abnormal returns (or market excess returns) before a public M&A announcement using:

$$AR_{it} = R_{it} - (\alpha_i + \beta_i R_{mt}) \tag{27}$$

where R_{it} is the compounded return of target firm *i* and R_{mt} the market-weighted All Ordinaries Index return for day *t*. Following Cai and Sevilir (2012) and also motivated by Schwert (1996), we estimate the market model parameters over 200 trading days ending two months before the public M&A announcement. We analyze cumulative abnormal returns for different time windows based on:

$$CAR_{it} = \sum_{-t}^{+t} AR_{it} \qquad t = -40, \ldots, +10 \tag{28}$$

Since we classify news articles or alerts as one part of our subsample, we investigate the relation between news/social media coverage and target run-ups; and not the magnitude of a target run-up itself. The phenomenon of target run-ups is documented in the literature, even

Table 4-3. Average and Median CAR-Differences between Subsamples

This table presents the average and median differences in firm characteristics. The Welch-test and Mann-Whitney-U-test were applied to test the mean and median differences, respectively. ***, **, and * describe significance at 0.1%, 1% and 5% level, respectively.

Differences of full media vs.			***Only news sample***		***HotCopper sample***		***No media sample***	
			Average	Median	Average	Median	Average	Median
Market cap. [mn. AUD]			-6,567.72**	-707.88***	1,879.95***	289.82***	1,920.71***	287.78***
Equity ratio			0.05	0.04	-0.07***	-0.16***	-0.11***	-0.22***
EPS ltm [AUD]			-15.88	-0.09*	-13.16	0.02***	-10.94	0.02***
Market-to-Book			-0.50	0.02*	0.06	0.31***	-0.03	0.07***
Scaled EBITDA ltm			0.02	0.00	0.13***	0.08***	0.18***	0.11***
Net analyst up-/downgrade			0.91	0.00	-1.23**	0.00***	-1.34**	0.00**
Only news sample vs.	***Full media sample***				***HotCopper sample***		***No media sample***	
	Average	Median			Average	Median	Average	Median
Market cap. [mn. AUD]	6,567.72**	707.88***			8,447.67***	997.70***	8,488.43***	995.66***
Equity ratio	-0.05	-0.04			-0.12**	-0.20***	-0.16***	-0.26***
EPS ltm [AUD]	15.88	0.09*			2.73*	0.11***	4.94	0.11***
Market-to-Book	0.50	-0.02*			0.56	0.29***	0.47	0.05***
Scaled EBITDA ltm	-0.02	0.00			0.11*	0.08***	0.15***	0.10***
Net analyst up-/downgrade	-0.91	0.00			-2.15	0.00	-2.25	0.00
HotCopper sample vs.	***Full media sample***		***Only news sample***				***No media sample***	
	Average	Median	Average	Median			Average	Median
Market cap. [mn. AUD]	-1,879.95***	-289.82***	-8,447.67***	-997.70***			40.76	-2.04
Equity ratio	0.07***	0.16***	0.12**	0.20***			-0.04*	-0.06**
EPS ltm [AUD]	13.16	-0.02***	-2.73*	-0.11***			2.22	0.00
Market-to-Book	-0.06	-0.31***	-0.56	-0.29***			-0.09	-0.24
Scaled EBITDA ltm	-0.13***	-0.08***	-0.11*	-0.08***			0.05*	0.03**
Net analyst up-/downgrade	1.23**	0.00***	2.15	0.00			-0.11	0.00
No media sample vs.	***Full media sample***		***Only news sample***		***HotCopper sample***			
	Average	Median	Average	Median	Average	Median		
Market cap. [mn. AUD]	-1,920.71***	-287.78***	-8,488.43***	-995.66***	-40.76	2.04		
Equity ratio	0.11***	0.22***	0.16***	0.26***	0.04*	0.06**		
EPS ltm [AUD]	10.94	-0.02***	-4.94	-0.11***	-2.22	0.00		
Market-to-Book	0.03	-0.07***	-0.47	-0.05***	0.09	0.24		
Scaled EBITDA ltm	-0.18***	-0.11***	-0.15***	-0.10***	-0.05*	-0.03**		
Net analyst up-/downgrade	1.34**	0.00**	2.25	0.00	0.11	0.00		

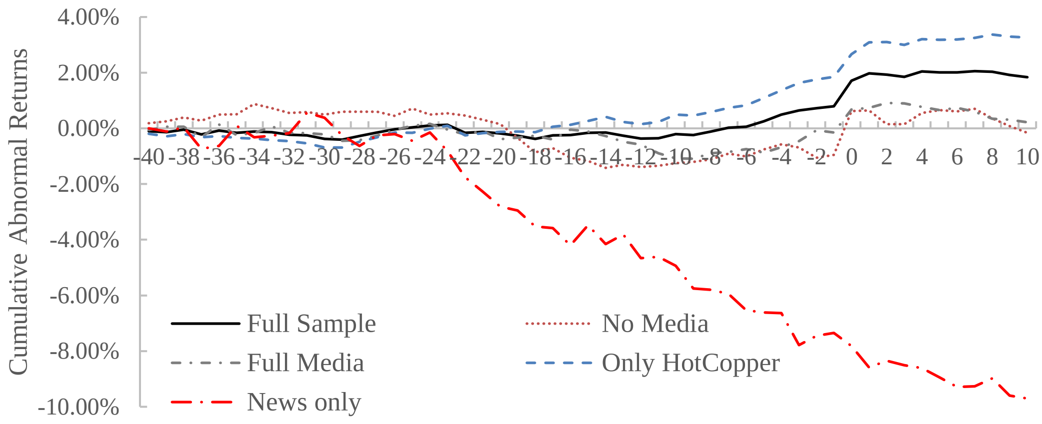

Figure 4-1. Average CARs around M&A Announcements

though it is still open to debate what causes these run-ups (Schwert, 1996). Figure 4-1 depicts the *CARs* in the time window of *[-40; +10]* around the public M&A announcement date on day t_0. It is striking that the subsample CARs develop significantly different depending on the degree of (social) media coverage. Table 4-4 and Table 4-5 show the average and median CARs around M&A announcements, respectively. In Table 4-4 Panel A, the full sample shows highly significant target premiums of around 2% depending on the time horizon. Nearly half of the premium (0.93%) results from abnormal returns occurring on the announcement date.

The lower level of M&A premiums in this full sample compared to the universal M&A literature mainly result from the filter applied. We also included M&A announcements for minority positions (<50%) and do not filter for deal value size (usually > $1-5 million), because we are also interested in individual investor trading behavior which relates to small trade volume and deal sizes. Splitting our full M&A sample into four groups based on whether a M&A announcement was previously covered in traditional or social media, we repeated the event study analysis for mean and median CARs. Most importantly, mean premiums starting 30 or 40 days before the M&A announcement date are significantly different from zero at a significance level of 0.1% for the subsamples only experiencing news signals (around -10%) or only grabbing social media attention (around +3 to 4%), while the full- and non-media sample mean premiums are not significantly different from zero. It is noticeable that a significant target run-up of +2.4% can only be observed for stocks of firms that solely received attention on the internet

Table 4-4. Average CARs around M&A Announcements

This table presents the average CARs based on the market model. We apply the standardized cross-sectional Boehmer-test. ***, **, and * describe significance at 1%, 5% and 10% level, respectively.

Window [-t]	[+t]	*Panel A* **Full sample**	*Panel B* **News & HC**	*Panel C* **News only**	*Panel D* **HC only**	*Panel E* **No media**
-40	10	1.83%*	0.22%	-9.70%**	3.26%***	-0.16%
-30	10	2.08%**	0.40%	-10.27%**	3.80%***	-0.74%
-20	10	1.96%***	0.31%	-7.43%*	3.43%***	-0.47%
-10	10	2.19%***	1.11%	-5.09%*	3.03%***	1.19%
-3	3	1.36%***	1.57%**	-1.86%	1.62%***	0.73%
-2	2	1.29%***	1.37%**	-0.56%	1.47%***	0.84%
-1	1	1.26%***	0.81%*	-1.14%	1.34%***	1.70%***
0	0	0.93%***	0.83%**	-0.47%	0.82%***	1.58%***
-1	-1	0.07%	-0.07%	0.11%	0.09%	0.10%
-2	-2	0.08%	0.39%**	0.34%	0.12%	-0.37%*
-3	-3	0.15%	0.21%	-1.15%***	0.26%**	-0.11%
-4	-4	0.25%**	0.20%	-0.03%	0.28%**	0.19%
-5	-5	0.19%	-0.13%	-0.09%	0.25%	0.25%
1	1	0.26%*	0.05%	-0.77%	0.42%***	0.02%
2	2	-0.05%	0.18%	0.24%	0.02%	-0.49%**
3	3	-0.08%	-0.01%	-0.16%	-0.11%	0.00%
4	4	0.19%	-0.13%	-0.10%	0.22%	0.40%
5	5	-0.03%	-0.12%	-0.33%	-0.03%	0.11%
-10	-1	1.14%***	0.75%	-2.73%*	1.62%***	0.39%
-20	-11	-0.23%	-0.80%	-2.34%	0.40%	-1.66%**
-30	-21	0.12%	0.08%	-2.83%*	0.37%	-0.28%
-40	-31	-0.25%	-0.17%	0.56%	-0.54%	0.59%
-20	-1	0.92%	-0.05%	-5.08%**	2.02%**	-1.27%
-30	-1	1.04%	0.03%	-7.91%**	2.39%**	-1.55%
-40	-1	0.79%	-0.14%	-7.34%**	1.85%	-0.96%
# Events		2,765	407	76	1,779	503

investment platform beginning 30 days before the M&A announcement. One potential explanation for this is based on the attention theory by Barber and Odean (2008). They assert that attention-grabbing events may increase disagreement between investors beliefs and investors only have limited options for portfolio rebalancing. This is consistent with our results and could explain the negative target run-down driven by sell activities of larger institutional investors. Bullish individual investors are able to buy the stock, whereas the bearish ones could only sell if they own the stock or otherwise initiate a short position. As a result, attention-grabbing events, or in this case, social media attention for smaller, underperforming and undervalued stocks would induce net purchases by individual investors and therefore increase stock returns. Limited portfolios and short-sale constraints would hinder individual investors to sell, even if they would like to. However, institutional investors would be able to buy and sell with regards to heterogeneity in beliefs (Barber and Odean, 2008). This leads to the second possible explanation why stocks of larger firms of the "news only" sample experience significantly negative target run-ups and premiums of up to -10.3%.

As described in Table 4-2, stocks in this subsample are usually large capitalization stocks that experience on average more analyst downgrades compared to the other subsamples. Even though the differences in analyst net scores are not significantly different to other subsamples, we find that the mean of -2.53 indicates that this smaller subsample is pre-dominated by stocks that are talked down by analysts.

Womack (1996) and Barber et al. (2001) state in their studies that unfavorable changes in analyst recommendations would negatively impact on the respective stock's return at the time of its release. Barber et al. (2001) also point out that institutional investors react more quickly to such changes in recommendations since smaller individual investors might receive such reports at a later stage. Palepu (1986) and Comment and Schwert (1995) discuss in their studies that accounting and stock market measures of firm's performances contribute to predicting takeover activity. Comment and Schwert (1995) argue that compared to non-M&A targets, potential target firms perform poorly (lower sales growth), possess inefficient capital structures (lower debt/equity ratios) and have lower market-to-book ratios. The latter is explained by different interpretations in prior literature, such as fewer growth options (Myers, 1977), market undervaluation of the target firm (Comment and Schwert, 1995), or inefficient management of the target firm (Lang et al., 1989). The higher target run-ups for firms belonging to the "Only HC" sample compared to firms also covered by traditional news media, might be explained by these factors.

Table 4-5. Median CARs around M&A Announcements

This table presents the median CARs based on the market model. We apply the non-parametric Corrado-Rank-test to test for significance. ***, **, and * describe significance at 1%, 5% and 10% level, respectively.

Window [-t] [+t]	*Panel A* Full sample	*Panel B* News & HC	*Panel C* News only	*Panel D* HC only	*Panel E* No media
-40 10	1.42%*	2.68%	-3.21%	1.98%**	-1.75%
-30 10	1.81%*	3.84%*	-0.21%	2.30%**	-1.73%
-20 10	1.58%**	3.25%*	-0.75%	1.89%***	-0.60%*
-10 10	1.38%***	2.89%**	-1.47%	1.69%***	0.00%
-3 3	0.70%***	1.32%***	-1.52%	0.78%***	0.04%
-2 2	0.50%***	1.15%***	-0.33%	0.51%***	0.00%
-1 1	0.36%***	0.79%***	-0.36%	0.34%***	0.14%***
0 0	0.10%***	0.17%***	-0.16%	0.09%***	0.15%***
-1 -1	0.03%	0.12%	-0.02%	0.03%	0.01%
-2 -2	0.02%	0.16%***	0.08%	0.03%	0.00%
-3 -3	0.02%	0.13%	-0.35%**	0.03%	0.00%
-4 -4	0.03%**	0.13%	-0.05%	0.05%**	0.00%
-5 -5	0.01%	0.02%	-0.02%	0.02%	0.00%
1 1	0.06%**	0.17%	0.24%	0.04%**	0.01%
2 2	0.00%	0.13%	0.13%	0.01%	-0.04%**
3 3	0.00%	0.12%	0.00%	0.00%	0.00%
4 4	0.03%	0.10%	-0.13%	0.03%	0.00%
5 5	0.01%	0.05%	0.02%	0.02%	0.00%
-10 -1	0.21%*	0.95%	-0.27%*	0.31%**	0.00%
-20 -11	0.00%	0.28%	0.18%	0.21%	-0.15%
-30 -21	0.00%	0.00%	-1.86%	0.00%	0.00%
-40 -31	0.00%	-0.32%	-0.01%	0.00%	0.00%
-20 -1	0.24%	0.86%	-0.49%	0.55%*	0.00%
-30 -1	0.23%	1.64%	-0.46%	0.76%	-1.21%
-40 -1	0.09%	1.64%	-0.86%	0.59%	-1.06%
#Events	2,765	407	76	1,779	503

Prior literature on US target run-ups has observed several different magnitudes, such as Sanders and Zdanowicz (1992) with run-ups of 7.4% or Schwert (1996) with run-ups of 13.3% prior to bid announcements. For robustness, we repeat our analysis and cleaned our sample for acquisition announcements in which the acquirer holds a minority or no stake on the target before the bid announcement and pursues a takeover resulting in an ownership of > 50%. This filter yields a total sample size of 352 bid announcements. Results are shown in Figure 4-2. We find an average target run-up of 4.9% starting 40 trading days before and ending one day before the

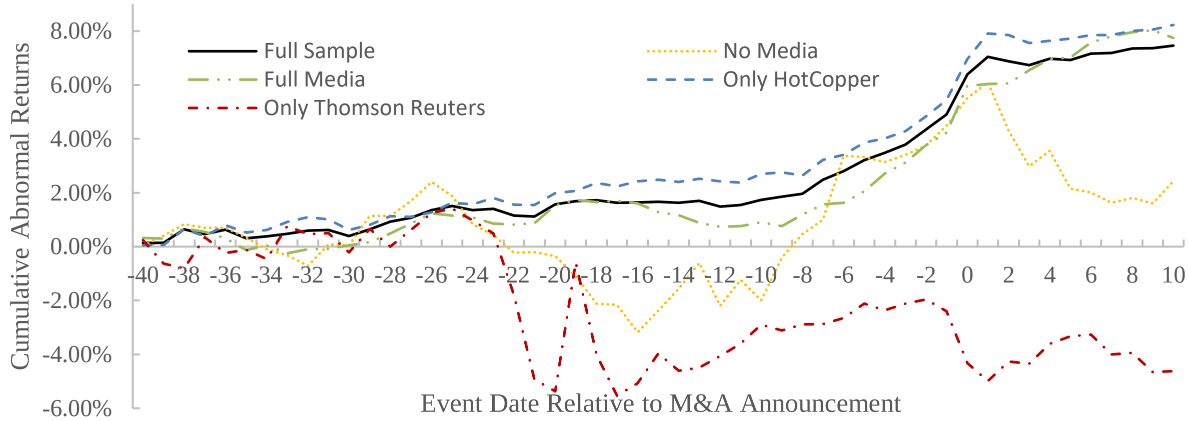

Figure 4-2. Average CARs around M&A Announcements – Change of Control

bid announcement date for the full sample. The "News only" subsample only consists of 5 events and might be confounded by other events around 20 days before the M&A announcement. Interestingly, the "No media" subsample of 27 events experiences a return reversal, showing an overreaction to takeover announcements. Consistent with Asquith (1983), this pattern is similar to patterns of unsuccessful merger announcements after the press release of the failure. Together, the results remain robust with higher levels of target run-ups and markups consistent with prior literature.[29]

Overall, we can summarize two main findings in this section from our analysis. First, small stocks that only receive attention in social media do not differ significantly or economic meaningfully from other small stocks which receive no (social) media attention at all. However, attention for these stocks is significantly related to higher target run-ups and markups around M&A announcements. Secondly, stocks only covered in social media are smaller (market capitalization), have less analyst coverage and changes in recommendations, perform weaker financially (EBITDA, EPS), have higher equity ratios, and lower market valuations (market-to-book) compared to stocks which (only) receive traditional media attention. The latter sample is characterized by a higher net number of analyst downgrades compared to other subsamples, and we find indications that institutional investors react with net sells for these stocks, resulting in target rundowns before M&A announcements.

[29] For further robustness tests, we apply the Fama-French-three factor model and repeat the analysis for the entire sample including minor shareholder ownership. We qualitatively find the same results even though of different levels for target run-ups and markups. Results are not tabulated here for brevity.

4.4. Multivariate Regression Analysis of Pre-Bid Run-Ups

To measure the impact of (un-)sophisticated individual and institutional investor attention on pre-bid run-ups while controlling for target, deal and other characteristics, we perform a multivariate regression analysis specified as follows:

$$Runup_{i,-30} = \alpha + \beta FullMedia_{i,-40} + \gamma News_{i,-40} + \delta SocialMedia_{i,-30} + \rho ASX_{i,-30} + \theta_1 AnUp_{i,-30} + \theta_1 AnDown_{i,-30} + \vartheta X + \mu Y + \varepsilon_{i,t}. \quad (29)$$

The dependent variable *Runup*$_{i,-30}$ is the target cumulative abnormal return (CAR) starting 30 trading days before until one trading day before the bid announcement for firm *i*. Our five key independent variables are: *FullMedia*$_{i,-40}$, which is one if there was at least one merger signal in the news or news alert plus random talks on firm *i* on the investment platform HotCopper in the two month period before the bid announcement and zero otherwise; *News*$_{i,-40}$, which is one if there was at least one merger signal in the news or news alert about firm *i* as well as no talks on the investment platform HotCopper in the two month period before the bid announcement and zero otherwise; *SocialMedia*$_{i,-30}$, which is one if there was no merger signal in the news or news alert but only talks on the investment platform HotCopper about firm *i* in the 30-trading day period before the bid announcement and zero otherwise; *ASX*$_{i,-30}$ equals one if there was a public ASX announcement about firm *i* in the 30-trading day period before the bid announcement and *AnalystUp*$_{i,-30}$/*AnalystDown*$_{i,-30}$ equals one if there was an upward/downward change in analyst recommendation for firm *i* in the 30-trading day period before the bid announcement and zero otherwise.

We include the following target characteristic control variables[30] in vector *X* of our regression: *LnSize* is the natural logarithm of firm *i*'s market capitalization in the month of the bid announcement, *Market-to-Book* is the market capitalization divided by the book value of equity for firm *i* and *EBITDA* is the firm's operative performance (EBITDA) of the last twelve month before the bid announcement scaled by total assets at the time of the bid announcement.

[30] Palepu (1986) and Comment and Schwert (1995) discuss whether accounting and stock performance measures identify potential target firms and their determinants on target run-ups vary. Comment and Schwert (1995) assert that target firms perform poorly and have higher costs of capital. Palepu (1986) concludes that the probability of acquisitions decreases with size, increases with lower market values or increases with inefficient management teams.

Additionally, we include dummy variables for deal and other characteristics in vector *Y* of our regression model, representing: acquisitions involving toeholds or in other words the acquirer's ownership in the target before the bid announcement (Jarrell and Poulsen, 1989; Betton and Eckbo, 2000; Stulz et al., 1990); takeover values of conglomerates or diversifying firms (Maquieira et al., 1998; Graham et al., 2002; Martynova and Renneboog, 2008); cross-border acquisitions (Porta et al., 1998; Brigida and Madura, 2012); hostile and friendly takeovers (Servaes, 1991; Schwert, 2000) or tender offers (Dodd and Ruback, 1977; Schwert, 1996). We further test whether public companies are more experienced with acquisitions resulting in less leakage (lower target run-ups) and also use public takeover (intention to increase ownership from minority to majority stakeholder with more than 50% of the shares) as a proxy for the expectation of efficient management and include these control variables.

We do not control whether the bid was successful or unsuccessful at a later stage of the process for several reasons. First, we are interested in the nature of media attention before a bid is publicly announced. It is, therefore, highly unlikely that investors can predict at that point in time whether a bid will be successful or not. Second, Asquith (1983) shows that market responses on average for both successful and unsuccessful were equal at the time of public announcements. However, this changes as time approaches the outcome date (successful merger versus failure). Thus, the market is expected to be uncertain at the press date whether a bid will be successful or not.

Regression Models 1 to 4 in Table 4-6 show the relation between media attention and target run-ups and form the foundation for our baseline results in regression Model 5 of Table 4-6. The coefficients on social media attention are positive and significant at the 5% level. Target run-ups which only catch social media attention on the investment platform HotCopper experience abnormal returns that are 3.7 percentage points higher compared to the mean run-up of -1.55% for firms with non-media attention. One can discuss the causality, whether social media activity about a firm increases net purchases of its stock and therefore cause target run-ups or whether target run-ups caused by other reasons grab the attention of individual investors on social media. However, taking the results of the previous section in consideration that firms that only receive social media attention do not differ from firms that receive no media attention at all, we argue that individual investors rather buy stocks that receive social media attention.

Table 4-6. Fixed-Effects Regressions on Target Price Run-Ups

This table reports the industry- and year-fixed effect regressions results. Standard errors are firm-clustered and reported in parentheses. The dependent variable is Run-up, the cumulative abnormal returns of the target from 30 days to one day before the bid announcement. The variable definitions are in the Appendix. The constant is not tabulated in this table. ***, **, and * describe significance at 1%, 5% and 10% level, respectively.

	Model 1	*Model 2*	*Model 3*	*Model 4*	*Model 5*	*Model 6*
Media attention proxy variables						
$FullMedia_{-40}$	-0.004			-0.010	0.006	0.007
	(0.020)			(0.020)	(0.022)	(0.022)
$News_{-40}$		-0.087*		-0.091*	-0.076	-0.085
		(0.049)		(0.050)	(0.052)	(0.052)
$SocialMedia_{-30}$			0.037**		0.037**	0.036**
			(0.017)		(0.018)	(0.018)
ASX_{-30}			0.002		-0.005	-0.006
			(0.020)		(0.021)	(0.022)
$AnUp_{-30}$						0.097***
						(0.027)
$AnDown_{-30}$						-0.093***
						(0.029)
Target characteristics						
$LnSize_{i,t}$	-0.000	0.001	-0.001	0.001	0.000	0.001
	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)	(0.005)
$Market\text{-}to\text{-}book_{i,t}$	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$EBITDA_{i,LTM}$	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Deal and other characteristics						
International	0.015	0.016	0.014	0.016	0.015	0.014
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Diversification	0.011	0.011	0.012	0.011	0.011	0.015
	(0.016)	(0.016)	(0.016)	(0.015)	(0.015)	(0.015)
Hostile	0.328	0.326	0.317	0.329	0.317	0.321
	(0.206)	(0.208)	(0.196)	(0.207)	(0.198)	(0.197)
Toehold	-0.020	-0.020	-0.021	-0.020	-0.020	-0.022
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Takeover	0.039***	0.036**	0.040***	0.037**	0.038**	0.037**
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.016)
Tender	0.015	0.015	0.014	0.016	0.015	0.015
	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)
PublicAcq	-0.029*	-0.028*	-0.029*	-0.028*	-0.029*	-0.028*
	(0.016)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Observations	1,872	1,872	1,872	1,872	1,872	1,872
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	1.97%	2.15%	2.25%	2.11%	2.30%	2.95%

This is in line with the findings by Huberman and Regev (2001). They find that public attention impacts on share prices, even though no genuinely new information was disseminated. We transfer this finding to our economic reasoning that for structurally similar companies, public attention raises interest in a firm's stock and therefore drive target run-ups as observed in our regression analysis. Results of Models 2 and 4, on the other hand, give indications that merger signals for firms with news and no other social media attention are negative and significant at the 10% level compared to firms with non-media attention. Consistent with Womack (1996) and Barber et al. (2001), who discuss the immediate impact of (un-)favorable changes of analyst recommendations at the date of its release for institutional investors, we include the indicator variables for positive or negative changes in analyst recommendations in the run-up period in our regression Model 6 in Table 4-6. It is interesting that analyst upgrades and downgrades are both positive and negative and significant at the 1% level, while the coefficient for social media attention ($SocialMedia_{-30}$) remains positive and significant at the 5% level. Also, both coefficients are almost at the same level with +0.097 for positive and -0.093 for negative changes in analyst recommendations, implying that investors do not put more weight on positive or negative analyst information. This consequently supports our hypothesis that analyst recommendations are an important information source for institutional investors who are able to buy and sell stocks with only low limitations on short selling. Since analyst reports are usually costly to source for individual investors, they are expected to impact firms that are larger and thus have more analyst coverage. Individual investors, on the other hand, look for other low-cost information sources and therefore increase their activity such as in internet investment platforms. Most strikingly in our results is the finding that merger signals on traditional media have different relations to target run-ups depending on the degree of total media coverage. Especially firms which only catch traditional news attention tend to experience negative run-ups. However, analyst recommendations appear to be the more important source of information for institutional investors.

The coefficients on the other control variables are mostly consistent with findings from the prior literature. We find no evidence that target characteristics of a firm could help to predict mergers or acquisitions. The coefficients for firm size (natural logarithm of market capitalization), market valuation (market-to-book) and operational performance (EBITDA of last twelve months scaled by total assets) are all low and insignificant. This is consistent with the findings of Palepu (1986) who concludes that financial models based on accounting and stock price information do not predict targets accurately.

We find no significant coefficient for hostile takeovers which is consistent with the findings from Schwert (2000), who asserts that friendly and hostile takeovers do not differ significantly. Consistent with the findings from Maquieira et al. (1998), we do not find significant economic differences for diversifying firms before bid announcements. Also, cross-border related bid announcements do not significantly relate to target run-ups, yet with a positive coefficient of 0.015. Porta et al. (1998) corroborate that investor protection and anti-insider trading laws are less pronounced than in the U.S. Similarly, Australia is also subject to strict anti-insider laws by the Corporations Act 2001. It is, thus, expected that information leakage could be more prominent in international deals and thus reinforce target run-ups. Most strikingly, we find that the coefficient for takeover expectations (when minority or non-shareholders intend to become a majority shareholder) of +0.0037 is significant at the 5% level. This finding supports the inefficient management theory, which suggests that the market anticipates a change in management prior to an official bid announcement. Furthermore, we find a negative and significant coefficient of -0.028 if the acquirer is a public firm. This could be explained by the fact that public firms are subject to stricter anti-insider trading laws. Shareholders of a public firm would, therefore, have fewer inclinations to disseminate price sensitive information before the official bid announcement illegally.

In sum, the results of our main Model 6 presented in Table 4-6 suggest that firms only receiving social media attention before a bid announcement experience significantly higher target run-ups compared to similar firms without media attention. Furthermore, analyst recommendations appear to be a significant informational source for institutional investors before bid announcements.

4.5. Net Buyers or Sellers for Attention-Grabbing Events

We test the hypothesis whether stockholders who post actively on social media before bid announcements act as net purchasers or sellers. Barber and Odean (2008) describe the attention hypothesis as a positive buy-sell imbalance for attention-seeking events. Thus, net purchases

Table 4-7. Regressions on Target Run-Ups including Ownership Disclosure

This table reports the industry- and year-fixed effect regressions results. Standard errors are firm-clustered and reported in parentheses. The dependent variable is $Runup_{-30}$, the cumulative abnormal returns of the target from 30 days to one day before the bid announcement. The constant of the regression parameters is not tabulated in this table. ***, **, and * describe significance at 1%, 5% and 10% level, respectively.

	Model 1	Model 2	Model 3
$FullMedia_{-40}$	0.007	0.012	0.016
	(0.022)	(0.022)	(0.022)
$News_{-40}$	-0.085*	-0.089*	-0.088*
	(0.052)	(0.052)	(0.052)
$SocialMedia_{-30}$	0.037*	0.045**	0.056**
	(0.021)	(0.020)	(0.026)
ASX_{-30}	-0.006	-0.003	-0.003
	(0.022)	(0.022)	(0.022)
$OwnerHigh_{-30}$	-0.003		0.015
	(0.017)		(0.024)
$OwnerLow_{-30}$		0.015	0.024
		(0.016)	(0.023)
$SocialMedia_{-30}$ x $OwnerHigh_{-30}$	-0.004		-0.022
	(0.038)		(0.041)
$SocialMedia_{-30}$ x $OwnerLow_{-30}$		-0.011	-0.021
		(0.061)	(0.064)
$AnUp_{-30}$	0.097***	0.097***	0.097***
	(0.027)	(0.027)	(0.027)
$AnDown_{-30}$	-0.093***	-0.092***	-0.092***
	(0.029)	(0.029)	(0.029)
Target characteristics	*Yes*	*Yes*	*Yes*
Deal characteristics	*Yes*	*Yes*	*Yes*
Observations	1,872	1,872	1,872
Industry FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Adjusted R^2	2.84%	2.88%	2.79%

would positively affect stock prices during such events. Merton (1987) claims that individual investors only hold a limited number of stocks in their portfolios compared to institutional investors. Individual investors only follow a few numbers of stocks based on scarce resources used to gather information. Hence, investors would only buy or sell the stocks they follow. On the other hand, individual investors would not buy a stock because these stocks have caught their attention. Individual investors would rather sell stocks they own (Barber and Odean, 2008). To clarify the relation between stock ownership of social media posters and target run-ups, we introduce two new variables, $OwnerHigh_{-30}$ and $OwnerLow_{-30}$, in our regression model and add two interaction terms to test the interactive relationship between ownership and social media attention with target run-ups. We classify our regression sample into four quartiles according to the share of social media posters who own the stock. We can test this because the investment platform HotCopper allows its users to disclose whether they own the stock or not. $OwnerHigh_{-30}$ / $Ownerlow_{-30}$ is an indicator variable equaling one for the highest / lowest quartile. We interact both variables with our social media indicator variable.

Results of our regression are shown in Table 4-7. Overall, our results remain robust. Model 3 suggests that firms who experience social media attention yield 5.6% higher cumulative abnormal returns before the bid announcement compared to similar firms who receive no media attention. Overall, we find a negative/positive relation for $OwnerHigh_{-30}$ / $OwnerLow_{-30}$ (including the interactive terms) which indicates that stock owners who post before bid announcements would sell/buy around attention-grabbing events. The results, however, are not significant. Nevertheless, these findings support the idea that individual investors who do not own the stock are not able to short sell and rather buy a stock with attention as well that individual investors who own the stock will rather sell the stock they already own. Altogether, the results give weak evidence for a positive net-buy-sell balance as we consider the positive and larger coefficient for social media attention.

4.6. Search Activity of Unsophisticated Individual Investors

Prior literature on investor attention makes a clear distinction between institutional and individual/retail investors (e.g., Barber and Odean, 2008). Institutional investors are defined by nature as sophisticated investors, who own dedicated resources for research and analysis. Individual investors, on the other hand, act very differently in their research, analysis and trading behavior. We, therefore, explicitly distinguish in our paper between sophisticated and unsophisticated individual investors. The first group dedicates a considerable share of their spare time to financial research, analysis, and discussions to derive trading decisions from that information. The second group, however, acts less rational in our definition, spends less time on financial analysis and would follow traditional media outlets with easy to access information.

Traditional literature on investor attention analyzed different proxies for investor attention, such as significant abnormal returns (Barber et al., 2001), trading volume (Barber and Odean, 2008; Gervais et al., 2001), or news media (Barber and Odean, 2008). However, market-based measures are prone to the argument of being the outcome of several different economic forces other than investor attention (a similar problem as described by Da et al. (2015) in relation with investor sentiment).

Lately, a new strand of literature around internet search queries with Google as a direct measure for investor attention has evolved (see Da et al., 2011; Vlastakis and Markellos, 2012; Vozlyublennaia, 2014). Da et al. (2011) claim in their seminal work that this measure “captures investor attention in a timelier fashion”, p. 1461. All of these studies generally relate internet search queries with individual investor attention. Finding the right measure of attention remains a substantial issue in answering the question of whether investor attention contributes to increasing (higher informativeness of prices) or decreasing (noise trading) market efficiency (Vozlyublennaia, 2014). Da et al. (2011) suggest that financial internet search queries are related to less sophisticated investors. However, alike from the best of our knowledge, all other studies they missed to clearly distinguish between unsophisticated and sophisticated individual investor attention in their empirical research.

We, therefore, introduce Google internet search queries as a direct measure for unsophisticated individual investor attention in our regression model. As a reminder, the direct measure *SocialMedia* describes firms which only experience sophisticated investor attention on the investment platform HotCopper. Vozlyublennaia (2014) also argues in their study that it is unlikely that sophisticated or professional investors would search for stock tickers on Google because their trading systems provide sufficient news data and information about the respective assets, already.

Altogether, our regression model now controls for indirect and direct measures of investor attention for institutional (news media and changes in analyst recommendations), sophisticated (investment platform HotCopper) and unsophisticated (Google search queries) individual investors. Motivated by prior literature, we examine the Google search queries for all individual stock tickers (e.g., “ASX MRE” for Minara Resources Ltd.) in our bid announcement sample. Google Trend returns the relative search volume in a search period scaled by the maximum number of searches. Based on Da et al. (2011), Da et al. (2015) and Siganos (2013), we first calculate the daily log differences in search volume and the abnormal search volume:

$$Google1_{i,t} = \ln(1 + Google_{i,t}) - ln(1 + Google_{i,t-1}) \tag{30}$$

and

$$Google2_{i,t} = \ln(1 + Google_{i,t}) - \ln[Median((1 + Google_{i,t-41}), \ldots, \ldots, \ln(1 + Google_{i,t-60})] \tag{31}$$

where $Google_{i,t}$ is the relative Google search query score on day *t* for firm *i* that we also adjust to a range between 1 and 2 for analysis purposes following Siganos (2013). $Google1_{i,t}$ describes the daily change in Google search volume and $Google2_{i,t}$ represents the abnormal Google search volume calculated as the deviation from the median Google search activity from -41 to – 60 days before the bid announcement. This calendar period (-40 to -1 days relative to the bid announcement) approximately matches the relevant target run-up period of 30 trading days before the bid announcement.

Table 4-8 reports the descriptives of the Google variables. Both measures show a positive skewness (0.0070 and 2.0637, respectively) and distinctive peaks (3.0110 and 4.3771), implying a non-normal distribution. Figure 4-3 and Figure 4-4 show the average changes and abnormal Google search inquiries forty days until one day before the bid announcement. The figures only show a slightly increasing trend of search volume towards the day of the bid announcement. We adapt the method by Siganos (2013) and define each merger signal ($Outlier_{i,t}$) as follows:

$$Outlier_{i,t} > Q3_i + 1.5 * (Q3_i - Q1_i) \tag{32}$$

where $Q3_i$ and $Q1_i$ are the upper or lower quartiles for target firm *i* over the period of -40 to -1 days before the bid announcement.

Da et al. (2011) point out in their study that internet users commonly get their information via internet search engines. Google dominates the Australian search engine market with a market share of 94% and thus presents most of the households' internet search queries in Australia.[31] We now repeat our main regression analysis and include the direct measure for unsophisticated

[31] http://gs.statcounter.com/search-engine-market-share/all/australia/2016, November 16th, 2017.

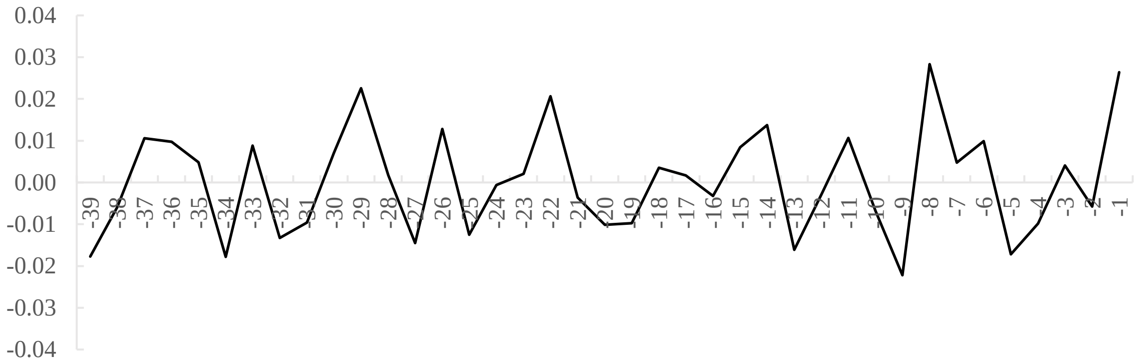

Figure 4-3. Google Ticker Search Volume around M&A Announcements

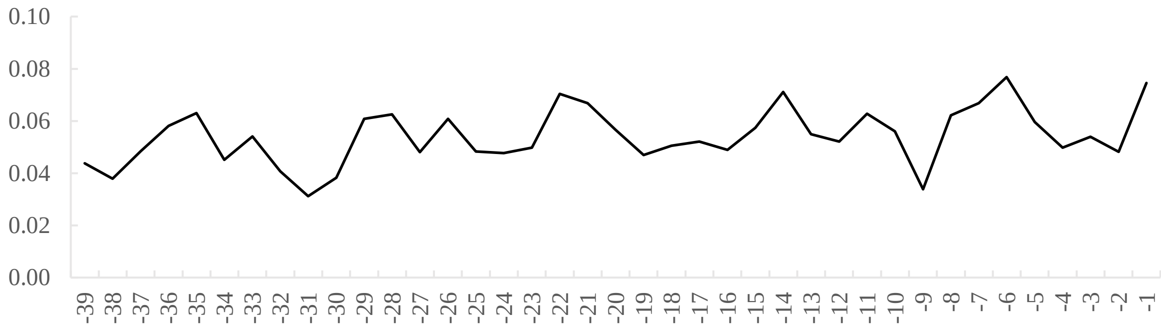

Figure 4-4. Abnormal Google Ticker Search Volume around M&A Announcements

Table 4-8. Google Descriptives

	Google1$_i$	**Google2$_i$**
Average	0.0008	0.0558
Median	0.0000	0.0000
Minimum	-0.6931	-0.4415
Maximum	0.6931	0.6931
Standard dev.	0.2413	0.1856
Skewness	0.0070	2.0637
Kurtosis	3.0110	4.3771

individual investors $Google1_{i,t}$ and alternatively $Google2_{i,t}$ in our model. Furthermore, we want to examine how Google search inquiries interactively relate to news and social media coverage. Thus, we include three interaction terms in the model. The results of this regression are presented in Table 4-9. Overall, our model remains robust. The coefficients for social media coverage and changes in analyst recommendations remain significant at the 5% and 1% level, respectively.

Table 4-9. Fixed-Effect Regression on Target Price Run-Ups – Google Merger Signals

This table reports the industry- and year-fixed effect regressions results. Standard errors are firm-clustered and reported in parentheses. The dependent variable is Runup-30, the cumulative abnormal returns of the target from 30 days to one day before the bid announcement. The constant of the regression parameters is not tabulated in this table. ***, **, and * describe significance at 1%, 5% and 10% level, respectively.

	Model 1	*Model 2*	*Model 3*	*Model 4*	*Model 5*	*Model 6*	*Model 7*
FullMedia$_{-40}$			0.007	0.007	0.025	0.008	0.029
			(0.022)	(0.022)	(0.021)	(0.021)	(0.021)
News$_{-40}$			-0.071	-0.071	-0.028	-0.081	-0.028
			(0.051)	(0.051)	(0.059)	(0.051)	(0.060)
SocialMedia$_{-30}$			0.037**	0.037**	0.038**	0.036*	0.037**
			(0.018)	(0.018)	(0.019)	(0.018)	(0.019)
ASX$_{-30}$			-0.008	-0.008	-0.003	-0.010	-0.003
			(0.022)	(0.022)	(0.023)	(0.022)	(0.023)
AnUp$_{-30}$						0.101***	0.109***
						(0.027)	(0.028)
AnDown$_{-30}$						-0.090***	-0.094***
						(0.029)	(0.029)
Google$_{1\text{-}30}$	-0.041**		-0.038*		0.006	-0.039*	0.015
	(0.021)		(0.021)		(0.039)	(0.021)	(0.041)
Google$_{2\text{-}30}$		-0.042**		-0.039*			
		(0.021)		(0.021)			
Google$_{1\text{-}30}$ x FullMedia$_{-40}$					-0.100*		-0.118**
					(0.059)		(0.060)
Google$_{1\text{-}30}$ x News$_{-40}$					-0.120		-0.150
					(0.099)		(0.098)
Google1$_{-30}$ x SocMedia$_{-30}$					-0.027		-0.036
					(0.044)		(0.045)
Target characteristics	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Deal characteristics	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Observations	1,872	1,872	1,872	1,872	1,872	1,872	1,872
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R-squared	2.19%	2.20%	2.43%	2.44%	2.52%	3.08%	3.27%

Additionally, models 1 to 4 yield negative and significant coefficients of around -0.04 for the distinctive Google attention measures. The results of our final *Model 7* suggest that increased Google search volume of unsophisticated individual investors reduces the generally positive impact of news and social media coverage. All of the attention related interactive terms are negative, but only significant for *Google1$_{i,-30}$ x FullMedia$_{i,-30.}$* The interactive term is only significant at the 5%-level for target firms with full media coverage. These stocks are generally mid to large size stocks in our sample. This is at some point consistent with the results from

Vozlyublennaia (2014), who find that (unsophisticated) individual investors pay more attention to indexes of larger rather than of small stocks. Also, Vozlyublennaia (2014) finds support that individual investors create positive or negative price pressure, depending on the information uncovered by the attention. The effect, however, is not long-lasting and speaks for a noise trading behavior.

4.7. Short-Term Target Markups around Bid Announcements

A substantial amount of literature examined the short-term stock reactions on bid announcements. While there is a general consensus on limited reactions on bidder returns, findings on the magnitude of target stock return reaction differ. Dodd (1980) finds that US targets experience a cumulative positive bid announcement effect of 12.2% one day before until one day after the bid announcement. For the same time period, the results by Asquith (1983) indicate a cumulative markup of around 7% for successful and unsuccessful bid announcements. In another study, Jarrell and Poulsen (1989) find for 172 tender offers a target share reaction of +20.5% in the comparable time period.[32] Even though our study is different from that conducted by Clarkson et al. (2006), who elaborated on the impact of merger rumors in Australia, we find results which are similar in their magnitude. For example, we find cumulative markups of 3% one day before until one day after the bid announcement whereas Clarkson et al. (2006) suggest in their study that firm's returns react by 5.32% the day before until the day of the release of a merger rumor.

We conduct a regression analysis with our main and control variables on the markups or also called unanticipated premiums within the time period of [-1, +1], [-1, +5] and [-1, +10]. The results are reported in Table 4-10. While social media attention significantly increases target pre-bid run-ups, target markups are significantly reduced. All investor attention related media and social media variables are significantly and negatively correlated with short-term markups around bid announcements for each markup period.

---

[32] Refer to Martynova and Renneboog (2008) for a review of literature on M&As and the short-term wealth effects around bid announcements.

Table 4-10. Target Markups around Bid Announcements

This table reports the industry- and year-fixed effect regressions results. Standard errors are firm-clustered and reported in parentheses. The dependent variable is *Markup*, the cumulative abnormal returns of the target from one day before to one, five and ten days after the bid announcement. The constant of the regression parameters is not tabulated in this table. ***, **, and * describe significance at 1%, 5% and 10% level, respectively.

	Model 1	*Model 2*	*Model 3*	*Model 4*	*Model 5*	*Model 6*
	[-1,+1]	[-1,+1]	[-1,+5]	[-1,+5]	[-1,+10]	[-1,+10]
$FullMedia_{-40}$	-0.015*	-0.021**	-0.018*	-0.028**	-0.020	-0.027*
	(0.008)	(0.009)	(0.011)	(0.012)	(0.013)	(0.014)
$News_{-40}$	-0.038**	-0.058**	-0.042	-0.071**	-0.036	-0.077**
	(0.018)	(0.024)	(0.026)	(0.031)	(0.032)	(0.033)
$SocialMedia_{-30}$	-0.014**	-0.016**	-0.022**	-0.023**	-0.025*	-0.027**
	(0.007)	(0.007)	(0.010)	(0.010)	(0.013)	(0.013)
ASX_{-30}	-0.015*	-0.020**	-0.013	-0.018	-0.011	-0.019
	(0.008)	(0.009)	(0.011)	(0.011)	(0.014)	(0.015)
$AnUp_{-30}$		-0.000		0.023*		0.027
		(0.010)		(0.014)		(0.017)
$AnDown_{-30}$		0.002		-0.012		-0.023
		(0.010)		(0.014)		(0.017)
$Google_{1-30}$		-0.035**		-0.044**		-0.070***
		(0.016)		(0.020)		(0.024)
$Google_{1-30}$ x $FullMedia_{-40}$		0.051***		0.063**		0.073**
		(0.020)		(0.025)		(0.030)
$Google_{1-30}$ x $News_{-40}$		0.066*		0.080		0.122*
		(0.035)		(0.052)		(0.063)
$Google_{1-30}$ x $SocMedia_{-30}$		0.033*		0.025		0.054*
		(0.018)		(0.023)		(0.028)
$RunUp_{-30}$	0.036**	0.037**	0.067**	0.067**	0.095***	0.093***
	(0.015)	(0.015)	(0.027)	(0.028)	(0.036)	(0.036)
Target characteristics	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Deal characteristics	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Observations	1,872	1,872	1,872	1,872	1,872	1,872
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	3.18%	3.21%	1.91%	2.05%	1.89%	2.10%

The effect is the strongest for firms which are only mentioned in traditional media compared to those experiencing no media attention (coefficient of -0.077 and significant at the 5%-level). Also, the effect is stronger for firms which have merger signals based on internet search queries before the bid announcement. Altogether, all four media variable coefficients are negative and significant, though at different significant levels and in their magnitude. Even though all interactions are positive and significant, the overall impact remains negative. Our regressions results reveal that target run-ups are positively and significantly related to all markup periods.

The impact strengthens and becomes more significant with longer markup periods ranging from +0.036 to +0.095. This result supports the markup pricing theory rather than the substitution hypothesis proposed by Schwert (1996), who finds evidence that observable target price run-ups replace markups around bid announcements. The markup pricing theory, on the other hand, suggests that target run-ups would increase markup pricing because bidders and targets revise their valuation as other bidders might be acquiring open shares in the market. Each increase in target run-up is thus added to the final deal price. Another possible explanation for the positive correlation of run-ups and markups is the hubris hypothesis stated by Roll (1986). In this hypothesis, the bidder is willing to pay for the takeover irrespectively of the underlying costs. As a result, bidder prices would drop subsequently after the bid announcement since the market would anticipate the market overvaluation.

Irrespective of undervaluation by targets and bidders or overestimation of the bidders, all media related investor attention proxy variables are negative and significant. One possible explanation is that investor attention reveals new information which is immediately incorporated into stock prices (in line with the efficient market hypothesis). Investors would, therefore, not be entirely surprised by the day of the bid announcement. Another reason could be that investor attention before bid announcements reduces uncertainty so that hubris bidder activity is reduced. For both cases, investor attention would contribute to increasing market efficiency before bid announcements.

4.8. Discussion on Investor Behavior around Bid Announcements

So far, the reader may be concerned that social media, traditional media, and internet search queries are endogenous, or to be more specific are prone to the reverse causality issue. For example, users on internet investment platforms learn in news articles about an upcoming merger and then search for further information on the internet. We have examined separate cases or subsamples according to their degree of total media coverage. Since Australia is also a

Table 4-11. Timing of First Media / Internet Attention

This table reports the number of firms receiving merger attention before bid announcements. Average and median days describe the first day of the respective merger signal before the bid announcement.

	Google1$_i$	***Google2$_i$***	***NewsSignal$_i$***	***NewsAtt$_i$***	***HCAtt***
With merger attention	434	430	483	753	1552
Without merger attention	2,331	2,335	2,282	2,012	1,213
Total number of events	2,765	2,765	2,765	2,765	2,765
Ratio covered from total	15.70%	15.55%	17.47%	27.23%	56.13%
Average days	-29	-29	-26	-34	-46
Median days	-34	-34	-24	-40	-53

highly regulated market in which price sensitive announcements are quickly released, we additionally controlled for public ASX announcements in our analysis. In our hypothesis, sophisticated individual investors tend to share and express their opinions on dedicated internet investment platforms, whilst unsophisticated individual investors instead start internet search queries (Da et al., 2011). Institutional investors, however, tend to extract their information from analyst opinions. Since it is unlikely that analysts build their costly reports based on internet investment platforms and that individuals have access to expensive analysis reports, both media attention instruments should have a lower probability on reverse causality. Our results strongly support our belief that smaller and underperforming firms which catch individual investor's attention before bid announcements experience net purchase balances and thus increasing stock returns. One might argue that especially internet search queries are endogenous to macroeconomic or firm-specific events. Google, as one of the main search engines, dominated the market for internet search queries in Australia with a market share of 94%[33] in 2016. Most Australians, including unsophisticated individual investors, release search inquiries via this channel. To address the reverse causality issue of our media attention measures, we analyze the timing effects of the first merger signal or activity within each measure. Table 4-11 reports the average and median days of the first merger signal or firm attention before the bid announcement. Most interestingly, more than half of the potential merger targets caught attention on the internet investment platform. 27% of the firms in our sample were covered by traditional newspapers

[33] http://gs.statcounter.com/search-engine-market-share/all/australia/2016, November 16th, 2017.

Table 4-12. Comparison of Attention Across Media and Internet Outlets

This table reports the number of agreement and disagreement on merger signals for pairwise comparison of our different investor attention measures.

Agree/Disagree [in #]	***Google1$_i$***	***Google2$_i$***	***NewsSig$_i$***	***NewsAtt$_i$***	***HCAtt$_i$***	***Total***
Google1$_i$	-	428/6	169/265	255/179	255/179	434
Google2$_i$	428/2	-	166/264	251/179	253/177	430
NewsSignal$_i$	169/314	166/317	-	483/0	358/125	483
NewsAtt$_i$	255/498	251/502	483/270	-	534/219	753
HCAtt$_i$	255/1297	253/1299	358/1194	534/1018	-	1552

Agreement of total [in %]	***Google1$_i$***	***Google2$_i$***	***NewsSig$_i$***	***NewsAtt$_i$***	***HCAtt$_i$***	***Total***
Google1$_i$	-	99%	39%	59%	59%	434
Google2$_i$	100%	-	39%	58%	59%	430
NewsSig$_i$	35%	34%	-	100%	74%	483
NewsAtt$_i$	34%	33%	64%	-	71%	753
HCAtt$_i$	16%	16%	23%	34%	-	1552

before the bid announcement. 17% of these firms were explicitly covered in a merger context. Finally, only 16% of the sample received significant attention based on specific internet searches. Most strikingly, first investor attention signals on the internet investment platform HotCopper leads other investor attention measures with on average 46 days before the bid announcement. General news attention, which we define as a random firm coverage on the press, follows as the second fastest attention channel for investors with on average 34 days before the bid announcement. Explicit news signals, however, appear on average 26 days before the bid announcement. Both Google measures show on average 29 days before the bid announcement first signals of an upcoming merger. These results speak for our hypothesis that sophisticated investors are able to identify potential targets. Not only do investors talk in more of half of the cases about the target firm, but also identify these companies at an earlier stage.

The results reported in Table 4-11 only give information on average and median days of first merger signals irrespectively, whether merger signals appeared on the same channel or only in individual media outlets. Table 4-12, therefore, additionally reports the agreement in attention (both outlets show merger signals) and disagreement of each media and internet measure. *Google1$_i$* / *NewsSig$_i$* (169/265) describes that out of 434 Google merger signals, 169 merger signals were also identified via traditional news. Nearly 60% of merger signals induced by internet search queries were also covered by news outlets and the investment platform HotCopper. On the other hand, only 16% of firms which caught attention on the investment platform HotCopper also experienced Google merger signals. This supports our hypothesis that unsophisticated and sophisticated individual investors use other instruments to identify the firm of interest.

4.9. Conclusion

A large number of studies observed the phenomenon of target price run-ups before the actual first bid announcement. Researchers, in general, find consensus on the rationality of markup pricing behavior on the day of the bid announcement. However, it is elusive why target firms experience significant price run-ups (Schwert, 1996). Studies evolved around the market expectations and insider trading theory, which try to explain whether market efficiency or illegal exploitation of private information contribute in explaining target run-ups. This study extends the existing literature and connects the market expectations and investor attention theory (Barber and Odean, 2008). We find evidence that investor attention is the prerequisite for market expectations on upcoming merger or acquisition activity.

Previous research relied on indirect measures of investor attention (e.g., news media, trading volume) to examine its relation to firm performances or specific events. New technologies and innovations, however, enable investors to participate and gather information in real-time and efficiently change their investment decisions. This also allows researchers to create direct measures of investor attention, such as posting activity on internet investment platforms or active internet search queries. Our study provides evidence that institutional and (un-)sophisticated individual investors use preferred channels to gather and disseminate information before bid announcements.

We find that smaller and underperforming stocks that only capture the attention of sophisticated individual investors on HotCopper experience a significantly stronger target run-up before bid announcements. Firms with similar fundamental characteristics without (social) media attention do not experience a significant run-up but rather a short-term announcement effect.

Large firms, on the other hand, are especially sensitive to analyst opinions. Positive and negative analyst upgrades have significant influences on target run-ups in the respective directions. Merger signals in traditional media appear only to play a minor role for institutional investors. Also, Google search activity before bid announcements of unsophisticated investors does not significantly or economically explain target run-ups.

In summary, we find distinctive drivers of institutional and individual investor attention which explain target run-ups consistent with the market expectations hypothesis. We show that media attention contributes to target run-ups before merger attention. Secondly, our results may be valuable to practitioners who aim to anticipate run-ups in their deal process to reduce their target premiums.

5. A News Sentiment Risk Factor in the Mass Media Zoo

ABSTRACT: We study the tone of news media and the cross-section of stock returns. Our results provide evidence that an equally-weighted long-short portfolio of stocks sorted by the tone of the news media coverage earns significant returns of 7.5% per year even after controlling for market, size, book-to-market, momentum, liquidity, profitability, and investment factors. Separating the effect of positive and negative media tones reveals that results are mainly driven by positive media tone which we refer to as a "premium on optimism".

5.1. Introduction

The digitalization of news media content away from traditional newspaper outlets is an inexorable trend which also affects the informational finance environment. It influences how information is disseminated in speed, quality, and quantity (e.g., Tetlock, 2007; Engelberg and Parsons, 2011; Hillert et al., 2014) . In the course of this development, news media tones about a firm's latest performances and seemingly unforeseeable events gain importance in the contemplation of stock markets. A strand of literature highlights how news media coverage influences investor behavior (e.g., Tetlock, 2007; Fang and Peress, 2009; Engelberg and Parsons, 2011). It, therefore, suggests itself that the underlying news media tone conclusively provide directional indications for future stock market movements. However, existing literature offers ambivalent implications on the role of news media tone in financial markets. Some findings suggest that market reactions on negative news media tones are more pronounced. For example, Chan (2003) demonstrates in his study that "losers" with news headline coverage experience a negative return drift compared to their size, book-to-market, and event-return-matched peer group, whereas "winners" with good news exhibited less drift. Moreover, only the fraction of negative words in newspaper articles is found to possess predictive power as argued by Tetlock (2007) and Tetlock et al. (2008). Contrarily, García (2013) observes in his study that both, the fractions of positive and negative words in a newspaper article predict the next day's return. In a momentum related study, Hillert et al. (2014) provide support for the hypothesis that investors tend to overreact to winner stocks with high media coverage and positive news media tone.

This study extends the news media related literature by analyzing the cross-sectional relation between news media tone and stock returns as well as shedding light on the short- to long-term behavior of stock returns resulting from news media tone coverage. In this study, we rely on an innovative news database with more than 120 million unique news publications consisting of news articles, news flashes, or press releases covering on average more than 6,800 U.S. firms per year (in the period between 2000 and 2017). Fang and Peress (2009) argue in their paper that it is unlikely that mass media contains new value-relevant information but that mass media supports the broader functional dissemination of news which in the end affects stock returns.

Our results provide evidence that an equally-weighted portfolio of stocks experiencing news media coverage with positive and negative tones (positive/negative news stocks, forth on) earns significant returns of 7.5% per year even after adjusting for widely accepted market risk factors.[34] In particular, an equally-weighted portfolio of positive and negative news stocks earn significantly higher returns, when stocks are smaller, less profitable and exhibit lower momentum. The premium ranges between 8.3% and 17.5% for these subsamples even after adjusting for widely accepted risk factors. Especially the subsample dimension for past losers reveals the most substantial economic significance in our results. Separating the effect of positive and negative media tones, our results suggest that returns are mainly driven by positive media tone which we refer to a "premium on optimism".

The behavioral finance theory provides several explanations for the observed "premium on optimism". First, it could be explained by a noise trading related premium in the spirit of De Long et al. (1990). If positive sentiment is a proxy for mispricing (e.g., due to limits of arbitrage), then rational or Bayesian investors would not trade against their positions in the presence of market frictions.[35] Noise trader's optimism might drive prices upwards away from its actual fundamentals. Rational investors or so-called arbitrageurs (short-)selling this asset must deal with the possibility that noise traders become even more optimistic and therefore continuously drive prices further away. As a result, arbitrageurs would realize a loss and therefore

[34] Adjusting for market, size, book-to-market, momentum, Pastor-Stambaugh liquidity factor, profitability, and investment.
[35] Whether news media tones serve as proxies for investor sentiment is discussed in more detail in the next section.

rather limit their willingness to open the originally intended short position. De Long et al. (1990) refer to this short-term deviation of asset prices as “noise trader risk”, in our context the premium on optimism. In the style of the attention theory by Barber and Odean (2008), investors rather only trade on stocks that they own. Conclusively, investors would rather buy stocks which they do not own based on the underlying sentiment than selling stocks which are not part of their private portfolio. The “impediments-to-trade” hypothesis stated by Fang and Peress (2009) signifies that market frictions prevent rational arbitrageurs to exploit mispricing. Negative sentiment can from this standpoint be less pronounced for stock prices because investors would rather be reluctant to open short positions according to the limits of arbitrage theory. Another theory, namely the prospect theory by Kahneman and Tversky (2013), alternatively explains the limited overreaction to negative media tones with the behavioral bias that investors are reluctant to realize losses. If negative media tones reflect past investor sentiment, one should expect low returns in the short-term but high returns in the longer time horizon. In this context, negative media tone, therefore, acts as a proxy for “loser stocks”, where losses should have been already realized in the past. In summary, the premium on optimism might, therefore, subsume a variety of elements commonly discussed in the behavioral finance literature.

The remainder of the paper is structured as follows. Section 5.2 discusses the theoretical background of news media tone and investor sentiment. Section 5.3 gives an overview of related literature. Section 5.4 describes the data used for this paper. Section 5.5 and 5.6 contain analysis and results on media tones and related market or firm returns. Section 5.7 give possible explanations for the results. Time variation of news sentiment loadings and the risk factor characteristic of news sentiment are discussed in Section 5.8. Section 5.9 concludes the results.

5.2. News Media Tone and Investor Sentiment

The traditional literature view on security returns that financial markets efficiently and rationally price assets with publicly accessible information has faced a versatile challenge from a broad body of behavioral literature. Some of the most commonly discussed market anomalies are, for example, event-based return predictability, short-term momentum, long-term return reversals, high volatility of stock returns in deviation to its fundamentals or post-return drifts as a result from earnings announcements (Daniel et al., 1998). The underlying drivers behind the aforementioned anomalies and thus the determinants of predictability, however, remain open to debate. Fama (1998) argues that chance deviations from fundamentals are elements of market efficiency and thus explain short-term appraisals of anomalies. Yet. behavioral research insists on the refutation that return patterns are strong and persistent (Daniel et al., 1998).

One of the first studies to link news media and stock returns were published by Cutler et al. (1989). Unexpectedly, the authors find only small market responses to macroeconomic news in coincidence with important political and world events. Two recent studies provide interesting findings on news sentiment and its link to stock returns. García (2013) studies the fraction of positive and negative words of financial news in the New York Times. He reports that news sentiment predicts the next day's market returns with subsequent reversals. In another study, Hillert et al. (2014) observe significantly stronger momentum effects for stocks with higher news media coverage. The effect is influenced by news sentiment and reverses after 24 months. Both papers, therefore, provide novel findings compared to the fundamental work of Tetlock (2007), who asserts that only media pessimism would be related to future stock returns.

To clarify the underlying theory of news media tones and investor sentiment for our study, we revisit and furthermore extend the hypotheses stated by Tetlock (2007). In his paper, he tests explicitly the hypothesis that negative media tone (media pessimism) is related to low investor sentiment, resulting in declining stock prices. With this regard, it remains ambiguous whether media tones forecast the future or reflect past investor sentiment. If negative media tone serves as a proxy for low past and future investor sentiment, then one should expect low future returns in the short-term and high future returns in the long-term. However, if media pessimism follows past investor sentiment, the long-term returns will be higher than the short-term returns.

Alternatively, when media pessimism reflects negative fundamental information which is not yet fully incorporated into the stock price, then one would still find negative stock reactions in the short-term however with the absence of subsequent return reversals. Lastly, if media pessimism is a measure for stale information which is already fully reflected in the stock price, then one would observe no impact of media pessimism on stock returns.[36]

[36] Please refer to Tetlock (2007), p. 1142 ff, for more details.

We extend the model by Tetlock (2007) by two basic assumptions. The latest research has shown that positive sentiment can be equally important in explaining future stock returns. We, therefore, argue that the same mechanics described above should also apply to positive media tones. Another theory, namely the prospect theory by Kahneman and Tversky (2013), alternatively describes the behavioral bias that investors are reluctant to realize losses. If negative media tone reflects in part past investor sentiment and information about past low returns, one should expect low returns in the short-term but high returns in the longer time horizon. In this context, negative media tone, therefore, acts as a proxy for “loser stocks”, where losses should have been already (partly) realized in the past. Since investors become more risk averse when they face potential losses, they might hold on potential loser portfolio stocks which additionally experience negative news media exposure. Past literature reports that (social) media outlets tend to be more bullish in terms of coverage (e.g., Antweiler and Frank, 2004; Leung and Ton, 2015) and, according to behavioral theory, investors tend to underweight information that does not confirm their private signal (e.g., bullish investors who experience negative media sentiment). This suggests that on average the impact of negative media tone must be lower than for positive media tones. The mechanisms and relationship between media tones, investor sentiment, and stock returns are summarized in Figure 5-1.

In this paper, we thus specifically test the hypothesis that news media tone is related to investor sentiment with asymmetric associations between separate media tones and stock returns. We use the term news sentiment as an equivalent to investor sentiment in our study forth on. One of the most common critics on behavioral theories refer to the limited range of application and that presented theories only apply to specific environments. The acceptance of new models, therefore, relies on the parsimonious character, explanation power for various anomalies in a different context and the generation of empirical implications (Daniel et al., 1998). Extending the simple model by Tetlock (2007), thus allows us to consider a variety of behavioral theories and explain anomalies in a broader context.

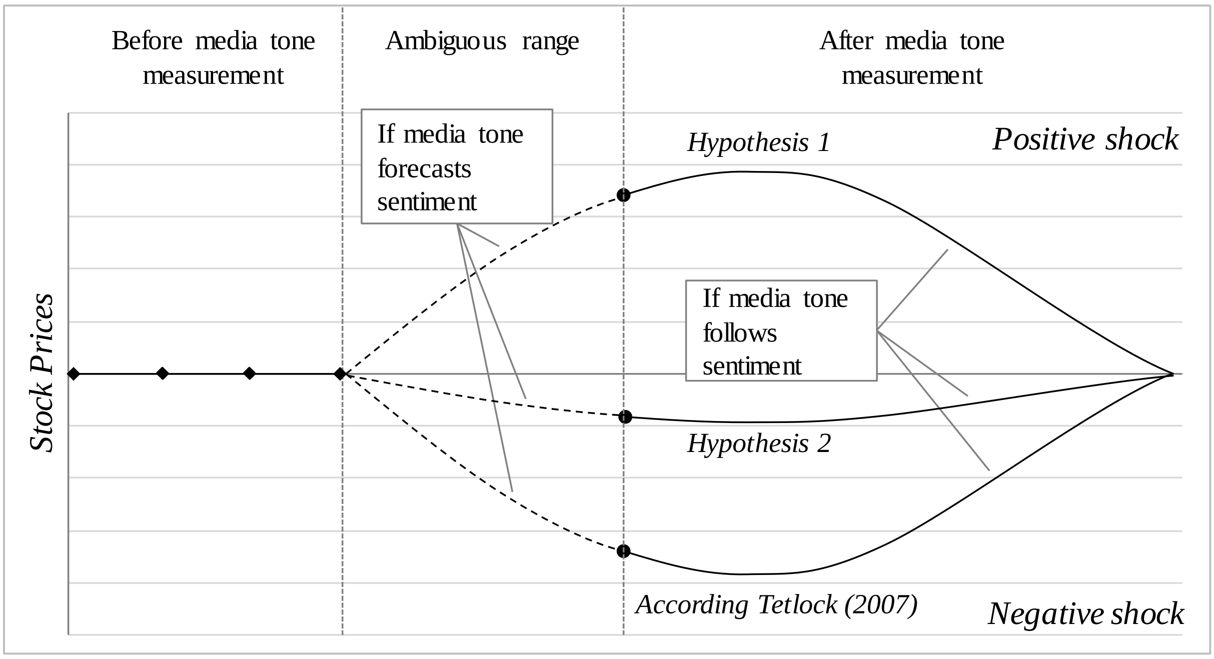

Figure 5-1. Impact of Media Tone on Stock Prices.

5.3. Literature Review

Our paper can be categorized into the strand of literature which deals with the relationship between media sentiment and stock returns and the behavioral research around the cross-sectional pattern of stock returns.

5.3.1. Media Sentiment and Stock Returns

Early studies on the impact of qualitative media content include Cutler et al. (1989), who argue that stock prices are affected by types of news (e.g., war, presidency) at some degree but that all return variations are not subject to qualitative news contents which are not associated with macroeconomic innovations. Thus, large market movements could not be observed on days without significant news events. Contrarily, Chan (2003) concludes in his study that especially "news losers" (stocks experiencing low past returns with coincidence in news headline coverage) exhibit a strong return drift.

In a seminal work by Tetlock (2007), the author examines the qualitative content of the Wall Street Journal news column "Abreast of the Market" by measuring media pessimism. In his major findings, media pessimism predicts downward pressure on stocks with a subsequent reversal in returns. In an extended study, Tetlock et al. (2008) further report that the fraction of negative words in all Wall Street Journal and Dow Jones News Service stories predicts low earnings and implies brief underreaction in stock returns. The authors, therefore, link the qualitative content of news articles with fundamental information incorporated into stock prices. To further investigate on the role of positive media content, García (2013) comprehensively studies the impact of positive and negative media sentiment of two columns of financial news from the New York Times in a broad time horizon along the 20th century. He provides evidence on return predictability for both, positive and negative news sentiment, which is especially pronounced in recession periods. More recently, Hillert et al. (2014) examined how momentum is associated with media coverage and sentiment. Their findings uncover a significant stronger momentum effect for stocks with news coverage in association with investor overreaction. News related momentum and subsequent return reversals are found to be more distinct when the underlying news sentiment matches the formation period return.

In a broader media sentiment context, the literature provides mixed results on return predictability of qualitative media content. The impact of internet stock message boards on US stock returns, for example, is documented to be economically insignificant as stated by Antweiler and Frank (2004). However, they find that posting volume predicts market volatility. In another study on a popular social media investing platform, Chen et al. (2014) document a strong relationship between social media pessimism expressed in articles and commentaries and future stock returns as well as earnings surprises. Extending the advocates on media pessimism as a return predictor, Da et al. (2015) innovatively constructed a “FEARS” sentiment measure based on daily internet search volumes. In particular, their sentiment measure predicts short-term market return reversals and increases in volatility.

Our paper is most of all related but highly distinctive to Tetlock (2007), who focuses on daily short-term reactions of stock returns to media pessimism of news articles only published in one section of the Wall Street Journal. We enumerate a broad variety of news stories in mass news media outlets and include positive and negative news sentiment in our analysis. Furthermore, our study provides new insights on cross-sectional return patterns based on sentiment differentials opposed to the time dimensional focus of Tetlock (2007). The role of news sentiment in a classical asset pricing setting will thus be in the limelight of our study.

5.3.2. Sentiment and the Cross-Section of Stock Returns

Our study is, furthermore, closely linked to the literature around behavioral asset pricing models and the explanation of cross-sectional returns. We do not aim to shed light on the convoluted factor zoo as described by Cochrane (2011), but primarily pursue to provide more clarity on the role of news sentiment in financial markets. Among those hundreds of papers, our study relates the most to the following two sentiment related asset pricing papers.

Baker and Wurgler (2006) proposed in their famous study a composite investor sentiment index based on six different proxies identified in the prior literature.[37] They, for example, document for low beginning-of-period sentiment relatively high subsequent returns for growth stocks, which are typically small, young, more volatile, and unprofitable. In another study, Hirshleifer and Jiang (2010) investigate equity and debt financing to explain cross-sectional return patterns which they refer to as common misvaluations of asset prices. In this context, the authors define a misvaluation factor "as any statistical common factor in stock returns that is substantially correlated with the common mispricing of individual stocks" (Hirshleifer and Jiang, 2010, p. 3042-43). In their zero-investment-portfolio, they go long on firms that repurchase stocks and short firms with new equity issues, resulting in a so-called UMO (undervalued minus overvalued) factor. Their findings suggest that UMO loadings predict cross-sectional returns for both, portfolios and individual firms.

Finally, our paper is closely related to Fang and Peress (2009) who focus on media coverage and its relation to cross-sectional return patterns. Firms with no media coverage are systemically found to generate what they call a "no-media premium". Their study, however, neglects the directional relationship of news content and stock returns. Our analysis in the later section will further provide details on the relevance of news sentiment in the cross-section of returns.

[37] Baker and Wurgler (2006) use six sentiment proxies: discount on closed-end funds, turnover of NYSE shares, number and average of IPO-first-day-returns, share of equity in total new issues, and the dividend premium.

5.4. Data Design and Descriptives

5.4.1. Sample Data

We obtained our monthly trading data (stock returns, market capitalization and trading volume) from the Center for Research in Security Prices (CRSP) and included all stocks from the New York Stock Exchange (NYSE), the American Stock Exchange (NYSE MKT, formerly AMEX), and the National Association of Securities Dealers Automated Quotations (NASDAQ). Since our study does not limit to the universe of large stocks, which are mainly listed on NYSE, we consider all three major stock exchanges for our further analysis which were listed between January 1st, 2000 and December 31st, 2017. To understand the comprehensive news sentiment impact on cross-sectional returns of all stocks, we also include smaller stocks and do not filter our data, for example, due to liquidity concerns (e. g. by price thresholds or size). Our analysis accounts for possible illiquidity drivers, which will be further presented and discussed in Sections 5.6 and 5.7. We have drawn our fundamental data such as book value of equity or operational profitability from the merged Compustat/CRSP database. As a consequence of the financial crisis 2008/09 several self-regulatory organizations (SROs) offer daily and monthly short sale information on individual securities to increase the overall transparency around short sale transactions.[38] To analyze how short sale activity relates to news sentiment, we obtained monthly short sale volume as reported by the Financial Industry Regulatory Authority (FINRA) for stocks traded on NYSE during January 1st, 2010 and December 31st, 2017.

We use the firm-level Composite Sentiment Score (forth on news sentiment) provided by RavenPack News Analytics (RPNA) to create an average monthly news sentiment score for an individual firm. RPNA gains in popularity amongst researchers but also for practitioners. Leading investment banks already use this proprietary dataset to investigate on global macro trends (equity index, FOREX, and sovereign bonds), improve their pair-trading strategies based on abnormal news volume and sentiment, or use CAPEX announcements effects for asset return predictions. RPNA covers more than 180,000 entities in more than 200 countries with its analytics. However, we focus on the US news sentiment coverage in this paper. The RPNA news sources can be divided into three distinct groups: Dow Jones, Web and PR edition. The former

[38] Further information on short sale volume and transactions: https://www.sec.gov/answers/shortsalevolume.htm

examines information from well-established financial news provider including the Dow Jones Newswires, regional editions of the Wall Street Journal, Market Watch and Barron starting from January 1st, 2000. The Web edition uses articles from generic business publishers, local news, blog sites, government and regulatory updates beginning from January 1st, 2007. Lastly, the PR edition adds press releases from newswires, and distribution networks to RPNA analysis covered since January 1st, 2004. A RPNA news story is, furthermore, classified into the categories full article, news flash, press release, and tabular material. Full articles possess both, a headline and a body text, with mostly textual material. News flash articles only have headlines and no body text. Press releases are news initially announced by a firm but distributed via news providers. Tabular material related news has a headline and a body text mostly dominated by tabular data. Table 5-1 presents summary statistics for the unique news stories of our sample. The total number of included news stories in the RPNA database increased significantly from less than 2 million articles in 2006 to almost 4 million articles in 2007 with the inclusion of the Web edition. The RPNA article universe more than tripled subsequently until 2017 with more than 14 million news stories. This might reflect either the increasing coverage of the RPNA algorithm or the actually increasing number of created news articles. We cannot disentangle this effect by now, but we assume that both elements contribute to the increasing number of news stories included. The most dominant news type is "full article" covering more than 80% from all news stories.

To match each news story with the relevant firm entity, RPNA assigns relevance scores between 0 and 100 to measure how strongly the news stories and the firm entities are related. The relevance score considers, for example, keywords, mentions, firm role in the specific event, or the text positioning (headline vs. body) in its calculations. A headline story about a firm results in a relevance score of at least 90 for the respective firm. According to RPNA, sentiment scores are especially applicable in settings with relevance scores higher than 90. We, therefore, followed their recommendation and only included firm-level news stories with relevance scores higher than 90. Other than previous studies on news media, we concentrate on media sentiment expressed via the mass media and not the coverage, or in other words the total number of news articles, itself. More details on the RPNA sentiment score construction are described in the next section.

Table 5-1. Summary Statistics for Total RPNA Stories

This table describes the total number of unique RPNA news stories by each year. Full articles possess both a headline and a body text with mostly textual material. News flash articles only have headlines and no body text. Press releases are news initially announced by a firm but distributed via news providers. Tabular material related news have a headline and a body text mostly dominated by tabular data. The numbers of news stories are denoted in thousands.

Year	Full article		News flash		Press release		Tabular material		Total
	#	% of total	#	% of total		% of total	#	% of total	
2000	435	59%	152	21%	24	3%	129	17%	740
2001	623	57%	269	25%	93	9%	107	10%	1,092
2002	337	40%	268	32%	144	17%	83	10%	832
2003	368	41%	245	28%	174	20%	100	11%	888
2004	750	41%	302	16%	596	33%	183	10%	1,831
2005	660	37%	314	18%	650	36%	171	10%	1,795
2006	590	33%	325	18%	674	38%	175	10%	1,764
2007	2,357	60%	488	12%	713	18%	384	10%	3,943
2008	3,498	68%	605	12%	719	14%	328	6%	5,149
2009	4,370	72%	757	12%	753	12%	225	4%	6,104
2010	4,547	71%	680	11%	989	15%	196	3%	6,412
2011	7,855	78%	877	9%	1,074	11%	256	3%	10,063
2012	10,433	80%	1,100	8%	1,214	9%	326	2%	13,074
2013	10,731	80%	966	7%	1,183	9%	556	4%	13,436
2014	10,428	79%	865	7%	1,360	10%	581	4%	13,234
2015	11,499	83%	612	4%	1,366	10%	371	3%	13,849
2016	11,704	82%	645	5%	1,522	11%	437	3%	14,307
2017	11,821	81%	716	5%	1,733	12%	362	2%	14,632
Total	93,005	76%	10,188	8%	14,982	12%	4,969	4%	123,145

5.4.2. News Sentiment

The news sentiment classified by RPNA relies on three proprietary methodologies, grouped into Traditional, Expert Consensus and Market Response. The traditional method falls back on keyword analysis and assigns stories, according to a rule base (with more than 12,000 phrase/word-level combinations) defined by RPNA, into sentiment clusters (positive, negative, neutral). A news story, typically about global equities and earnings evaluations in the traditional methodology, is then assigned into a positive or negative sentiment cluster. If a news story cannot be assigned to any of these two clusters, it is classified as neutral.

The Expert Consensus methodology rests on a training classification algorithm developed based on the results of manually classified sentiments by financial experts. A large news dataset was provided to financials experts who evaluated the possible positive, negative or neutral impact on asset returns in the subsequent hours. The resulting training sets were then used for the generation of an automated computer classification algorithm. For this methodology, classifiers mainly examined short commentaries and editorials on global equity markets, news about mergers and acquisitions, stories about corporate announcements, and news on earnings releases.

The Market Response methodology examines the short-term impact of news stories on the subsequent stock price volatility in the next two hours. RPNA provides a so-called Composite Sentiment Score (CSS or news sentiment forth on) that combines the three methodologies described before. The direction of the media sentiment (positive / negative) results from the classification strategy of the Traditional and Expert Consensus methodology. The overall strength of the score (range between 0 and 100, where 50 denotes neutral) is determined by the Market Response methodology.[39] The sentiment data provided by RPNA is available on daily firm-level. For our analyses, we aggregated the sentiment data on a monthly basis and weighted each news item by its relevance. We only included news stories with a relevance score higher than 90. One news item can cover more than one firm. Table 5-2 describes the yearly summary statistics of the total media sentiment in our sample. The average news sentiment was exceptionally low in the year after the Dotcom crash in 2001 and in the year of the financial crisis in 2008 with scores of 49.30 and 49.79, respectively. Also, the statistics document a higher standard deviation in news sentiment and thus widespread disagreement on news sentiment in periods of financial downward pressure. The share of neutral news items remained fairly stable during the sample period. However, we find that the share of negative and positive news items developed asymmetrically, with an increasing (decreasing) share of positive (negative) sentiment from 32% (18%) in 2000 to 41% (11%) in 2017. On average, news items in the RPNA database covered 6,858 firms per year during the sample period.

[39] Information are drawn from the RavenPack News Analytics 4.0 User Guide and Service Overview. For further details please contact RavenPack News Analytics. Alternatively, RPNA also provide another sentiment score, called the Event Sentiment Score (ESS). ESS systematically compares news stories with categories classified by financial experts. The strength of the score is based on survey data from industry experts.

Table 5-2. Summary Statistics for Media Sentiment

This table summarizes the statistics for Ravenpack Composite Sentiment Scores (CSS) for our sample firms by year. Sentiment describes the average CSS in the respective year. Sentiment scores exceeding 50 are considered positive, equal 50 are referred as neutral, and below 50 are defined as negative. A Ravenpack story has a unique identifier but can cover more than one firm. The statistics only include news stories with a relevance score higher than 90. Ravenpack categorizes stories into full article, news flash, press release, and tabular material. Firms denote the total number of firms covered by Ravenpack in the respective year.

	Sentiment		# stories and covered firms (in thousands)							
			Negative		Neutral		Positive			
Year	Mean	Std.	#	(% total)	#	(% total)	#	(% total)	# Total	# firms
2000	49.742	5.069	79.9	18.2	219.0	50.03	139.0	31.74	437.8	6,097
2001	49.302	5.846	131.4	21.10	300.8	48.28	190.8	30.63	623.0	5,904
2002	49.354	6.358	98.1	19.38	230.0	45.45	178.0	35.17	506.1	5,791
2003	49.742	5.991	87.3	16.70	239.7	45.88	195.5	37.42	522.5	5,782
2004	50.079	5.130	200.3	17.84	515.5	45.89	407.4	36.27	1,123.3	6,829
2005	50.165	4.826	200.5	16.84	574.9	48.30	414.9	34.86	1,190.4	7,127
2006	50.309	4.915	169.3	14.98	539.6	47.73	421.7	37.30	1,130.7	7,274
2007	50.159	5.404	337.0	17.45	881.3	45.63	713.2	36.92	1,931.6	7,281
2008	49.785	5.703	416.0	19.16	982.7	45.27	772.3	35.57	2,171.0	7,159
2009	49.962	5.204	398.6	15.78	1225.4	48.52	901.7	35.70	2,525.7	6,917
2010	50.455	4.445	345.6	12.95	1303.5	48.83	1020.3	38.22	2,669.5	6,941
2011	50.505	4.370	483.9	12.36	1882.1	48.09	1548.1	39.55	3,914.1	6,917
2012	50.464	4.484	673.2	12.57	2574.4	48.08	2107.3	39.35	5,354.9	7,269
2013	50.645	4.355	631.7	11.78	2488.9	46.39	2244.3	41.83	5,364.9	7,412
2014	50.612	4.308	558.3	11.94	2171.4	46.44	1946.2	41.62	4,675.9	7,089
2015	50.457	4.560	587.4	12.35	2239.1	47.07	1930.6	40.58	4,757.2	7,089
2016	50.389	4.544	584.5	12.55	2247.1	48.24	1826.8	39.22	4,658.4	7,253
2017	50.634	4.283	575.2	11.21	2449.9	47.74	2106.3	41.05	5,131.5	7,316
			6558.4	13.47	23065.5	47.37	19064.5	39.16	48,688.4	

The changing data structure and composition of RPNA may raise the concern that media sentiment expressed via RPNA does not consistently reflect the actual investor sentiment in the same manner in relation to market movements. To alleviate this concern, we examine in the next section how the Composite Sentiment Score relates to US market activities, namely the Standard & Poor’s (S&P) 500 and Russell 3000 index movements.

5.5. RPNA Sentiment and US-Market Indices

Is news sentiment generally an appropriate predictor for market returns? To approach the answer, we first conduct a vector autoregression (VAR) analysis in advance to the asset pricing tests in later sections. Before we explain our VAR results, we first illustrate the correlation of our media sentiment measure and general market movements in Figure 5-2. We construct a value-weighted 90-day moving average of the cumulative change in media sentiment and compare the measure with the S&P 500 and the Russell 3000 market development. It is evident that the news sentiment score captures market movements quite close with indications for predictability of news sentiment, especially before the financial crisis in 2008. The stagnation period in 2014/2015 was accompanied by seemingly negative overreactions of the news sentiment but a quick recovery of sentiment forth on. To formalize the relationship between news sentiment and market returns, we divide our sample period into bearish and bullish time periods according to minimum and maximum levels of market indices and conduct a pairwise correlation test.[40] Table 5-3 describes the correlation between the average aggregated news sentiment and respective market variables, the market index and the monthly market excess return. Overall, the results yield high, positive and significant correlations for the total sample period. However, the table also indicates in line with the findings by García (2013) that results are stronger for bearish periods. More particularly, the correlations are especially pronounced between news sentiment and monthly excess returns for both market indexes, ranging between 0.51 and 0.57.

Furthermore, we perform a VAR analysis to examine the lead-lag relationship between the excess market return and our media sentiment measure according to the following specifications:

$$Mkt - Rf_t = \alpha_1 + \sum_{j=1}^{L} \Theta_{1,j}, Mkt - Rf_{t-j} + \sum_{j=1}^{L} \delta_{1,j}, CSS_{t-j} + \varepsilon_{1t} \tag{33}$$

$$CSS_t = \alpha_2 + \sum_{j=1}^{L} \Theta_{2,j}, Mkt - Rf_{t-j} + \sum_{j=1}^{L} \delta_{2,j}, CSS_{t-j} + \varepsilon_{2t} \tag{34}$$

[40] García (2013) reports higher sensitivity of market returns to news sentiment especially during recessions.

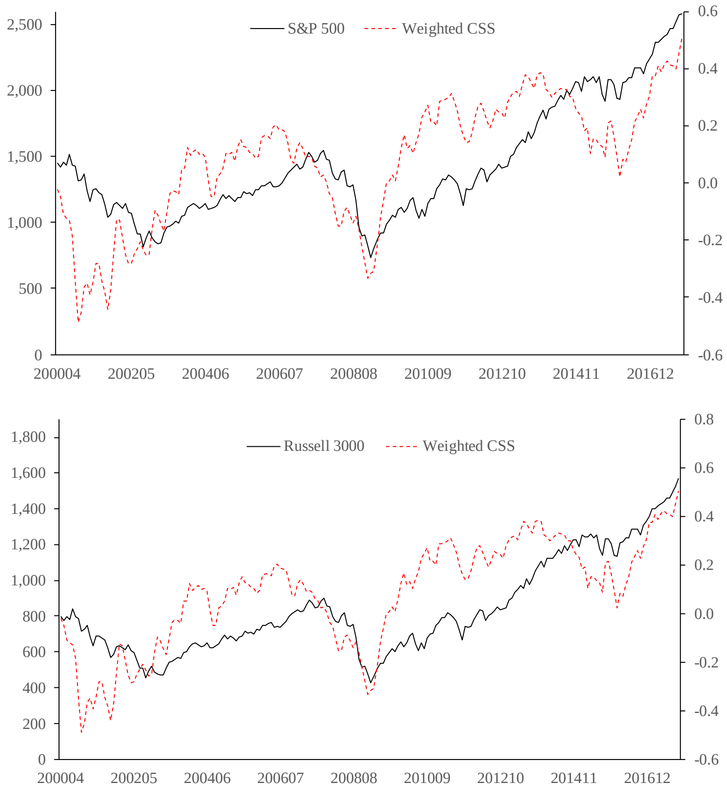

Figure 5-2. Aggregated News Sentiment and Index Developments

where *Mkt-Rf*$_t$ is the excess market return (S&P 500 and Russell 3000) and *CSS*$_t$ is the average aggregated monthly news sentiment score. In order to apply the optimal lead-lag regression specification for our VAR analysis, we first compute the optimal lag length L based on the standard criterion used in the literature. Table 5-4 presents the results of the specification tests which suggest an optimal lag length of 3 for the endogenous variables in our analysis. As we aim to evaluate the economic importance of market reactions following a one-standard deviation shock to news sentiment, we normalize our news sentiment measure with zero-means and unit-variance for our VAR analysis. The results are shown in Table 5-5.

The null hypothesis that the three lags of aggregated news sentiment score do not forecast market excess returns can be rejected on the 5%-significance level (p-value of 0.016 and 0.021 for S&P 500 and Russell 3000 related market excess returns, respectively). This strongly implies that aggregated news sentiment is somehow related to future market excess returns. The table moreover pronounces, next to the significance, the economic meaningful impact of

Table 5-3. Correlation between Sentiment and Market Variables

This table reports the Pearson correlation coefficients between the average aggregated monthly news sentiment and market variables, market index and excess market returns, for the S&P500 and the Russell3000. ***, **, and * describe significance at 1%, 5% and 10% level, respectively.

	Monthly Aggregated News Sentiment				
	Jan. 00 - Aug. 02 (Bearish)	*Sep. 02 - Jul. 07 (Bullish)*	*Aug. 07 - Feb. 09 (Bearish)*	*Mar. 09 - Nov.17 (Bullish)*	*Jan. 00 - Nov. 17 (Total period)*
SP500	0.03	0.59***	0.83***	0.46***	0.56***
$SP500_{exret}$	0.51***	0.14	0.57**	-0.07	0.35***
Russell 3000	0.03	0.59***	0.83***	0.47***	0.60***
Russell 3000_{exret}	0.51***	0.15	0.55**	-0.04	0.35***

today's aggregated news sentiment on the next month’s market excess returns. A one-standard-deviation change in the aggregated sentiment score results on average in a 1.3% (or about 130 basis points) market excess return movement in the following month. Furthermore, we observe a mutual Granger-causality relationship with the aggregated sentiment score as the regressand. Taking both directions into account, one can subsume that the aggregated news sentiment is somehow associated with market excess returns from the previous month as well as the previous three months aggregated media sentiment (p-values of 0.024 and 0.043 for the Chi^2-tests with SP500 and Russell 3000 market excess returns, respectively). This adumbrates a positive short-term correlation and thus a short-term momentum of news sentiment in a three months period.

In summary, news sentiment seems to be associated with future returns on the aggregated level. We also find indications for a short-term momentum of news sentiment. It is, therefore,

Table 5-4. LAG-Order Selection Statistics for VAR (Russell 3000)

This table shows the results for the LAG-order tests for the VAR analysis according to distinct criterion. * indicates the lag order selected by each criterion. Following criterion are applied: FPE = Final prediction error, AIC = Akaike information criterion, SBIC = Schwarz information criterion, and HQIC = Hannan-Quinn information criterion.

Lag	Likelihood-Ratio	DoF	p-Value	FPE	AIC	HQIC	SBIC
0				0.001640	-0.737439	-0.724372	-0.705129
1	218.582089	4	0.0000	0.000590	-1.759682	-1.720481	-1.662753
2	35.457902	4	0.0000	0.000516	-1.892973	-1.827637	-1.731425
3	87.176535	4	0.0000	0.000352*	-2.277325*	-2.185855*	-2.051158*
4	2.621815	4	0.6230	0.000361	-2.251217	-2.133614	-1.960432

Table 5-5. Excess Market Return Prediction with Aggregate Sentiment Score

This table reports the VAR results with three lags for each endogenous variable (market excess return and news sentiment). The news sentiment score is weighted by relevance and market capitalization and normalized with zero-mean and unit-variance. ***, **, and * describe significance at 1%, 5% and 10% level, respectively.

	Model 1: $Mkt\text{-}Rf_{SP500}$	*Model 2: CSS*	*Model 3: $Mkt\text{-}Rf_{Russell}$*	*Model 4: CSS*
$Mkt\text{-}Rf_{t-1}$	0.056	2.534***	0.068	2.293***
	(0.768)	(3.036)	(0.932)	(2.824)
$Mkt\text{-}Rf_{t-2}$	-0.116	-0.171	-0.110	-0.204
	(-1.626)	(-0.208)	(-1.540)	(-0.255)
$Mkt\text{-}Rf_{t-3}$	0.067	-0.081	0.057	-0.080
	(0.935)	(-0.099)	(0.793)	(-0.101)
CSS_{t-1}	0.013**	0.300***	0.013**	0.302***
	(2.405)	(4.971)	(2.370)	(4.993)
CSS_{t-2}	-0.003	0.057	-0.004	0.057
	(-0.607)	(0.892)	(-0.628)	(0.902)
CSS_{t-3}	0.000	0.573***	-0.000	0.573***
	(0.015)	(9.626)	(-0.021)	(9.559)
χ^2-stat *Mkt-Rf (all lags)*		9.405		8.133
p-Value *Mkt-Rf*		0.024		0.043
χ^2-stat *CSS (all lags)*	10.323		9.778	
p-Value *CSS*	0.016		0.021	

plausible at this point to assume that news sentiment moves market returns. The next sections will further examine the return predictability of news sentiment from the cross-sectional perspective.

5.6. News Sentiment and the Cross-section of Stock Returns

In this section, we examine in detail the cross-sectional relationship between news sentiment and stock returns. First, we examine raw return differentials in a univariate analysis. We then examine excess returns in multivariate spanning regressions while controlling for widely accepted risk factors.

5.6.1. Return Differentials in Univariate Analysis

Table 5-6 represents the average monthly returns in the next months of stocks sorted by firm-characteristics and news sentiment. Our firm characteristics are chosen according to well-accepted factors such as size, book-to-market or momentum. Since firm characteristics might be subject to skewed distributions, we divide each firm into tercile groups to ensure a reliable sample size for each group.[41] However, the sentiment sorting does not follow tercile groups in this section but we sort the sample stocks according to true positive (CSS > 50), neutral (CSS = 50), and negative sentiment (CSS < 50). We conjecture that if sentiment affects individual asset returns and subsume to a possible market-wide systematic component, we independently from percentile distributions, should expect return differentials between positive and negative sentiment related portfolios. Stocks that do not experience any news media coverage are assigned a neutral media sentiment score. The equal-weighted average returns are computed for the portfolios for the subsequent month. Former studies (Diether et al., 2002; Chan, 2003; Kumar and Lee, 2006; Fang and Peress, 2009) also reported equal-weighted average returns. The main motivation for our use of equal-weighted returns results from common findings of the sentiment related literature that especially smaller stocks are sensitive to changes in sentiment. The overall measurement of sentiment, therefore, favors the usage of equal-weighted returns to prevent the overweighting of possibly less sentiment sensitive large capitalization stocks.

The first row of Table 5-6 presents the unconditional average next month returns for negative, neutral and positive media sentiment portfolios with values of 0.51%, 0.70%, and 1.38%, respectively. Therefore, the average monthly return differential for an equal-weighted long-short portfolio, that goes long on positive sentiment stocks and shorts negative sentiments stocks equals 0.87% (corresponds to an annual return of 10.44%). The difference is highly significant with t-statistics of 20.45 and of high economic importance. For comparison, in a related study on media coverage, Fang and Peress (2009) identified a monthly return differential of 0.39% between stocks that are not covered by news media compared to stocks with high media attention.[42]

[41] Terciles 1 and 3 refer to the lowest and highest values.
[42] Fang and Peress (2009) describe their finding as the no-media premium.

Table 5-6. Stock Returns and Sentiment: Predictive Return Differentials

This table shows the average monthly raw returns sorted by negative, neutral and positive sentiment. Average raw returns are denoted in percentages. In each month, we divide our sample stocks into negative, neutral or positive sentiment bins. Sentiment is measured by the Ravenpack Composite Sentiment Score (CSS) based on categories developed according to more than 330 market moving events. Return differentials of positive and negative sentiment bins are calculated based on equal-weighted average next month returns of the sentiment portfolios. Return differentials are also computed for subsamples of firms sorted by size, book-to-market, momentum, volatility, operative profitability, investments, and liquidity.

	Average monthly returns (%)					**Share of sent. in obs. (%)**		
	Sentiment				t-Statistics	Sentiment		
	Neg.	Neut.	Pos.	Pos - Neg.	Pos - Neg.	Neg.	Neut.	Pos.
All stocks	0.51	0.70	1.38	0.87	20.45	0.15	0.55	0.30
Panel A: Size								
1	-0.97	-0.14	-0.09	0.88	7.03	0.11	0.72	0.17
2	1.03	1.20	1.47	0.44	5.77	0.16	0.55	0.29
3	1.22	1.40	1.77	0.55	11.80	0.20	0.37	0.43
SMB (1-3)	-2.19	-1.55	-1.85					
Panel B: Book-to-Market								
1	1.41	1.84	2.14	0.73	7.00	0.20	0.29	0.50
2	0.76	0.82	1.32	0.57	6.46	0.21	0.31	0.48
3	-0.36	-0.24	0.14	0.51	4.54	0.17	0.52	0.31
HML(3-1)	-1.77	-2.08	-1.99					
Panel C: Momentum								
1	-0.04	0.50	1.70	1.74	17.71	0.18	0.56	0.26
2	0.57	0.63	1.14	0.57	10.58	0.12	0.60	0.28
3	1.12	1.01	1.34	0.22	3.79	0.15	0.49	0.36
WML(3-1)	1.16	0.51	-0.36					
Panel D: Volatility								
1	0.72	0.71	1.01	0.29	8.10	0.11	0.61	0.28
2	0.42	0.81	1.17	0.75	14.84	0.18	0.46	0.36
3	0.62	0.85	2.17	1.55	14.25	0.19	0.54	0.27
RMS(3-1)	-0.10	0.14	1.16					
Panel E: Profitability in t-1								
1	-0.70	-0.23	0.45	1.16	8.18	0.19	0.55	0.27
2	0.83	1.29	1.37	0.54	6.03	0.20	0.39	0.40
3	2.10	2.30	2.08	-0.02	-0.27	0.20	0.36	0.45
RMW(3-1)	2.80	2.53	1.62					
Panel F: Investment in t-1								
1	2.30	2.34	2.31	0.01	0.14	0.17	0.45	0.38
2	1.03	1.06	1.31	0.29	4.64	0.18	0.45	0.36
3	-0.60	-0.49	0.31	0.91	9.73	0.20	0.49	0.30
CMA(1-3)	2.90	2.83	2.00					
Panel G: Liquidity								
1	0.23	0.55	0.79	0.56	6.24	0.10	0.71	0.19
2	0.83	0.63	1.16	0.33	4.32	0.16	0.54	0.30
3	0.66	0.86	1.68	1.02	16.45	0.21	0.39	0.40
IML(1-3)	-0.43	-0.31	-0.89					

Consequently, only sorting stocks by news sentiment produce cross-sectional return differentials. In one of the most accepted studies on news sentiment, Tetlock (2007) argues that only media pessimism is reliably related to stock asset prices. The return differential for our unconditional sentiment groups, however, imply that positive short-term premiums are associated with positive news sentiment stocks which we refer to a “premium on optimism”. The unconditional results are generally supported by our findings on double-sorted portfolios in Panels A – G. With the exceptions for two portfolios, the return differentials between positive and negative news sentiment stocks are all positive and highly significant (t-statistics ranging from 4.32 and 17.71). Thus, independently from firm characteristics, return differentials seem to be persistently existent in the cross-section. Interestingly, monthly returns do not significantly differ for profitable or so-called robust firms (tercile 3) amongst each news sentiment group, all averaging slightly above 2%. Additionally, firms with low investments seem to be less sensitive to sentiment. These firms are likely attributed as value firms with lower growth ambitions. Most interestingly, “loser stocks” with low momentum are highly sensitive to positive news sentiment. An equally-weighted long-short portfolio for this subsample generates significant annual returns of 21% which underpins the economic importance of our results. High momentum stocks, on the other hand, are less sensitive to news sentiment in our analysis. Our momentum related findings, therefore, differ to the findings by Hillert et al. (2014). The authors report a stronger momentum effect for a portfolio of winner and loser stocks with high news coverage and matching media tones in the medium- to long-term perspective.

Overall, our results are highly consistent with the general view of the common behavioral literature that especially lower capitalization stock, high volatility, and unprofitable stocks are more sensitive to sentiment (e.g., Baker and Wurgler, 2007). The return differentials in this section are mostly driven by positive news sentiment, further denoted as a premium on optimism. This finding differs from most media sentiment related literature, which predominantly documents return predictability of media pessimism.

5.6.2. Factor Spanning Regressions for Long-Short Portfolios

To formally analyze the explanatory power of long-short-portfolio returns by controlling for well-accepted risk factors, we apply factor spanning regressions as used in the standard asset pricing related literature (e.g., Baker and Wurgler, 2006; Fang and Peress, 2009; Hirshleifer and Jiang, 2010). Each month, we assign a stock into a positive, neutral, and negative sentiment group according to the average monthly news sentiment score. We then create zero-investment portfolios which go long on stocks with positive and shorts stocks with negative news sentiment and calculate the average next month return. Repeating this approach, we obtain a time-series of portfolio returns. We define this portfolio return as the PNM (positive minus negative) factor forth on. Before conducting spanning tests by regressions of our PNM returns against other widely-accepted risk factors, we first examine the Spearman rank correlations between all factors. The results are reported in Table 5-7. This table reveals significant correlations between PNM and five out of seven widely-accepted risk factors. At this point, we are specifically concerned about the high correlation between our sentiment-based PNM and the Carhart momentum factor, since momentum until now is the most prominent widely-accepted risk factor associated with the behavioral literature. If momentum already captures part of potential systematic sentiment elements, then one would expect insignificant Alpha coefficients in the subsequent spanning regressions. The same also applies to the other risk factors.

Thus, our spanning regression model is specified as follows:

$$PNM_t = \alpha + \beta \boldsymbol{X} + u_t \tag{35}$$

where PNM is the equally-weighted portfolio return for going long on positive and short in negative sentiment stocks. X represents a vector of widely-accepted risk factors including SMB_t, HML_t, MOM_t, CMA_t, RMW_t,[43] and $PS\ LIQ_t$.[44] Table 5-8 summarizes the baseline results of our spanning regressions for the full sample period. The multivariate analysis confirms the premium on optimism even after controlling for common risk factors. We also extend the traditional five-factor models (incl. Pastor-Stambaugh liquidity, profitability, and investment) and use all seven factors, since the correlation analysis revealed high correlations amongst PNM

[43] SMB (small minus big), HML (high minus low), MOM (Carhart momentum), CMA (conservative minus aggressive) and RMW (robust minus weak) downloaded from Kenneth R. French website.
[44] Pastor-Stambaugh liquidity factor downloaded from https://faculty.chicagobooth.edu/lubos.pastor/research/.

Table 5-7. Spearman Rank Correlations of Risk Factors

This table summarizes the Spearman rank correlations between widely accepted risk factors. The risk factors include premiums based on market (MkrRf), size (SMB), book-to-market (HML), Carhart momentum (MOM), Pastor-Stambaugh liquidity (PS LIQ), profitability (RMW), investments (CMA) and news sentiment (PNM). ***, **, and * describe significance at 1%, 5% and 10% level, respectively.

	PMN	*MktRF*	*SMB*	*HML*	*MOM*	*PS LIQ*	*RMW*
MktRF	-0.30***						
SMB	-0.24***	0.31***					
HML	-0.18***	-0.01	0.11				
MOM	0.34***	-0.30***	0.06	-0.16**			
PS LIQ	-0.03	0.06	0.22***	-0.07	0.13*		
RMW	0.37***	-0.57***	-0.39***	0.13*	0.23***	-0.07	
CMA	-0.06	-0.09	0.10	0.51***	0.02	-0.13*	0.07

and most of the widely-accepted risk factors. The alpha coefficient continuously decreases from 91 to 63 basis points per points with increasing number of explaining risk factors, yet remains highly significant at the 1%-level for all specifications. Widely-accepted risk factors, therefore, only capture about 25% of the alpha compared to the basic market model. For comparison, Fang and Peress (2009) only reported a no-media premium of 23 basis points using the Carhart four-factor model extended by the Pastor-Stambaugh liquidity factor. Our results remain robust for Newey-West adjusted standard errors. The loadings in Table 5-8 are quite intriguing. The zero-investment strategy that longs stocks with positive and shorts stocks with negative sentiment has a negative exposure to small stocks but a positive exposure to momentum and profitable stocks according to coefficients of the size (SMB), momentum (MOM), and the profitability (RMW) factors.

Panels B and C of Table 5-8 separately document the intercepts for the long and short leg of the zero-investment portfolio. The alpha intercept only remains significant for the long (positive sentiment) leg of the portfolio and confirms the results of the univariate analysis that premiums mainly result from the premium on optimism. These findings support the hypothesis that investors tend to react to confirming public signals which are potentially biased towards positive sentiment whereas investors hold on potential losers.

To further assess the importance of firm characteristics in cross-sectional return patterns along with news sentiment, we repeat the spanning regression tests for different subsamples

Table 5-8. Factor-Based Time-Series Regressions for Zero-Investment-Portfolios

This table reports the results of the factor-based time-series regressions on equal-weighted average returns of portfolios that go long on stocks with average positive and shorts stocks with average negative sentiment in the previous month. In each month, a stock's sentiment is classified as positive (negative) if the average Ravenpack Composite Sentiment Score exceeds 50 (is lower than 50). The equally-weighted portfolios are formed at the beginning of the month and held for one month. The long-short portfolio returns are regressed on different widely accepted risk factors. T-statistics are in parenthesis. ***, **, and * describe significance at 1%, 5% and 10% level, respectively.

	Model 1 CAPM	*Model 2* FF 3-Factor	*Model 3* Carh. 4-Factor	*Model 4* PS 5-factor	*Model 5* FF 5-Factor	*Model 6* All 7-Factor
	PMN	PMN	PMN	PMN	PMN	PMN
Panel A: Full sample						
Mkt - rf	-0.1656***	-0.1170***	-0.0597**	-0.0710**	-0.0182	0.0070
	(-5.5282)	(-4.1607)	(-2.0310)	(-2.3544)	(-0.5899)	(0.2196)
SMB		-0.2614***	-0.2935***	-0.2939***	-0.1266***	-0.1665***
		(-6.7254)	(-7.7940)	(-7.5144)	(-3.0140)	(-3.8233)
HML		0.0322	0.0649*	0.0796**	-0.1123**	-0.0491
		(0.8673)	(1.8034)	(2.1535)	(-2.3302)	(-0.9741)
MOM			0.1093***	0.1047***		0.0877***
			(4.7305)	(4.4546)		(3.8916)
PS LIQ				0.0526		0.0246
				(1.6088)		(0.7926)
RMW					0.3453***	0.3002***
					(6.5440)	(5.5704)
CMA					0.0466	0.0197
					(0.6799)	(0.2833)
Intercept	0.0084***	0.0091***	0.0086***	0.0079***	0.0069***	0.0063***
	(6.4316)	(7.5475)	(7.5161)	(6.5553)	(5.9723)	(5.4193)
Observations	215	215	215	204	215	204
Adj. R^2	0.12	0.28	0.34	0.35	0.39	0.43
Panel B: Only positive sentiment						
Intercept	0.0095***	0.0075***	0.0082***	0.0079***	0.0080***	0.0080***
	(5.9378)	(6.7373)	(8.6410)	(8.1279)	(6.8374)	(7.7801)
Panel C: Only negative sentiment						
Intercept	0.0011	-0.0016	-0.0004	0.0001	0.0011	0.0016
	(0.4603)	(-1.0116)	(-0.3627)	(0.0601)	(0.7108)	(1.3137)

based on firm characteristics matching common risk factors. Results are tabulated in Table 5-9. Interestingly, the premium on equity effect is not concentrated among small stocks, but alpha intercepts are highly significant in both tercile groups. The premium for small stocks, however, is slightly exceeding the premium on optimism for large stocks. Prior literature often documented that return anomalies often occur among small firms (e.g., Fang and Peress, 2009).

Table 5-9. Spanning Regression by Firm Characteristics

This table reports the intercepts (alphas) of the factor-based time-series regressions on equal-weighted average returns of portfolios that go long on stocks with average positive and shorts stocks with average negative sentiment in the previous month sorted by different firm characteristics. The zero-investment portfolios are regressed on excess market returns (CAPM), the Fama-French (1993) three-factor, Carhart (1997) Momentum-factor, the Pastor-Stambaugh (2003) liquidity-factor, the Fama-French (1995) five-factor and the combination of all factor-models. The equally-weighted portfolios are formed at the beginning of the month and held for one month. T-statistics are in parenthesis. ***, **, and * describe significance at 1%, 5% and 10% level, respectively. For brevity, we exclude the mid tercile results but are shown in Appendix A3.

CAPM	FF 3-Factor	Carh. 4-Factor	PS 5-factor	FF 5-Factor	All 7-Factor
Panel A: Firm Size					
Small					
0.0089***	0.0094***	0.0090***	0.0089***	0.0069***	0.0069***
(5.4852)	(6.0605)	(5.9718)	(5.8487)	(4.6035)	(4.6626)
Big					
0.0054***	0.0062***	0.0058***	0.0055***	0.0047***	0.0046***
(4.7612)	(5.6585)	(5.6444)	(5.4080)	(4.3452)	(4.4415)
Panel B: Book-to-Market					
Low					
0.0054**	0.0062***	0.0059**	0.0056**	0.0040	0.0039
(2.1904)	(2.6410)	(2.5009)	(2.3781)	(1.6513)	(1.6332)
High					
0.0001	0.0022	0.0021	0.0030	0.0002	0.0009
(0.0167)	(0.5523)	(0.5311)	(0.7488)	(0.0506)	(0.2264)
Panel C: Momentum					
Loser					
0.0159***	0.0166***	0.0161***	0.0156***	0.0151***	0.0146***
(10.6976)	(11.2902)	(11.3336)	(10.5037)	(9.9646)	(9.5387)
Winner					
0.0008	0.0010	0.0008	-0.0000	-0.0006	-0.0012
(0.5142)	(0.7543)	(0.6020)	(-0.0030)	(-0.4502)	(-0.8609)
Panel D: Profitability					
Weak					
0.0097***	0.0101***	0.0100***	0.0098***	0.0086***	0.0085***
(5.1063)	(5.2878)	(5.2051)	(5.0435)	(4.3575)	(4.2777)
Robust					
0.0023*	0.0030**	0.0025**	0.0026**	0.0017	0.0018
(1.7455)	(2.2902)	(2.0740)	(2.0767)	(1.3029)	(1.4220)
Panel E: Investment					
Conservative					
0.0020	0.0024*	0.0022*	0.0020	0.0007	0.0007
(1.4442)	(1.8046)	(1.6825)	(1.5437)	(0.5282)	(0.4939)
Aggressive					
0.0076***	0.0082***	0.0078***	0.0075***	0.0061***	0.0060***
(5.0109)	(5.5860)	(5.4897)	(5.2704)	(4.2331)	(4.2013)

We, therefore, provide new evidence that news sentiment conceivably impacts on firms independently from size. The potential refutation that news sentiment effects are only another measure for size effects can thus be antagonized.

Most strikingly, the premium on optimism is highly significant and of major significant importance for low momentum stocks with an alpha intercept of 146 basis points per month, confirming the results from the univariate analysis while controlling for widely-accepted risk factors. It also becomes evident that the premium on equity effect is closely related to unprofitable but high-growth firms. To put our results into other words, investors carefully choose their stocks amongst past losers which fundamentally also generated low profits in the past. Firm commitments in growth, namely investments, and an underlying positive news sentiment are thus strategically anchor points for investors to earn significant returns. Altogether, the subsample spanning regressions, for the most part, confirm the results from our previous univariate analysis.

5.6.3. Robustness Tests with Quintiles-Based Long-Short Portfolios

In this section, we repeat our univariate analysis from Section 5.6.1 but divide the sentiment groups by quintiles instead of three groups according to the monthly average Ravenpack sentiment score (>50 = positive, 50 = neutral, missing = neutral, <50 = negative). Table 5-10 presents the results of the robustness test. Overall, the quintile-based analysis confirms and moreover underpins our previous results. The return differentials between the lowest and the highest news sentiment quintiles are even larger for every single panel and the overall sample. For example, the lowest momentum tercile group in Panel C generates an average monthly return differential of 1.94% between high and low news sentiment firms (compared to 1.74% if sentiment groups are divided according to average positive, neutral, or negative monthly news sentiment). The results in Table 5-10 also document that return differentials are mostly driven by positive sentiment and thus support the hypothesis of premiums on optimism.

Additionally, we create an equally-weighted zero-investment portfolio which buys the highest news sentiment quintile and sells the lowest news sentiment stocks and calculate the average next month's return. We then conduct spanning regressions on the portfolio returns and control for widely-accepted risk factors as described in the previous sections. Table 5-11 summarizes the results of the spanning regression test. The alpha intercept in Model 6 remains highly significant at the 1%-level with an average return of 6.8 basis points per month. Thus, extending the three-factor model by the momentum, liquidity, profitability and investment factors does not fully capture the explanatory power on return differentials for news sentiment based zero-investment portfolios. The findings in Panel B and C, furthermore, substantiate that the return differentials mainly result from positive news sentiment and thus underpin our premium on optimism hypothesis.

In summary, both sentiment group classifications, either by absolute average monthly news sentiment or by news sentiment quintiles, support the premium on optimism hypothesis.

Table 5-10. Stock Returns and Sentiment: Predictive Return Differentials by Quintiles

This table shows the average monthly raw returns sorted by sentiment quintiles. Average raw returns are denoted in percentages. In each month, we divide our sample stocks according to sentiment quintile bins. Sentiment is measured by the Ravenpack Composite Sentiment Score (CSS) based on categories developed according to more than 330 market moving events. Return differentials between the highest and lowest sentiment quintile bins are calculated based on equal-weighted average next month returns of the sentiment quintile portfolios. Return differentials are also computed for subsamples of firms sorted by size, book-to-market, momentum, volatility, operative profitability, investments, and liquidity.

	Average monthly returns (%)							**Share of sent. (%)**				
	Sentiment						t-Stats	Sentiment				
	Q1	Q2	Q3	Q4	Q5	Q5-Q1	Q5-Q1	Q1	Q2	Q3	Q4	Q5
All stocks	0.52	0.17	0.90	1.24	1.46	0.94	19.43	15.08	15.28	40.15	10.02	19.47
Panel A: Size												
1	-0.93	-0.80	0.04	-0.54	0.06	0.99	7.27	10.93	16.21	55.54	4.30	13.01
2	1.05	0.36	1.48	1.19	1.59	0.55	6.48	15.68	14.07	40.82	8.93	20.51
3	1.21	0.73	1.71	1.60	1.87	0.66	12.58	19.93	11.55	25.99	16.03	26.49
SMB	-2.13	-1.53	-1.67	-2.14	-1.81							
Panel B: Book-to-Market												
1	1.40	1.45	2.07	1.96	2.25	0.85	7.10	20.08	10.93	18.52	20.09	30.38
2	0.73	0.28	1.17	1.18	1.42	0.69	6.95	21.18	11.46	19.71	19.30	28.34
3	-0.37	-0.90	0.16	0.05	0.21	0.58	4.52	16.92	19.52	32.78	12.30	18.47
HML	-1.78	-2.35	-1.91	-1.92	-2.05							
Panel C: Momentum												
1	-0.04	-0.55	0.79	1.29	1.90	1.94	17.44	18.06	11.85	44.36	8.46	17.27
2	0.56	0.40	0.75	1.12	1.14	0.58	9.76	12.23	20.45	39.70	9.67	17.95
3	1.15	0.46	1.20	1.29	1.36	0.21	3.27	15.20	13.27	36.13	12.07	23.33
WML	1.19	1.01	0.42	0.01	-0.54							
Panel D: Volatility												
1	0.72	0.48	0.86	0.95	1.04	0.32	8.35	10.43	24.11	37.56	10.19	17.72
2	0.41	0.23	1.01	1.13	1.19	0.78	14.13	17.71	11.51	34.68	12.82	23.29
3	0.65	-0.59	1.14	1.81	2.32	1.67	13.77	18.76	9.27	45.05	8.10	18.83
RMS	-0.07	-1.07	0.27	0.86	1.28							
Panel E: Profitability in t-1												
1	-0.68	-1.22	0.07	0.26	0.56	1.23	7.60	18.73	12.97	41.70	9.08	17.52
2	0.84	0.15	1.62	1.21	1.46	0.62	6.14	19.85	8.99	30.69	13.74	26.72
3	2.08	0.96	2.65	1.81	2.20	0.12	1.32	19.26	7.15	28.77	13.95	30.87
RMW	2.75	2.18	2.58	1.55	1.64							
Panel F: Investment in t-1												
1	2.30	1.00	2.67	2.13	2.39	0.08	0.98	17.19	8.94	35.88	11.63	26.36
2	1.03	-0.12	1.39	1.14	1.40	0.37	5.26	18.07	10.12	35.36	12.33	24.11
3	-0.58	-1.33	-0.28	0.14	0.39	0.98	9.08	20.24	10.14	39.47	10.12	20.03
CMA	2.89	2.34	2.95	1.99	2.00							
Panel G: Liquidity												
1	0.24	0.03	0.70	0.70	0.82	0.58	5.88	10.29	15.88	54.84	4.56	14.42
2	0.84	-0.03	0.86	1.00	1.23	0.38	4.50	15.67	14.05	39.79	9.39	21.09
3	0.65	0.03	1.21	1.39	1.86	1.20	16.77	20.58	11.86	27.88	15.28	24.41
IML	-0.41	0.00	-0.51	-0.70	-1.03							

Table 5-11. Time Series Regressions for Zero-Investment-Portfolios by Quintiles

This table reports the results of the factor-based time-series regressions on equal-weighted average returns of portfolios that go long on stocks within the highest average sentiment quintile and shorts stocks within the lowest average sentiment quintile in the previous month. The equally-weighted portfolios are formed at the beginning of the month and held for one month. The long-short portfolio returns are regressed on different widely accepted risk factors. T-statistics are in parenthesis. ***, **, and * describe significance at 1%, 5% and 10% level, respectively.

	Model 1	*Model 2*	*Model 3*	*Model 4*	*Model 5*	*Model 6*
	CAPM	FF 3-Factor	Carhart 4-Factor	PS Liq. 5-factor	FF 5-Factor	All 7-Factor
	PMN	PMN	PMN	PMN	PMN	PMN
Panel A: Full sample						
Mkt - rf	-0.1822***	-0.1304***	-0.0653**	-0.0778**	-0.0303	-0.0019
	(-5.7036)	(-4.3524)	(-2.1001)	(-2.4479)	(-0.9114)	(-0.0565)
SMB		-0.2835***	-0.3199***	-0.3160***	-0.1495***	-0.1925***
		(-6.8449)	(-8.0332)	(-7.6609)	(-3.3003)	(-4.1432)
HML		0.0120	0.0492	0.0665*	-0.1364***	-0.0592
		(0.3028)	(1.2922)	(1.7065)	(-2.6250)	(-1.1011)
MOM			0.1242***	0.1192***		0.1026***
			(5.0829)	(4.8092)		(4.2650)
TradeLiq				0.0485		0.0214
				(1.4068)		(0.6467)
RMW					0.3461***	0.2913***
					(6.0847)	(5.0643)
CMA					0.0558	0.0213
					(0.7538)	(0.2869)
Intercept	0.0089***	0.0097***	0.0092***	0.0083***	0.0075***	0.0068***
	(6.3929)	(7.5833)	(7.5842)	(6.5961)	(6.0351)	(5.4956)
Observations	215	215	215	204	215	204
Adj. R^2	0.13	0.28	0.36	0.36	0.38	0.43
Panel B: Only positive sentiment (Q5)						
Intercept	0.0099***	0.0081***	0.0087***	0.0084***	0.0086***	0.0085***
	(6.3322)	(7.2000)	(8.8875)	(8.2165)	(7.2728)	(7.9148)
Panel C: Only negative sentiment (Q1)						
Intercept	0.0010	-0.0016	-0.0005	0.0001	0.0011	0.0016
	(0.4475)	(-1.0301)	(-0.3810)	(0.0422)	(0.6981)	(1.3034)

5.7. In Explanation for the News Sentiment Effect

In the following, we propose different explanations for the news sentiment effect: Overreaction to news sentiment, the persistence of noise trader risk, and momentum of news sentiment.

5.7.1. Overreaction and Return Reversal

One potential explanation for the news sentiment effect is that we observe a short-term overreaction of stock prices to a streak positive news sentiment with subsequent return reversals and the underestimation of negative news due to a feeling of cognitive dissonance to previous actions.

As argued by Daniel et al. (1998), overreaction results from the confirmation of private actions by following public signals and thus increase the (over-)confidence of investors. However, disconfirming public signals on prior private signals (e.g., bullish sentiment based on private balance sheet analysis followed by negative news sentiment) only cause confidence to fall moderately. Both patterns are referred to the self-attribution theory. If news sentiment tends to be more bullish on average, then one should expect stock returns to be more sensitive against positive news sentiment ultimately.[45] This circumstance can be confirmed by our descriptive statistics in Table 5-2. Even in alleged bearish macroeconomic time periods (2000 or 2008) the share of positive exceeded the share of negative news stories about a firm by far (31.7% vs. 18.2% and 35.6% vs. 19.2%, respectively). From the momentum perspective, a sustained streak of positive news will, therefore, let prices overshoot and cause a long-term return reversal (Hong and Stein, 1999).

To provide further evidence for this explanation, we analyze the impact of variation in formation and holding periods on zero-investment PNM portfolio returns and report the results in Table 5-12. We generally find that stocks overreact to news sentiment and experience return reversals after a holding period of up to 24 months. We find that short-term portfolio rebalancing based on news sentiment generally generates superior excess returns up to 24 months. The alpha intercepts for all factor models remain significant at the 1%-level for formation periods

[45] Antweiler and Frank (2004) report on their study that investor sentiment expressed on internet message boards is on average more bullish then bearish. Out of 1,000 hand-coded messages, they find 25% of messages to be bullish (buy) whereas only 5.5% of messages are classified as bearish (sell).

up to 3 months and holding horizons of up to 24 months. Yet, the economic relevance of the portfolio returns diminishes with increasing holding horizon. When investors base their decisions on mid-term news momentum (formation period of 6-12 months), we even find stronger evidence for return reversals. Alphas for our long-short strategy become insignificant after 24 months with formation periods of 6 or 12 months. At first glance, this is in line with theoretical models on overreaction and unconditional momentum (if only based on past returns). Let us suppose that a streak of positive news will let prices overshoot. Price correction will take place and reverse prices near its fundamentals. Thus, one can assume that positive sentiment will be attributed to overconfidence and bad news may be neglected as a consequence of cognitive dissonance. Figure 5-3 summarizes the previous results for the Fama-French three-factor alpha intercepts of our long-short strategy and depicts a return reversal for all formation periods after 24 months.

To assess the mechanisms behind the portfolio returns, we divide the analysis into the long and short legs of the portfolio and illustrate the three-factor alpha intercepts for different formation and holding periods in Figure 5-4. The long-short premium diminishes within a 24 months period due to a reversed drift of negative sentiment stocks while positive stock premiums remain remarkably stable (results are similar when we apply the four-factor model). Referring to prior literature, our results provide a tendency to support the hypothesis that media optimism on individual firms can subsume persistent and systematic risk factors in asset pricing. Media pessimism, on the other hand, does not drive return premiums for long-short portfolios. Stocks experiencing media pessimism seem to be less sensitive to sentiment in a 3 months horizon, but then generate a significant positive premium in the longer run. Supporting the theory stated by Tetlock (2007), that if media pessimism follows investor sentiment, one should expect high future returns. We, therefore, find an asymmetric media sentiment effect. An explanation for the observed media pessimism result can also be provided by the theory by Barber and Odean (2008). Stocks are less sensitive to negative sentiment because investors 1) rather hold to losers and do not want to realize losses or 2) investors usually only possess a small number of stocks and won’t engage in short selling activity due to their risk aversion.

Table 5-12. Variation of Formation and Holding Periods

This table shows the compound and mean returns for the zero-investment portfolio that buys stocks with an average positive and sells stocks with an average negative sentiment during the formation period (N=1, 3, 6, 12). Average monthly alphas are reported for holding periods between 1 and 24 months. ***, **, and * describe significance at 1%, 5% and 10% level, respectively.

	Returns		Alphas					
Holding period	Compound returns	Average return	CAPM	FF 3-Factor	Carhart 4-Factor	PS Liq. 5-factor	FF 5-Factor	All 7-Factor
Panel A: Formation period = 1 Month								
1 month	0.0087 ***	0.0087 ***	0.0084***	0.0091***	0.0086***	0.0079***	0.0069***	0.0063***
3 months	0.0485 ***	0.0162 ***	0.0163***	0.0165***	0.0164***	0.0161***	0.0159***	0.0157***
6 months	0.0517 ***	0.0086 ***	0.0084***	0.0084***	0.0083***	0.0081***	0.0082***	0.0080***
9 months	0.0494 ***	0.0055 ***	0.0054***	0.0054***	0.0053***	0.0052***	0.0052***	0.0052***
12 months	0.0478 ***	0.0040 ***	0.0038***	0.0038***	0.0037***	0.0036***	0.0037***	0.0037***
24 months	0.0373 ***	0.0016 ***	0.0018***	0.0018***	0.0017***	0.0017***	0.0018***	0.0018***
Panel B: Formation period = 3 Months								
1 month	0.0252 ***	0.0252 ***	0.0270***	0.0278***	0.0271***	0.0263***	0.0251***	0.0245***
3 months	0.0492 ***	0.0164 ***	0.0175***	0.0177***	0.0174***	0.0169***	0.0170***	0.0165***
6 months	0.0509 ***	0.0085 ***	0.0090***	0.0090***	0.0087***	0.0085***	0.0086***	0.0085***
9 months	0.0468 ***	0.0052 ***	0.0057***	0.0057***	0.0055***	0.0053***	0.0053***	0.0053***
12 months	0.0394 ***	0.0033 ***	0.0038***	0.0037***	0.0035***	0.0033***	0.0035***	0.0033***
24 months	0.0018	0.0001 ***	0.0014***	0.0013***	0.0012***	0.0011***	0.0013***	0.0013***
Panel C: Formation period = 6 Months								
1 month	0.0158 ***	0.0158 ***	0.0181***	0.0188***	0.0177***	0.0170***	0.0156***	0.0150***
3 months	0.0314 ***	0.0105 ***	0.0118***	0.0120***	0.0115***	0.0111***	0.0111***	0.0107***
6 months	0.0317 ***	0.0053 ***	0.0061***	0.0061***	0.0057***	0.0055***	0.0056***	0.0054***
9 months	0.0254 ***	0.0028 ***	0.0036***	0.0035***	0.0032***	0.0030***	0.0031***	0.0030***
12 months	0.0121 ***	0.0010 ***	0.0019**	0.0018**	0.0015**	0.0013*	0.0014*	0.0013*
24 months	-0.0379 ***	-0.0016 ***	0.0002	0.0002	0.0000	-0.0000	0.0001	0.0001
Panel D: Formation period = 12 Months								
1 month	0.0100 ***	0.0100 ***	0.0124***	0.0131***	0.0119***	0.0114***	0.0097***	0.0093***
3 months	0.0199 ***	0.0066 ***	0.0080***	0.0082***	0.0077***	0.0073***	0.0072***	0.0069***
6 months	0.0180 ***	0.0030 ***	0.0039***	0.0039***	0.0034***	0.0033***	0.0033***	0.0032***
9 months	0.0083 ***	0.0009 ***	0.0020**	0.0019**	0.0015*	0.0014*	0.0015*	0.0014*
12 months	-0.0083 ***	-0.0007 ***	0.0006	0.0005	0.0002	0.0000	0.0002	0.0000
24 months	-0.0660 ***	-0.0028 ***	-0.0005	-0.0006	-0.0007	-0.0008*	-0.0006	-0.0006

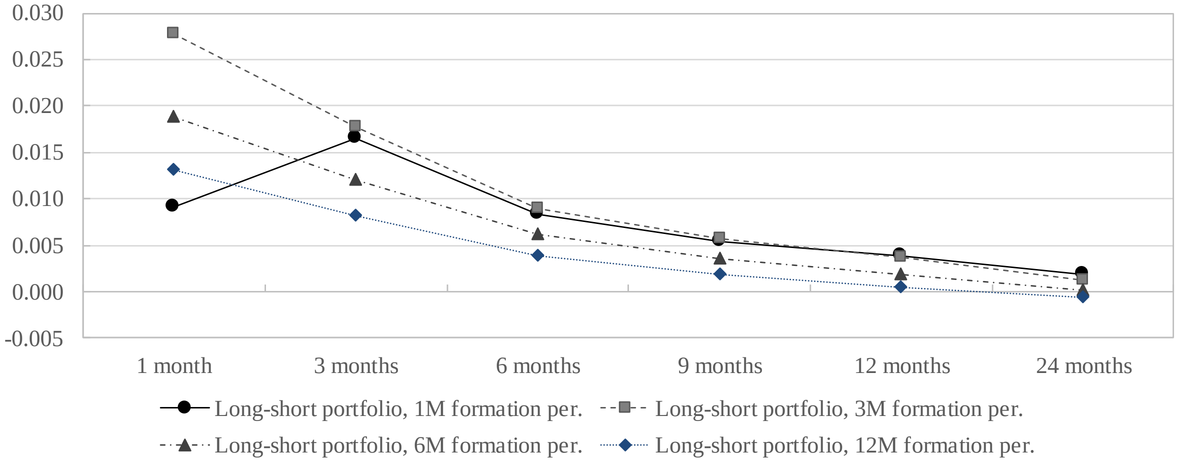

Figure 5-3. Time Horizon Analysis of News Sentiment Effect – Long-Short Portfolio

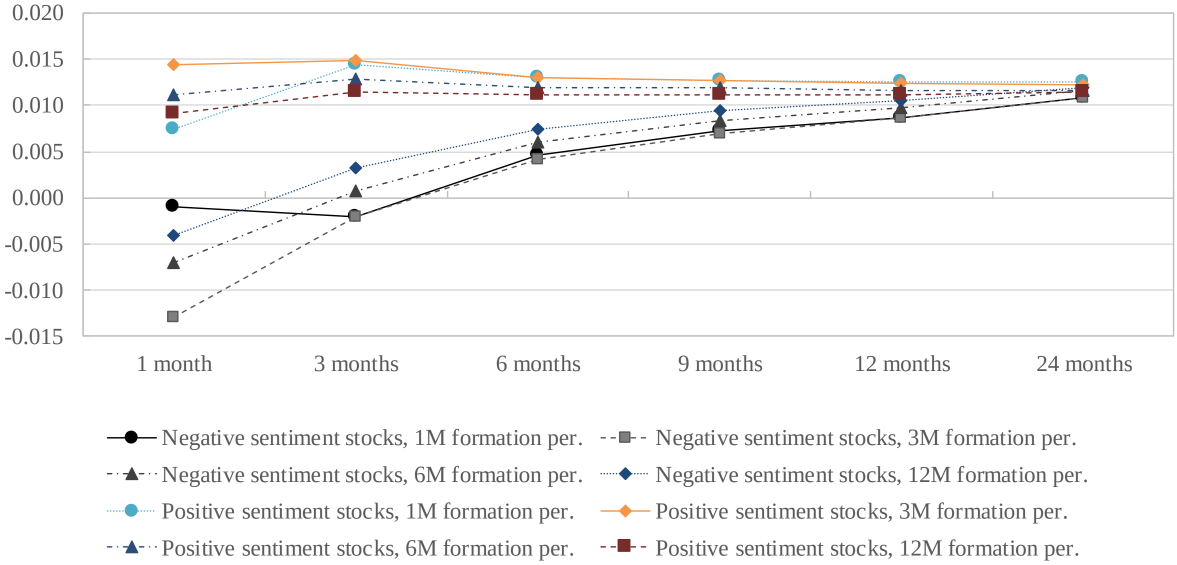

Figure 5-4. Time Horizon Analysis of News Sentiment Effect – Long and Short Legs

In summary, our results in this section propose that the news sentiment effect is not caused by overreactions to positive news and the underestimation of negative news. Instead, we find a persistent premium on positive news sentiment and a return reversal pattern for stocks associated with negative news sentiment, which confirms our premium on optimism hypothesis.

5.7.2. Limits in Arbitrage and Noise Trader Risk

Another possible explanation for the news sentiment effect is the limitation of arbitrage for rational investors to correct the mispricing of assets caused by irrational trades. In one of the most influential studies on noise trading, De Long et al. (1990) argue that short-term investors dealing with arbitrage trades against noise traders face the potential risk that a noise trader's opinion might even further change away from the mean (also if there is no fundamental risk) for a longer time horizon, so a rational investor must consider this unjustified divergence when taking a position against the noise trader. They refer this risk as noise trader risk.

In a figurative meaning to our study, when irrational investors and relatedly news sentiment are predominantly bullish in the long-term, then risk-averse and short-horizon orientated rational investors would not trade against the noise trader's position. As a result, one must expect a persistent return premium on stocks associated with positive sentiment. Corroboratively, this stable premium on optimism, or potential noise trader risk, could be observed in our previous analysis as shown in Figure 5-4.

In this section, we thus analyze the impact of NYSE short selling activity on news sentiment exposed stocks to further understand the importance of short selling activity in a market with noise traders and traders with Bayesian beliefs. We hypothesize that in times of high market sentiment total short sales should be independent of news sentiment. We, thus, perform a univariate and multivariate spanning regression analysis on short sale volume and news sentiment.[46] The results are reported in Table 5-13 and Table 5-14. The results show that long-short portfolios generate significant same-month premiums independently from the level of short selling activity. The more sophisticated implication, however, arises from next month return's implications. In comparison to previous results, long-short premiums disappear in the subsequent month for stocks with high short selling activity. For low short-sale terciles, alpha intercepts remain significant at the 5%-level ranging from 5.4 to 6.2 basis points depending on the model. However, alphas become insignificant for the Fama-French five-factor and the combined seven-factor model.

[46] We source short selling data from http://www.finra.org

Table 5-13. Stock Returns and Sentiment: Short Sale Volume

This table shows the average monthly raw returns sorted by negative, neutral and positive sentiment and by short sale volume terciles. Average raw returns are denoted in percentages. In each month, we divide our sample stocks into negative, neutral or positive sentiment bins. Sentiment is measured by the Ravenpack Composite Sentiment Score (CSS) based on categories developed according to more than 330 market moving events. Return differentials of positive and negative sentiment bins are calculated based on equal-weighted average current and next month returns of the sentiment portfolios.

	Average monthly returns (%)					**Share of sent. in obs. (%)**		
	Sentiment				t-Statistics	Sentiment		
	Neg.	Neut.	Pos.	Pos - Neg.	Pos - Neg.	Neg.	Neut.	Pos.
Panel A: Short sales volume and contemp. returns								
1	-1.12	0.15	1.63	2.75	26.08	0.12	0.59	0.29
2	-1.11	0.98	2.87	3.99	43.35	0.17	0.44	0.40
3	-1.96	1.04	3.09	5.04	59.47	0.18	0.32	0.50
LMH (1-3)	0.84	-0.90	-1.46					
Panel B: Short sales volume and next month returns								
1	0.29	0.18	0.97	0.68	6.45	0.12	0.59	0.29
2	0.91	1.00	2.00	1.10	11.84	0.17	0.45	0.39
3	0.66	1.12	2.12	1.46	17.22	0.17	0.33	0.49
LMH (1-3)	-0.36	-0.93	-1.14					

In conclusion, we only find slight evidence that short selling activity reduces the impact of news sentiment after controlling for widely-accepted risk factors. The former section underpinned that media optimism seem to be persistent and therefore a systematic character in asset pricing. It is yet unclear whether media optimism transfers value-relevant information which is not yet incorporated in asset prices or whether impediments to short selling systematically prevent rational investors from correcting mispricing which then results in a systematic deviation from the firm’s fundamentals. The truth will be an element of both explanations. The next section further examines, whether media optimism is more persistent compared to media pessimism, we analyze unconditional and conditional sentiment transition matrices in the following section.

Table 5-14. Spanning Regression for Short Sale Volume Subsample

This table reports the intercepts (alphas) of the factor-based time-series regressions on equal-weighted average returns of portfolios that go long on stocks with average positive and shorts stocks with average negative sentiment in the previous month sorted by short-sale volume. The zero-investment portfolios are regressed on excess market returns (CAPM), the Fama-French (1993) three-factor, Carhart (1997) Momentum-factor, the Pastor-Stambaugh (2003) liquidity-factor, the Fama-French (1995) five-factor and the combination of all factor-models. The equally-weighted portfolios are formed at the beginning of the month and held for one month. T-statistics are in parenthesis. ***, **, and * describe significance at 1%, 5% and 10% level, respectively.

Model 1	*Model 2*	*Model 3*	*Model 4*	*Model 5*	*Model 6*
CAPM	FF 3-Factor	Carhart 4-Factor	PS Liq. 5-factor	FF 5-Factor	All 7-Factor
Panel A: Short sale volume with contemp. monthly returns					
Low short sale volume					
0.0280***	0.0277***	0.0274***	0.0289***	0.0277***	0.0289***
(21.8697)	(21.7063)	(21.7414)	(22.5347)	(21.1655)	(22.1024)
High short sale volume					
0.0483***	0.0475***	0.0470***	0.0460***	0.0476***	0.0460***
(30.0586)	(31.8922)	(32.9408)	(30.6393)	(31.5183)	(30.0684)
Panel B: Short sale volume with next month returns					
Low short sale volume					
0.0054**	0.0062***	0.0059**	0.0056**	0.0040	0.0039
(2.1904)	(2.6410)	(2.5009)	(2.3781)	(1.6513)	(1.6332)
High short sale volume					
0.0001	0.0022	0.0021	0.0030	0.0002	0.0009
(0.0167)	(0.5523)	(0.5311)	(0.7488)	(0.0506)	(0.2264)

5.7.3. Persistence of Sentiment and Crowd Momentum

In the previous sections, we provided evidence on the persistence of a premium on optimism over a two-year time horizon. A stable premium on optimism implicitly conditions a directional continuity of investor or news sentiment. We, thus, examine unconditional (independently from previous news sentiment) and conditional (previous month news sentiment) news sentiment migration matrices to shed light on the persistence or momentum of positive news sentiment which possibly supports explaining the news sentiment effect. Hence, we conjecture that positive sentiment develops a stronger sentiment momentum (positive sentiment follows positive sentiment in the subsequent months) compared to negative news momentum and that the effect is even stronger for conditional sentiment momentum.

Table 5-15. Unconditional Sentiment Transition Matrices

This table shows the unconditional migration probability for the next months (T = 1,3,6,9,12).

		Sentiment in t+1	
	Negative	**Neutral**	**Positive**
Negative	39.34%	8.08%	52.58%
Neutral	24.27%	30.62%	45.10%
Positive	26.03%	7.12%	66.85%
		Sentiment in t+3	
	Negative	**Neutral**	**Positive**
Negative	45.69%	5.56%	48.75%
Neutral	22.33%	34.22%	43.45%
Positive	23.95%	5.45%	70.60%
		Sentiment in t+6	
	Negative	**Neutral**	**Positive**
Negative	43.60%	4.89%	51.52%
Neutral	22.92%	32.52%	44.56%
Positive	25.15%	4.55%	70.30%
		Sentiment in t+9	
	Negative	**Neutral**	**Positive**
Negative	42.37%	4.41%	53.22%
Neutral	22.88%	32.24%	44.88%
Positive	25.71%	4.01%	70.29%
		Sentiment in t+12	
	Negative	**Neutral**	**Positive**
Negative	43.33%	3.71%	52.96%
Neutral	21.65%	34.89%	43.46%
Positive	25.02%	3.51%	71.47%

The probability loadings for the unconditional and conditional news sentiment transition matrices are presented in Table 5-15 and Table 5-16. A characteristic of all matrices is the higher probability load on the matrix diagonals in particular for positive news sentiment. The unconditional probability that the current month's positive news sentiment remains positive in the next twelve months ranges between 67% and 71%, while probability loadings for negative sentiment only range between 39% and 46%. The momentum is even stronger for probability loadings in the conditional sentiment transition matrices in Table 5-16. If an average positive news sentiment month follows another average positive news sentiment month, it is more likely that the next month's (or the month in t+3, t+6, t+9, t+12) news sentiment about that firm is optimistic again. The probability loadings then range between 72% and 74%. In addition, it is then less likely that the current positive monthly news sentiment reverses into negative subsequently. Supplementary, we document similar findings for probability loadings associated with negative news sentiment. Previous month news pessimism, therefore, increases the probability that negative news sentiment persists in the following months. The probability loadings, however, are less pronounced compared to such related to positive news sentiment.

In total, we find strong implications for the persistence of positive news sentiment which equalizes to a positive news momentum. It is more likely that positive sentiment persists in the short- to mid-term compared to negative news sentiment. The higher probability of positive news momentum, therefore, supports the hypothesis that stable premiums on optimism and limits on arbitrage are accompanied or fortified by the momentum of positive news sentiment.

Table 5-16. Conditional Sentiment Transition Matrices

Panel A: Positive sentiment momentum

	Negative	**Neutral**	**Positive**
Sentiment in t+1			
Negative	33.95%	6.99%	59.05%
Neutral	24.36%	20.79%	54.85%
Positive	22.70%	5.71%	71.59%
Sentiment in t+3			
Negative	41.60%	4.81%	53.59%
Neutral	23.15%	24.58%	52.27%
Positive	21.28%	4.51%	74.21%
Sentiment in t+6			
Negative	39.56%	4.31%	56.13%
Neutral	24.42%	21.72%	53.86%
Positive	22.66%	3.72%	73.62%
Sentiment in t+9			
Negative	38.83%	3.94%	57.23%
Neutral	25.10%	20.42%	54.48%
Positive	23.39%	3.30%	73.31%
Sentiment in t+12			
Negative	40.89%	3.17%	55.94%
Neutral	25.09%	21.93%	52.99%
Positive	23.12%	2.87%	74.01%

Panel B: Neutral sentiment momentum

	Negative	**Neutral**	**Positive**
Sentiment in t+1			
Negative	32.60%	21.88%	45.52%
Neutral	15.17%	53.88%	30.95%
Positive	24.19%	21.89%	53.91%
Sentiment in t+3			
Negative	42.97%	14.73%	42.29%
Neutral	12.39%	59.47%	28.14%
Positive	21.64%	16.11%	62.24%
Sentiment in t+6			
Negative	41.96%	13.55%	44.49%
Neutral	11.83%	61.14%	27.04%
Positive	23.21%	14.72%	62.07%
Sentiment in t+9			
Negative	39.82%	12.26%	47.92%
Neutral	11.40%	62.29%	26.31%
Positive	24.52%	13.28%	62.19%
Sentiment in t+12			
Negative	43.59%	10.85%	45.55%
Neutral	9.45%	66.79%	23.76%
Positive	23.03%	11.43%	65.53%

Panel C: Negative sentiment momentum

	Negative	**Neutral**	**Positive**
Sentiment in t+1			
Negative	47.88%	6.70%	45.43%
Neutral	35.98%	18.27%	45.75%
Positive	35.23%	6.33%	58.45%
Sentiment in t+3			
Negative	51.65%	5.00%	43.35%
Neutral	31.75%	22.94%	45.32%
Positive	31.21%	5.30%	63.49%
Sentiment in t+6			
Negative	49.21%	4.45%	46.34%
Neutral	32.69%	19.30%	48.00%
Positive	31.76%	4.65%	63.60%
Sentiment in t+9			
Negative	47.30%	4.03%	48.67%
Neutral	31.94%	18.14%	49.91%
Positive	31.64%	4.10%	64.26%
Sentiment in t+12			
Negative	46.43%	3.48%	50.09%
Neutral	30.15%	20.93%	48.93%
Positive	30.05%	3.58%	66.37%

5.8. Time Variation of Loadings and Systematic Risk

So far, we have demonstrated that news optimism exhibits persistent characteristics throughout all of our analysis. In this section, we test whether the news sentiment effect is unconditional on time variation and whether news sentiment captures systematic risk in replacement of the widely discussed momentum factor. First, we split our sample into four different time regimes and repeat the PNM factor spanning regressions to understand the time variation in factor loadings and alphas. The results are reported in Table 5-17. We find that the news sentiment effect is only significant in the bull market period between March 2009 and December 2016 with an alpha coefficient of 115 basis points while controlling for seven risk factors.[47] One might argue that news sentiment in this study is only a timely limited factor which cannot be seen as a systematic risk factor. Classical studies on asset pricing often based their overall sample on time periods starting in 1960s until the early 2000s (e.g., Fama and French, 1992; Fama and French, 2015). Yet, behavioral aspects in asset pricing models became more popular in the late 1990's for researchers.[48] If researchers take into account that news media coverage increased dramatically with the expansion of the internet usage, one must accept the fact that variation in systematic risk is also conditional on advances in technology. Neglecting the rapid change in technological advances and investor research behavior, one would prefer to turn a blind eye to the acceptance of academic progress than adapting the general logic behind the financial market behavior. In our context, only the RPNA data set more than tripled their available overall news database between 2009 and 2017.

To further elaborate on the topic of systematic risk components, we furthermore conduct a traditional Fama-MacBeth regression analysis on the firm-level. Momentum was so far and still is a widely discussed and accepted phenomenon and considered to bear a behavioral component in asset pricing models. One weakness of this measure is, however, that researchers use

---

[47] Market, SMB, HML, MOM, PS LIQ, RMW and CMA. Pastor-Stambaugh liquidity factor only available until 2016.

[48] Hirshleifer and Jiang (2010) examine market mispricing associated with equity and debt financing (initial public offerings, seasoned equity offerings, debt offerings, equity repurchases and debt repurchases) for the period between 1970 and 2008. Kumar and Lee (2006) study on retail investor transactions and its relation to systematic risk over 1991 to 1996. In a longer time horizon between 1962 and 2001, Baker and Wurgler (2006) analyzed how investor sentiment (indirectly proxied by well-known market variables) is reflected in the cross-section of returns.

an output factor to measure its impact on another output variable. RPNA offers a direct measure of news sentiment and thus could replace an indirect measure, as return momentum, when possible. Table 5-18 shows the results of the Fama-MacBeth regression analysis which also examines potential differences of loadings in bullish (Sep. 2002 – Jul 2007 and Mar. 2009 – Nov. 2017) and bearish market periods (Jan. 2000 – Aug. 2002 and Aug. 2007 – Feb. 2009). In the first step, we estimate the market excess loadings for each firm based on the previous year. We then apply the loadings as the independent variable in the second step to measure the market risk premium and the impact of additional independent control variables (size, book-to-market, momentum, average monthly news sentiment in the previous month). When including the short selling variable in our model, we find that the current month's news sentiment is significantly associated with a firm's excess returns in the next month.[49] The results also suggest that growth firms are associated with positive returns. Momentum, however, does not exhibit any significance as shown in the Table.[50]

Overall, the results infer that a direct measure of investor sentiment, in our study more specifically news sentiment, should be taken into consideration as an elementary behavioral component of asset pricing models.

[49] NYSE short selling data is only available since August 2009 on www.finra.org.

[50] For robustness, we also conducted tests on fama French 25 size and book-to-market portfolios. Since portfolio loadings are expected to be more robust and stable compared to individual sentiment loadings, we would expect a positive risk premium for PMN portfolios. However, in intabulated results, we find no evidence on portfolio level.

Table 5-17. Spanning Regressions in Time Periods

This table reports the results of the factor-based time-series regressions on equal-weighted average returns of portfolios that go long on stocks with average positive and shorts stocks with average negative sentiment in the previous month. In each month, a stock's sentiment is classified as positive (negative) if the average Ravenpack Composite Sentiment Score exceeds 50 (is lower than 50). The equally-weighted portfolios are formed at the beginning of the month and held for one month. The long-short portfolio returns are regressed on different widely accepted risk factors. T-statistics are in parenthesis. ***, **, and * describe significance at 1%, 5% and 10% level, respectively.

	Panel A: Jan. 00 - Aug. 02		Panel B: Sep. 02 - Jul. 07		Panel C: Aug. 07 - Feb. 09		Panel D: Mar. 09 - Dez. 16	
	Bear market		*Bull market*		*Bear market*		*Bull market*	
	FF 5-Factor PMN	7-Factor PMN	FF 5-Factor PMN	7-Factor PMN	FF 5-Factor PMN	7-Factor PMN	FF 5-Factor PMN	7-Factor PMN
Mkt - Rf	-0.0950	-0.0749	-0.1033	-0.0663	-0.1771**	-0.1670**	-0.0585	-0.0160
	(-0.7342)	(-0.5510)	(-1.5147)	(-0.9875)	(-2.9480)	(-3.0558)	(-1.4792)	(-0.4513)
SMB	-0.1146	-0.0608	0.1800**	0.0379	0.3168**	0.3476***	-0.1313**	-0.1415**
	(-0.9837)	(-0.4432)	(2.1016)	(0.4068)	(2.5780)	(3.2887)	(-2.1107)	(-2.5078)
HML	0.0800	0.1941	-0.0361	-0.0430	-0.2546***	-0.2399***	-0.3262***	-0.1205
	(0.4497)	(0.9559)	(-0.3099)	(-0.3884)	(-3.6576)	(-3.2670)	(-4.3851)	(-1.5966)
MOM		0.0459		0.1752***		0.1263**		0.1439***
		(0.7027)		(3.0237)		(2.6873)		(5.0838)
LIQ		-0.1009		0.0302		-0.0123		0.0524
		(-0.8684)		(0.4716)		(-0.2616)		(1.1592)
RMW	0.3493**	0.3619**	0.2043	0.0380	0.1580	-0.0441	0.0664	0.1253
	(2.2026)	(2.2461)	(1.4400)	(0.2636)	(0.8712)	(-0.2634)	(0.6961)	(1.4866)
CMA	-0.1489	-0.3052	0.0873	0.1291	-0.4743**	-0.3596*	0.2180*	0.0328
	(-0.6550)	(-1.1093)	(0.4577)	(0.7009)	(-2.3990)	(-1.9882)	(1.7949)	(0.2908)
Intercept	0.0016	0.0026	0.0011	0.0012	-0.0002	0.0007	0.0116***	0.0115***
	(0.3195)	(0.4974)	(0.6298)	(0.6655)	(-0.0530)	(0.2598)	(8.0226)	(9.1084)
# Obs.	32	32	59	59	19	19	94	94
Adj .R^2	0.64	0.63	0.14	0.24	0.69	0.79	0.35	0.50

Table 5-18. Fama-MacBeth Regressions: Cross Section of Individual Stock Returns

This table shows the Fama-MacBeth regression results on firm-level. The dependent variable is the monthly excess return of an individual stock. β_{MKT} is estimated with the prior year excess market return. The month-end market capitalization is captured by ln(Size), ln(BE/ME) describes the natural logarithm of the book-to-market ratio with book value of equity from the year-end of the previous fiscal year, $MOM_{-12;-2}$ is the past 12 to previous 2 months cumulative raw return, CSS_{-1} is the average monthly composite news sentiment score, and ln(short) is the natural logarithm of the monthly short-sale volume of a firm. Robust Newey-West t-Statistics are reported in parenthesis.

Model / Period	**β_{MKT}**	**ln(Size)**	**ln(BE/ME)**	**$MOM_{-12;-2}$**	**CSS_{-1}**	**ln(Short)**
Jan.'00 – Nov.'17	-0.010	0.006	-0.050	-0.049		
	(-0.14)	(1.12)	(-1.62)	(-0.34)		
	-0.088	0.052	-0.451	-0.477	0.181	
	(-0.13)	(1.1)	(-1.57)	(-0.39)	(1.17)	
Jan.'00 – Aug.'02 &	0.090	0.059	-0.389	-0.026	0.164	
Aug.'07 – Feb.'09	(-0.04)	(1.14)	(-1.64)	(-0.08)	(1.34)	
Sep.'02 – Jul.'07 &	-0.623	0.033	-0.635	-1.830	0.231	
Mar.'09. – Nov.'17	(-0.42)	(1.03)	(-1.36)	(-1.26)	(0.67)	
Jan.'10 – Nov.'17	0.010	-0.015	-0.050	-0.058	0.034	0.016
	(0.04)	(-0.01)	(-2.22)	(-0.58)	(2.29)	(0.42)

5.9. Conclusion

We study the tone of news media and the cross-section of stock returns. Our results provide evidence that an equally-weighted long-short portfolio of stocks sorted by the tone of the news media coverage earns significant returns of 7.5% per year even after controlling for market, size, book-to-market, momentum, liquidity, profitability, and investment factors. Separating the effect of positive and negative news media tones reveals that results are mainly driven by positive news media tone which we refer to as a “premium on optimism”.

We offer several explanations for this result. First, we find a persistent premium on positive news sentiment and a return reversal pattern for stocks associated with negative news sentiment. Hence, the short-term overreaction of investors does not serve as a possible explanation for short- and long-term premiums of news optimism. Second, we only find weak evidence for sentiment related arbitrage transactions by informed investors. This supports the view on limitations of arbitrage in an environment of high investor sentiment which ultimately results in persistent mispricing or the premium on optimism. Third, we document a generally stronger manifestation of positive news sentiment on the firm-level even in bearish market environments. Finally, we find that news momentum is more likely for positive rather than negative news sentiment which favors the presence of a stable premium on optimism.

6. Overall Summary and Conclusion

The overarching goal of this dissertation is to shed light on the role of investor attention and sentiment in financial markets. In order to achieve this, we employ a variety of empirical methods in different natural setups to specifically disentangle the fundamental and noisy components conveyed by internet postings and news announcements on (social) media channels. In this context, we do not only apply standard methodologies but also enhance the empirical setups with the comprehensive application of data sources from various innovative media channels, including social media platforms, media press releases and internet search queries.

As described in the introduction, the research in this field gains more and more in importance since the way how information travels has changed significantly in recent decades. Social networks enable retail investors to interact and communicate in real-time, individuals have quick and easy access to publicly available information via the internet. Additionally, traditional media press, which used to reach a broad population on a daily base, experiences a digital transformation with far-reaching consequences. For example, traditional media press circulation in the US dropped significantly from around 63 million in its peak time in the 1970s to around 31 million in 2017. Concurrently, the number of social media users on the internet is expected to grow annually at 11% until 2021.[51] Consequently, financial markets undergo an inevitable change in terms of information dissemination and processing (Puppis et al., 2017).

Chapter 2 of this dissertation provides the basic foundation for our theoretical and empirical work. We introduce the reader to the main concepts of efficient markets, also called the neoclassical finance theory. The transition to behavioral finance is then explained based on the concepts of limitations on arbitrage, noise trading, and investor sentiment. We also summarize the most important psychological concepts which help to explain cognitive biases, beliefs or preferences of investors that finally lead investors to trade on presumably noisy information. Seminal studies in the field of behavioral finance evolved in particular in the 1990s, yielding three constitutional behavioral theories. These theories aim to provide a comprehensive and unified framework in order to explain market anomalies based on cognitive biases or as a result

[51] Please refer to the introduction in Section 1 for figures and sources.

of the interaction between different agents in financial markets. Barberis et al. (1998) developed a parsimonious model of investor sentiment, which explains the two pervasive patterns of overreaction of stock prices to news and the underreaction of stock prices to news events, such as earnings announcements. The authors base their model on the psychological concepts of representativeness and conservatism to describe how investors form their beliefs and ultimately trade on these. In the second cognitive-based model, Daniel et al. (1998) propose a model on market over- and underreaction based on two different behavioral biases compared to Barberis et al. (1998), namely investor overconfidence and self-attribution. In their findings, overconfidence of investors is related to an overreaction of stock returns, whereas biased self-attribution drives short-term momentum and earnings drifts. The later finding is closely related to the underreaction of stock prices. In the third established model, Hong and Stein (1999) deviate from the cognitive approach and develop an interaction-based model in which two heterogeneous groups of agents, newswatchers and momentum traders, give reasons for under- and overreaction of stock prices to news signals. In this model, newswatchers slowly anticipate public news signals resulting in the underreaction of stock prices to news signals. Momentum traders follow the short-term signals of serial-autocorrelated returns and trade based on historical return information. Another group of momentum trader, who anticipate this short-term momentum at a later stage, will also trade based on past information and cause stock prices to overshoot. Stock returns will consequently reverse in the long-term in accordance to the market anomaly of market overreaction. Therefore, timing effects of market entries from momentum traders play an important role when practitioners aim to implement a trading strategy based on this model.

Chapter 2 finally provides an overview of the information flow between different stakeholders in financial markets. It becomes evident that different types of investors make their decisions based on distinctive types of information depending on their resources and dedication. Individual or retail investors lack the resources to source costly information from brokers or analysts. Social media, therefore, provides a platform where investors can interact and exchange their opinions at low costs and in real-time.

The first empirical work in this dissertation, hence, explores the role of investor sentiment expressed on internet message boards in financial markets. We analyze messages posted on the

Australian financial internet message board HotCopper. Our results show that bullish sentiment on social media is positively related to stock returns up to ten trading days, but the effect diminishes after a month. Furthermore, the event study results indicate that abnormal returns are positively related to bullish sentiment on the same day, but returns reverse on a subsequent day. Both results support the pattern of stock price overreaction to bullish investor sentiment. This presumably observed contemporaneous overreaction in firm's abnormal return is reduced by short selling activities. We argue that only informed sellers initiate short selling activities when they believe that sentiment diverges far beyond a firm's fundamentals. Hence, short sellers arbitrage against noisy sentiment traders.

Contrarily, we find that negative sentiment incorporates value-relevant information about stock underperformances with a negative correlation of up to one month. Different to the implications of positive sentiment postings, we find indications that negative sentiment predicts the underperformance of stocks in the short-term. Abnormal return reversals remain absent after days of abnormal high postings with negative sentiment. The impact of negative sentiment is much more economically meaningful compared to messages with positive sentiment. The later view is additionally supported by our last analysis in the first empirical work. In particular, we document that sentiment homogeneity (agreement amongst investors) predicts negative earnings surprises. This is in conformity with the perception that negative sentiment expressed on social media conveys value-relevant information not only in general but also around firm-specific events, such as earnings announcements. The asymmetric role of investor sentiment in financial markets, thus, creates various avenues for further research in this field.

To further understand how individual and institutional investors use information sources and how this translates to financial markets, we examine in our second empirical work the impact of media and internet coverage on target run-ups before bid announcements. Previous research relied on indirect measures of investor attention (e.g., news media, trading volume) to examine its relation to firm performances or specific events. New technologies and innovations, however, enable investors to participate and gather information in real-time and efficiently change their investment decisions. This also allows researchers to create direct measures of investor attention, such as posting activity on internet investment platforms or active internet search queries. Our study provides evidence that institutional and (un-)sophisticated individual investors use preferred channels to gather and disseminate information before bid announcements. We find that smaller and underperforming stocks that only capture the attention of sophisticated individual investors on the internet message board HotCopper but no other media channel, experience a significantly stronger target run-up before bid announcements. In this connection, investor sentiment expressed on HotCopper fails to predict target run-ups. It is rather the attention or the social media coverage itself, which in the M&A context, significantly relate to target price run-ups before bid announcements. Firms with similar fundamental characteristics but missing (social) media attention do not experience a significant run-up before the bid announcement. On the other hand, large firms are especially sensitive to analyst opinions. Positive and negative analyst upgrades have significant influences on target run-ups in the respective directions. Merger signals in traditional media press only play a minor role for institutional investors. Also, Google search activity of unsophisticated investors does not explain target run-ups.

As we compare the role of investor sentiment in the first and second empirical of this dissertation, we find ambiguous implications. However, one explanation for the minor role of investor sentiment around bid announcements is the fact that M&A announcements are by nature related to positive returns, in particular for the target firm. The pure social media attention is in this case sufficient to predict a target price run-up before the actual bid announcement date. Traditional media press, on the other hand, is not found to be important in M&A settings and our results rather suggest that information disseminated via news media possesses stale characteristics.

The uttering surprise that news coverage does not seem of economic importance as opposed to prior findings in the literature (e.g., Fang and Peress, 2009), raised our interest to further explore the role of news sentiment within a broader context. Hence, we refer to the innovative and comprehensive RavenPack news database in our third empirical work to investigate the role of news sentiment and its explanatory power for cross-sectional returns. We first conduct a vector autoregressive analysis on news sentiment and its relation to market returns (S&P 500 and Russell 3000 market indices) and find that an aggregated news sentiment score significantly predicts the next month's market return. We then perform a univariate analysis in the spirit of the common asset pricing literature. To do so, we first double-sort stocks by common firm characteristics and news sentiment. Hence, we examine the average return differentials for firms with positive and negative news exposure according to various firm characteristics (e.g., size, book-to-market, momentum, profitability). The unconditional zero-investment portfolio that goes long on (shorts) stocks with positive (negative) news sentiment earns on average a return of 0.87% in the next month. This corresponds to an annual return of 10.44%. In particular, unprofitable stocks with smaller market capitalization, low return momentum, and higher investments are more sensitive to news sentiment.

Furthermore, we conduct factor spanning regressions on the resulting positive-minus-negative factor (average return of positive news sentiment stocks minus the average return of negative news sentiment stocks) and find that our results remain robust. The zero-investment portfolio yields annual returns of about 7.5% even after controlling for other widely-accepted risk-factors, such as market, size, book-to-market, momentum, profitability, and investment. The separation of the portfolio according to positive and negative sentiment legs, however, reveals that our results are mainly driven by positive news sentiment, which we refer to as the premium on optimism. We repeat the regression on subsamples according to different firm characteristics and find confirming evidence for our previous results. Hence, we again find evidence that unprofitable, smaller stocks, with low return momentum, and high investments exhibit a higher exposure to news sentiment than other subsamples.

To better understand the long-term impact of news sentiment, we construct similar zero-investment portfolios based on different holding and formation periods. We find that return differentials for zero-investment portfolios reverse after two years in accordance to the overreaction theory. Separating the news sentiment effect, however, shows that the positive sentiment portfolio leg generates persistent positive returns, whereas the negative sentiment leg experiences return reversals. The analysis of the unconditional and conditional transition matrices of news sentiment provides further evidence that positive news sentiment is more persistent than negative news sentiment in a twelve-month period. Thus, it is more likely that positive news occur if the firm experienced positive news signals in the previous months. Finally, our Fama-MacBeth regression results show that news sentiment significantly predicts future returns on firm-level while, amongst others, controlling for market premiums and return momentum. This finding, therefore, supports the hypothesis that news sentiment captures risk and thus partly explains the cross-section of returns.

Hence, we can conclude that news sentiment plays a crucial role in the explanation of cross-sectional returns. Positive news sentiment remains persistent in the long-term and is closely linked to a premium on optimism. However, stock markets overreact to negative news sentiment in a two-year horizon, causing returns to reverse in the long-term after experiencing a negative news shock. The premium on optimism is consistent with the traditional finance theory. In this view, news sentiment reflects a risk component which ultimately translates into higher returns. Stock markets, on the other hand, overreact to negative news sentiment in accordance with the behavioral finance theory. Thus, we assert that the behavioral finance theory must be seen as a complementary element to the traditional finance view.

Comparing the later results with the findings of our second empirical work, it seems ambiguous whether news coverage does play a major role in financial markets. The asset pricing tests, however, provide stronger support for the hypothesis that news sentiment and coverage influence financial markets.

Reviewing the leading research questions raised in the introduction of this dissertation, we can provide answers from different perspectives based on robust and comprehensive empirical test settings:

1. What role does investor sentiment play in financial markets? Do investors solely follow the market or do opinions and beliefs of investors predict future returns? We can confirm that investor sentiment expressed on (social) media plays a significant role in financial markets, yet with different implications depending on the directional components of investor sentiment. For example, our empirical results support the view that bullish sentiment expressed on social media is closely related to the overreaction of stock prices. Previous studies in this field have already documented that investors on social media platforms, in particular, tend to express their bullish opinions. Hence, it becomes even more essential for rational investors to distinguish the fundamental from the noisy components in social media postings.

On the other hand, negative sentiment expressed on social media seems to contain value-relevant information and thus contributes to increasing market efficiency. Social media users on financial platform share opinions on own analysis, reports and market trends. It is thus possible that the otherwise bullish investors on social media especially share value-relevant information if they feel that a firm's fundamentals develop in negative directions. This assumption is supported by the finding that negative sentiment homogeneity (investors agree on negative sentiment) particularly predicts negative earnings surprises

The implications of media press sentiment compared to social media sentiment for financial markets, however, are mixed. Our asset pricing tests in our third empirical work in this dissertation, document a positive premium on optimism for news sentiment, which is also persistent in the long-term. However, negative news sentiment seems to be related to stock price overreactions with return reversals in the long-term. This raises the questions whether media sentiment merely captures a risk factor that is not observed otherwise or whether media sentiment reflects a behavioral bias which ultimately translates to stock price movements. Even though our results support the hypothesis that positive media sentiment incorporates a media risk factor, we cannot rule out the fact that there are other potential sources which release value-relevant information preliminary. In total, our results support the noise trading theory that behavioral biases and limitations on arbitrage can ultimately yield a noise-based risk factor.

We also cannot preclude that some investors follow historical market performances which are partly reflected by investor sentiment. Our VAR analysis indicates a bilateral relationship of investor sentiment and stock returns. In summary, we can conclude that investor sentiment significantly impacts on financial markets.

2. How does (social) media relate to financial markets in the general daily context and specifically around news events, such as earnings or M&A announcements? The results of our empirical analyses reveal a general pattern of stock price overreaction to internet message board sentiment. Also, we find that negative internet message board sentiment forecasts earnings surprises. Hence, social media seemingly conveys fundamental information surrounding firm-specific events. This is not surprising since individual investors become more and more professionalized and share their analysis and valuable opinions in social media platforms. The convergence pattern of investor sentiment agreement potentially indicates that either investors get convinced by other's opinions or just exit the discussion.

Additionally, we find media coverage to help to identify target run-ups before M&A announcements. The effect is even stronger for firms that only experience social media coverage but lack a traditional media press coverage. Our result that smaller and underperforming stocks do not experience run-ups in the absence of (social) media coverage supports the hypothesis that (social) media contributes to increasing market efficiency or somehow impacts on stock prices. The general positive premium on optimism found for media sentiment, however, contradicts the findings on sentiment from internet message boards. Positive news sentiment premiums are persistent over a two-year period, whereas firms tend to overreact to negative news sentiment.

We can conclude that investor sentiment affects financial markets differently depending on the situation (general daily context vs. firm-specific events) and media channel (traditional media press vs. social media). General stock price overreaction of small-capitalization stocks to internet message board sentiment might be evidence for the tendency of bullish speculators to share their noisy opinion on social media platforms. The detailed analysis and discussions of financial reports and other available information, on the other hand, potentially help individual investors to better understand a firm's fundamental., which manifests itself in the predictability of earnings surprises from negative sentiment on internet message boards.

3. What kind of firms are more sensitive to investor sentiment than others? We also learned from our empirical work that social media users especially focus their trades on smaller and underperforming stocks. Our panel regressions have also shown that large stocks are not significantly affected by internet message board sentiment. Similar, we find that especially small stocks are prone to target run-up identification with social media coverage. Additionally, we find in our asset pricing setup, that smaller stocks are also more sensitive to news sentiment. Furthermore, the media sentiment analysis documents that especially growth stocks with high investments, low past return momentum, high return volatility, and low operational margins are sensitive to sentiment. In total, all results of this work confirm the anecdotal evidence that smaller, riskier stocks are more sensitive to investor sentiment.

4. Does arbitrage stabilize financial markets against noise traders? We find implications that arbitrage, more specifically short selling, mitigates the overreaction of stock prices to bullish investor sentiment on internet message boards. The overreaction, however, does not fully diminish in the presence of short selling. Limitations on arbitrage seem to hinder informed traders to fully exploit arbitrage opportunities. Another reason could be a risk-aversion of informed traders because bullish investors could even further push prices away from fundamentals so that they are facing a higher short-term risk, also referred to as the noise trader risk. In addition, we only find weak evidence that short selling reduces the impact of news sentiment on financial markets. This supports the view on the limitations of arbitrage for rational investors, who fail to take opposite positions, be it due to risk-aversion or limitations, such as transaction costs.

Our findings have wide-ranging implications not only for investors but also for firms and regulators. Firms must actively learn to anticipate (social) media movements which potentially rest on spurious information. Hence, the firms' investor relation departments increasingly need to deal with (un)justifiable disputes with the general public and can hence learn to early anticipate social media activity in favor of the firm's reputation. Regulators are also interested in identifying market manipulation to not only detect fraud activity but also stabilize financial markets. However, one must clearly distinguish the intention of influential individuals or the herding behavior of a crowd. Our empirical work is based on the most common methodologies in this field of research but, furthermore, combines different media perspectives to create a unified picture of investor sentiment. One main avenue for further researchers is the understanding of institutional and individual information sourcing. In this dissertation, it remains open to debate how and when investors source their information from different media channels. It is evident that investors will source their information in all available public channels covering media press, social media press or other sources from the internet. The mechanism on how investor beliefs are formed based on media channels remains elusive.

Additionally, future research could further evaluate in which specific situations financial markets are more sensitive to investor sentiment disseminated via social media. It is, furthermore, open to debate why negative sentiment could be more reliable in conveying value-relevant information to the market compared to rather noisy information spread by bullish investors. It is thus an academic imperative to find a parsimonious but comprehensive theoretical framework, which explains the asymmetric impact of investor sentiment on financial markets independent from the question, whether cognitive biases or other market drivers cause the stock price deviation from fundamentals in connection with investor sentiment. Despite the supporting evidence that (social) media attention significantly relates to target run-ups, future research could further investigate on the different roles of institutional and individual investors on target run-ups and its long-term impact. Our study only explores the short-term period of one to two months before the actual bid announcements. Future studies could additionally focus on the long-term horizon and the consequences of failed bid announcements. Furthermore, this extended analysis would help clarify whether (social) media coverage contributes in market efficiency and whether media coverage captures valuable information or rather induces a noisy short-term overreaction in the light of a potential upcoming merger event.

Finally, future research must, therefore, further explore the role of news sentiment in distinctive ways. First, researchers can shed light on the limitations of arbitrage which explain why our observed premium on optimism can persist in the long-term. If limitations of arbitrage hinder informed traders to trade against noisy information, then researchers would find stronger support for the noise trader theory by DeLong et al. (1990). Second, it still remains unclear why positive sentiment should be more persistent than negative news sentiment. Since previous studies often documented a predominantly bullish (social) media presentation, the mechanism behind this observation remains unclear. We can conclude that news sentiment plays an important role in the explanation of the cross-section of returns. However, the mechanisms and limitations which might explain the persistence of the premium on optimism remain unsolved.

In summary, we believe that the empirical findings of this dissertation contribute significantly to the elucidation of the role of investor sentiment in financial markets and extend the baseline for further research. Furthermore, this dissertation provides important insights and implications for practitioners, including investors, regulators and listed firms.

References

Abrigo, M. R. and Love, I., 2015. Estimation of panel vector autoregression in Stata: A package of programs. *manuscript available on http://paneldataconference2015.ceu.hu/Program/ Michael-Abrigo. pdf*.

Aitken, M. J., Frino, A., McCorry, M. S. and Swan, P. L., 1998. Short sales are almost instantaneously bad news: Evidence from the Australian Stock Exchange. *The Journal of Finance* 53(6), 2205-2223.

Akbas, F., 2016. The calm before the storm. *Journal of Finance* 71(1), 225-266.

Alpert, M. and Raiffa, H., 1982. A progress report on the training of probability assessors.

Andrei, D. and Hasler, M., 2015. Investor attention and stock market volatility. *Review of Financial Studies* 28(1), 33-72.

Antweiler, W. and Frank, M. Z., 2004. Is all that talk just noise? The information content of internet stock message boards. *Journal of Finance* 59(3), 1259-1294.

Asquith, P., 1983. Merger bids, uncertainty, and stockholder returns. *Journal of Financial Economics* 11(1-4), 51-83.

Asquith, P., Mikhail, M. B. and Au, A. S., 2005. Information content of equity analyst reports. *Journal of Financial Economics* 75(2), 245-282.

Bajo, E. and Raimondo, C., 2017. Media sentiment and IPO underpricing. *Journal of Corporate Finance* 46, 139-153.

Baker, M. and Stein, J. C., 2004. Market liquidity as a sentiment indicator. *Journal of Financial Markets* 7(3), 271-299.

Baker, M. and Wurgler, J., 2000. The equity share in new issues and aggregate stock returns. *Journal of Finance* 55(5), 2219-2257.

Baker, M. and Wurgler, J., 2004. A catering theory of dividends. *Journal of Finance* 59(3), 1125-1165.

Baker, M. and Wurgler, J., 2006. Investor sentiment and the cross-section of stock returns. *Journal of Finance* 61(4), 1645-1680.

Baker, M. and Wurgler, J., 2007. Investor sentiment in the stock market. *Journal of Economic Perspectives* 21(2), 129-151.

Bamber, L. S., 1987. Unexpected earnings, firm size, and trading volume around quarterly earnings announcements. *Accounting Review* 62(3), 510-532.

Banz, R. W., 1981. The relationship between return and market value of common stocks. *Journal of Financial Economics* 9(1), 3-18.

Barber, B., Lehavy, R., McNichols, M. and Trueman, B., 2001. Can investors profit from the prophets? Security analyst recommendations and stock returns. *Journal of Finance* 56(2), 531-563.

Barber, B. M. and Odean, T., 2002. Online investors: Do the slow die first? *Review of Financial Studies* 15(2), 455-488.

Barber, B. M. and Odean, T., 2008. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies* 21(2), 785-818.

Barberis, N., Shleifer, A. and Vishny, R., 1998. A model of investor sentiment. *Journal of Financial Economics* 49(3), 307-343.

Barberis, N. and Thaler, R.; 2003. A survey of behavioral finance.

Bartov, E., 1992. Patterns in unexpected earnings as an explanation for post-announcement drift. *Accounting Review* 67(3), 610-622.

Basu, S., 1977. Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis. *Journal of Finance* 32(3), 663-682.

Baytas, A. and Cakici, N., 1999. Do markets overreact: international evidence. *Journal of Banking & Finance* 23(7), 1121-1144.

Beaver, W. H., 1968. The information content of annual earnings announcements. *Journal of Accounting Research*, 67-92.

Berkman, H., Dimitrov, V., Jain, P. C., Koch, P. D. and Tice, S., 2009. Sell on the news: Differences of opinion, short-sales constraints, and returns around earnings announcements. *Journal of Financial Economics* 92(3), 376-399.

Bernard, V. L. and Thomas, J. K., 1989. Post-earnings-announcement drift: Delayed price response or risk premium? *Journal of Accounting Research* 27, 1-36.

Betton, S. and Eckbo, B. E., 2000. Toeholds, bid jumps, and expected payoffs in takeovers. *Review of Financial Studies* 13(4), 841-882.

Black, F., 1972. Capital market equilibrium with restricted borrowing. *The Journal of Business* 45(3), 444-455.

Black, F., 1986. Noise. *Journal of Finance* 41(3), 528-543.

Black, F. and Scholes, M., 1972. The valuation of option contracts and a test of market efficiency. *Journal of Finance* 27(2), 399-417.

Bloomfield, R., O'Hara, M. and Saar, G., 2009. How noise trading affects markets: An experimental analysis. *Review of Financial Studies* 22(6), 2275-2302.

Bollen, J., Mao, H. and Zeng, X., 2011. Twitter mood predicts the stock market. *Journal of Computational Science* 2(1), 1-8.

Brigida, M. and Madura, J., 2012. Sources of target stock price run-up prior to acquisitions. *Journal of Economics and Business* 64(2), 185-198.

Brown, G. W. and Cliff, M. T., 2004. Investor sentiment and the near-term stock market. *Journal of Empirical Finance* 11(1), 1-27.

Burkart, R.; 2002. Kommunikationswissenschaft. Böhlau Verlag, Köln.

Cai, Y. and Sevilir, M., 2012. Board connections and M&A transactions. *Journal of Financial Economics* 103(2), 327-349.

Campbell, J. Y. and Shiller, R. J., 1988. The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies* 1(3), 195-228.

Campbell, J. Y. and Shiller, R. J., 1988. Stock prices, earnings, and expected dividends. *Journal of Finance* 43(3), 661-676.

Carhart, M. M., 1997. On persistence in mutual fund performance. *Journal of Finance* 52(1), 57-82.

Chakrabarty, B., Moulton, P. C. and Shkilko, A., 2012. Short sales, long sales, and the Lee–Ready trade classification algorithm revisited. *Journal of Financial Markets* 15(4), 467-491.

Chan, W. S., 2003. Stock price reaction to news and no-news: Drift and reversal after headlines. *Journal of Financial Economics* 70(2), 223-260.

Chen, H., De, P., Hu, Y. and Hwang, B.-H., 2014. Wisdom of crowds: The value of stock opinions transmitted through social media. *Review of Financial Studies* 27(5), 1367-1403.

Chen, X., Cheng, Q. and Lo, K., 2010. On the relationship between analyst reports and corporate disclosures: Exploring the roles of information discovery and interpretation. *Journal of Accounting and Economics* 49(3), 206-226.

Choi, D., 1991. Toehold acquisitions, shareholder wealth, and the market for corporate control. *Journal of Financial and Quantitative Analysis* 26(3), 391.

Chopra, N., Lakonishok, J. and Ritter, J. R., 1992. Measuring abnormal performance: do stocks overreact? *Journal of Financial Economics* 31(2), 235-268.

Clarkson, P., Joyce, D. and Tutticci, I., 2006. Market reaction to takeover rumour in Internet Discussion Sites. *Accounting and Finance* 46(1), 31-52.

Cochrane, J. H.; 2001. Asset pricing. Princeton University Press, Princeton.

Cochrane, J. H., 2011. Presidential address: Discount rates. *Journal of Finance* 66(4), 1047-1108.

Comment, R. and Schwert, G. W., 1995. Poison or placebo? Evidence on the deterrence and wealth effects of modern antitakeover measures. *Journal of Financial Economics* 39(1), 3-43.

Conrad, J. and Kaul, G., 1993. Long-term market overreaction or biases in computed returns? *Journal of Finance* 48(1), 39-63.

Cutler, D. M., Poterba, J. M. and Summers, L. H., 1989. What moves stock prices? *Journal of Portfolio Management* 15, 4-12.

Da, Z., Engelberg, J. and Gao, P., 2011. In search of attention. *Journal of Finance* 66(5), 1461-1499.

Da, Z., Engelberg, J. and Gao, P., 2015. The sum of all FEARS investor sentiment and asset prices. *Review of Financial Studies* 28(1), 1-32.

Daniel, K., Hirshleifer, D. and Subrahmanyam, A., 1998. Investor psychology and security market under-and overreactions. *Journal of Finance* 53(6), 1839-1885.

Das, S. R. and Chen, M. Y., 2007. Yahoo! for Amazon: Sentiment extraction from small talk on the web. *Management Science* 53(9), 1375-1388.

De Bondt, W. F. and Thaler, R., 1985. Does the stock market overreact? *Journal of Finance* 40(3), 793-805.

De Bondt, W. F. and Thaler, R., 1995. Handbooks in operations research and management science. 9, 385-410.

De Bondt, W. F. and Thaler, R. H., 1987. Further evidence on investor overreaction and stock market seasonality. *Journal of Finance* 42(3), 557-581.

De Long, J. B., Shleifer, A., Summers, L. H. and Waldmann, R. J., 1990. Noise trader risk in financial markets. *Journal of Political Economy* 98(4), 703-738.

DeMarzo, P. M., Vayanos, D. and Zwiebel, J., 2003. Persuasion bias, social influence, and unidimensional opinions. *Quarterly Journal of Economics* 118(3), 909-968.

Diamond, D. W. and Verrecchia, R. E., 1987. Constraints on short-selling and asset price adjustment to private information. *Journal of Financial Economics* 18(2), 277-311.

Diether, K. B., Malloy, C. J. and Scherbina, A., 2002. Differences of opinion and the cross section of stock returns. *Journal of Finance* 57(5), 2113-2141.

Dodd, P., 1980. Merger proposals, management discretion and stockholder wealth. *Journal of Financial Economics* 8(2), 105-137.

Dodd, P. and Ruback, R., 1977. Tender offers and stockholder returns: An empirical analysis. *Journal of Financial Economics* 5(3), 351-373.

Drake, M. S., Roulstone, D. T. and Thornock, J. R., 2012. Investor information demand: Evidence from Google searches around earnings announcements. *Journal of Accounting Research* 50(4), 1001-1040.

Dreman, D. N.; 1979. Contrarian investment strategy: The psychology of stock market success. Random House Incorporated, New York.

Edmans, A., Garcia, D. and Norli, Ø., 2007. Sports sentiment and stock returns. *Journal of Finance* 62(4), 1967-1998.

Edwards, W.; 1968. Conservatism in human information processing. In: B. Kleinmutz: Formal representation of human judgment. Wiley, New York, 17-52.

Engelberg, J. E. and Parsons, C. A., 2011. The causal impact of media in financial markets. *Journal of Finance* 66(1), 67-97.

Engelberg, J. E., Reed, A. V. and Ringgenberg, M. C., 2012. How are shorts informed? *Journal of Financial Economics* 105(2), 260-278.

Fama, E. F., 1965. The behavior of stock-market prices. *Journal of Business* 38(1), 34-105.

Fama, E. F., 1970. Efficient capital markets: A review of theory and empirical work. *Journal of Finance* 25(2), 383-417.

Fama, E. F., 1998. Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics* 49(3), 283-306.

Fama, E. F., Fisher, L., Jensen, M. C. and Roll, R., 1969. The adjustment of stock prices to new information. *International Economic Review* 10(1), 1-21.

Fama, E. F. and French, K. R., 1988. Dividend yields and expected stock returns. *Journal of Financial Economics* 22(1), 3-25.

Fama, E. F. and French, K. R., 1992. The cross-section of expected stock returns. *Journal of Finance* 47(2), 427-465.

Fama, E. F. and French, K. R., 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33(1), 3-56.

Fama, E. F. and French, K. R., 1998. Value versus growth: The international evidence. *Journal of Finance* 53(6), 1975-1999.

Fama, E. F. and French, K. R., 2004. The capital asset pricing model: Theory and evidence. *Journal of Economic Perspectives* 18(3), 25-46.

Fama, E. F. and French, K. R., 2015. A five-factor asset pricing model. *Journal of Financial Economics* 116(1), 1-22.

Fama, E. F. and MacBeth, J. D., 1973. Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy* 81(3), 607-636.

Fang, L. and Peress, J., 2009. Media coverage and the cross-section of stock returns. *Journal of Finance* 64(5), 2023-2052.

Figlewski, S. and Webb, G. P., 1993. Options, short sales, and market completeness. *Journal of Finance* 48(2), 761-777.

Francis, J., Schipper, K. and Vincent, L., 2002. Earnings announcements and competing information. *Journal of Accounting and Economics* 33(3), 313-342.

Frankel, R., Kothari, S. P. and Weber, J., 2006. Determinants of the informativeness of analyst research. *Journal of Accounting and Economics* 41(1–2), 29-54.

Frazzini, A. and Lamont, O. A., 2008. Dumb money: Mutual fund flows and the cross-section of stock returns. *Journal of Financial Economics* 88(2), 299-322.

French, K. R., Schwert, G. W. and Stambaugh, R. F., 1987. Expected stock returns and volatility. *Journal of Financial Economics* 19(1), 3-29.

Froot, K. A., Scharfstein, D. S. and Stein, J. C., 1992. Herd on the street: Informational inefficiencies in a market with short-term speculation. *Journal of Finance* 47(4), 1461-1484.

Gao, Y. and Oler, D., 2012. Rumors and pre-announcement trading: Why sell target stocks before acquisition announcements? *Review of Quantitative Finance and Accounting* 39(4), 485-508.

García, D., 2013. Sentiment during recessions. *Journal of Finance* 68(3), 1267-1300.

Garfinkel, J. A. and Sokobin, J., 2006. Volume, opinion divergence, and returns: A study of post–earnings announcement drift. *Journal of Accounting Research* 44(1), 85-112.

Gervais, S., Kaniel, R. and Mingelgrin, D. H., 2001. The high-volume return premium. *Journal of Finance* 56(3), 877-919.

Graham, B.; 1965. The intelligent investor: A book of practical counsel. Harper & Brothers Publishers, New York.

Graham, J. R., Lemmon, M. L. and Wolf, J. G., 2002. Does corporate diversification destroy value? *Journal of Finance* 57(2), 695-720.

Griffin, D. and Tversky, A., 1992. The weighing of evidence and the determinants of confidence. *Cognitive Psychology* 24(3), 411-435.

Grossman, S., 1976. On the efficiency of competitive stock markets where trades have diverse information. *Journal of Finance* 31(2), 573-585.

Gupta, A. and Misra, L., 1989. Public information and pre-announcement trading in takeover stocks. *Journal of Economics and Business* 41(3), 225-233.

Hamilton, J. D. and Susmel, R., 1994. Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics* 64(1), 307-333.

Heimer, R. Z., 2016. Peer pressure: Social interaction and the disposition effect. *Review of Financial Studies* 29(11), 3177-3209.

Hendershott, T., Livdan, D. and Schürhoff, N., 2015. Are institutions informed about news? *Journal of Financial Economics* 117(2), 249-287.

Hillert, A., Jacobs, H. and Müller, S., 2014. Media makes momentum. *Review of Financial Studies* 27(12), 3467-3501.

Hirshleifer, D. and Jiang, D., 2010. A financing-based misvaluation factor and the cross-section of expected returns. *Review of Financial Studies* 23(9), 3401-3436.

Hong, H. and Stein, J. C., 1999. A unified theory of underreaction, momentum trading, and overreaction in asset markets. *Journal of Finance* 54(6), 2143-2184.

Huberman, G. and Regev, T., 2001. Contagious speculation and a cure for cancer: A nonevent that made stock prices soar. *Journal of Finance* 56(1), 387-396.

Ivković, Z. and Jegadeesh, N., 2004. The timing and value of forecast and recommendation revisions. *Journal of Financial Economics* 73(3), 433-463.

Jarrell, G. A. and Poulsen, A. B., 1989. Stock trading before the announcement of tender offers: insider trading or market anticipation? *Journal of Law, Economics, & Organization* 5(2), 225-248.

Jegadeesh, N. and Kim, W., 2010. Do analysts herd? An analysis of recommendations and market reactions. *Review of Financial Studies* 23(2), 901-937.

Jegadeesh, N. and Titman, S., 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance* 48(1), 65-91.

Jegadeesh, N. and Titman, S., 2002. Cross-sectional and time-series determinants of momentum returns. *Review of Financial Studies* 15(1), 143-157.

Jones, C. M. and Lamont, O. A., 2002. Short-sale constraints and stock returns. *Journal of Financial Economics* 66(2-3), 207-239.

Kahneman, D., 1979. Prospect theory: An analysis of decisions under risk. *Econometrica* 47, 278.

Kahneman, D. and Tversky, A., 1972. Subjective probability: A judgment of representativeness. *Cognitive Psychology* 3(3), 430-454.

Kahneman, D. and Tversky, A.; 2013. Prospect theory: An analysis of decision under risk. In: L. MacLean and W. Ziemba: Handbook of the fundamentals of financial decision making: Part I. World Scientific, New Jersey, 99-127.

Kamstra, M. J., Kramer, L. A. and Levi, M. D., 2003. Winter blues: A SAD stock market cycle. *American Economic Review* 93(1), 324-343.

Kandel, E. and Pearson, N. D., 1995. Differential interpretation of public signals and trade in speculative markets. *Journal of Political Economy* 103(4), 831-872.

Kaplan, A. M. and Haenlein, M., 2010. Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons* 53(1), 59-68.

Keim, D. B., 1983. Size-related anomalies and stock return seasonality: Further empirical evidence. *Journal of Financial Economics* 12(1), 13-32.

Keown, A. J. and Pinkerton, J. M., 1981. Merger announcements and insider trading activity: An empirical investigation. *Journal of Finance* 36(4), 855-869.

Kim, O. and Verrecchia, R. E., 1994. Market liquidity and volume around earnings announcements. *Journal of Accounting and Economics* 17(1-2), 41-67.

Kim, S.-H. and Kim, D., 2014. Investor sentiment from internet message postings and the predictability of stock returns. *Journal of Economic Behavior & Organization* 107, 708-729.

King, M. R., 2009. Prebid run-ups ahead of Canadian takeovers: How big is the problem? *Financial Management* 38(4), 699-726.

Klein, R. W. and Bawa, V. S., 1977. The effect of limited information and estimation risk on optimal portfolio diversification. *Journal of Financial and Quantitative Analysis* 12(4), 669-669.

Kumar, A. and Lee, C., 2006. Retail investor sentiment and return comovements. *Journal of Finance* 61(5), 2451-2486.

La Porta, R., 1996. Expectations and the cross-section of stock returns. *Journal of Finance* 51(5), 1715-1742.

Lakonishok, J., Shleifer, A. and Vishny, R. W., 1992. The impact of institutional trading on stock prices. *Journal of Financial Economics* 32(1), 23-43.

Lang, L. H., Stulz, R. M. and Walkling, R. A., 1989. Managerial performance, Tobin's Q, and the gains from successful tender offers. *Journal of Financial Economics* 24(1), 137-154.

Lee, C. M., Shleifer, A. and Thaler, R. H., 1991. Investor sentiment and the closed-end fund puzzle. *Journal of Finance* 46(1), 75-109.

Lee, L. F., Hutton, A. P. and Shu, S., 2015. The role of social media in the capital market: Evidence from consumer product recalls. *Journal of Accounting Research* 53(2), 367-404.

Lemmon, M. and Portniaguina, E., 2006. Consumer confidence and asset prices: Some empirical evidence. *Review of Financial Studies* 19(4), 1499-1529.

LeRoy, S. F. and Porter, R. D., 1981. Stock price volatility: Tests based on implied variance bounds. *Econometrica* 49(97), 113.

Leung, H. and Ton, T., 2015. The impact of internet stock message boards on cross-sectional returns of small-capitalization stocks. *Journal of Banking & Finance* 55, 37-55.

Lintner, J., 1965. Security prices, risk, and maximal gains from diversification. *Journal of Finance* 20(4), 587-615.

Liu, L. X., Sherman, A. E. and Zhang, Y., 2009. The role of the media in initial public offerings. *Working paper.*

Liu, P., Smith, S. D. and Syed, A. A., 1990. Stock price reactions to the Wall Street Journal's securities recommendations. *Journal of Financial and Quantitative Analysis* 25(3), 399-410.

Lord, C. G., Ross, L. and Lepper, M. R., 1979. Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology* 37(11), 2098.

Loughran, T. and McDonald, B., 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance* 66(1), 35-65.

Lütkepohl, H.; 2005. New introduction to multiple time series analysis. Springer Verlag, Berlin.

Maletzke, G.; 1963. Psychologie der Massenkommunikation. Verlag Hans-Bredow-Institut, Hamburg.

Malkiel, B. G., 2003. The efficient market hypothesis and its critics. *Journal of Economic Perspectives* 17(1), 59-82.

Maquieira, C. P., Megginson, W. L. and Nail, L., 1998. Wealth creation versus wealth redistributions in pure stock-for-stock mergers. *Journal of Financial Economics* 48(1), 3-33.

Markowitz, H., 1952. Portfolio selection. *Journal of Finance* 7(1), 77-91.

Markowitz, H.; 1959. Portfolio selection: Efficient diversification of investments. Cowles Foundation Monograph No. 16. John Wiley & Sons, New York.

Marsh, T. A. and Merton, R. C., 1986. Dividend variability and variance bounds tests for the rationality of stock market prices. *American Economic Review* 76(3), 483-498.

Martynova, M. and Renneboog, L., 2008. A century of corporate takeovers: What have we learned and where do we stand? *Journal of Banking & Finance* 32(10), 2148-2177.

Merton, R. C., 1987. A simple model of capital market equilibrium with incomplete information. *Journal of Finance* 42(3), 483-510.

Meulbroek, L. K., 1992. An empirical analysis of illegal insider trading. *Journal of Finance* 47(5), 1661-1699.

Mikkelson, W. H. and Ruback, R. S., 1985. An empirical analysis of the interfirm equity investment process. *Journal of Financial Economics* 14(4), 523-553.

Miller, E. M., 1977. Risk, uncertainty, and divergence of opinion. *Journal of Finance* 32(4), 1151-1168.

Mitchell, M., Pulvino, T. and Stafford, E., 2002. Limited arbitrage in equity markets. *Journal of Finance* 57(2), 551-584.

Modigliani, F. and Miller, M. H., 1958. The cost of capital, corporation finance and the theory of investment. *American Economic Review* 48(3), 261-297.

Murray, P. G., 1994. Pre-announcement share price run-ups and abnormal trading volume in takeover target's shares: A test of the market speculation hypothesis. *Pacific-Basin Finance Journal* 2(2), 319-348.

Myers, S. C., 1977. Determinants of corporate borrowing. *Journal of Financial Economics* 5(2), 147-175.

Narayanamoorthy, G., 2006. Conservatism and cross-sectional variation in the post–earnings announcement drift. *Journal of Accounting Research* 44(4), 763-789.

Neal, R. and Wheatley, S. M., 1998. Do measures of investor sentiment predict returns? *Journal of Financial and Quantitative Analysis* 33(4), 523-547.

Neuberger, C.; 2008. Internet und Journalismusforschung. Theoretische Neujustierung und Forschungsagenda. In: T. Quandt and W. Schweiger: Journalismus online-Partizipation oder Profession? VS Verlag für Sozialwissenschaften, Wiesbaden, 17-42.

Nofer, M. and Hinz, O., 2015. Using Twitter to predict the stock market. *Business & Information Systems Engineering* 57(4), 229-242.

Nofsinger, J. R. and Sias, R. W., 1999. Herding and feedback trading by institutional and individual investors. *Journal of Finance* 54(6), 2263-2295.

Olsen, R. A., 1996. Implications of herding behavior for earnings estimation, risk assessment, and stock returns. *Financial Analysts Journal* 52(4), 37-41.

Palepu, K. G., 1986. Predicting takeover targets: A methodological and empirical analysis. *Journal of Accounting and Economics* 8(1), 3-35.

Petersen, M. A., 2009. Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies* 22(1), 435-480.

Porta, R. L., Lopez-de-Silanes, F., Shleifer, A. and Vishny, R. W., 1998. Law and finance. *Journal of Political Economy* 106(6), 1113-1155.

Puppis, M., Schenk, M. and Hofstetter, B.; 2017. Medien und Meinungsmacht. vdf Hochschulverlag AG, Zürich.

Qiu, L. and Welch, I., 2006. Investor sentiment measures. *Working paper*.

Ramiah, V., Xu, X. and Moosa, I. A., 2015. Neoclassical finance, behavioral finance and noise traders: A review and assessment of the literature. *International Review of Financial Analysis* 41, 89-100.

Rapp, H.-W.; 2000. Der tägliche Wahnsinn hat Methode. Behavioral Finance: Paradigmenwechsel in der Kapitalmarktforschung. Schäffer-Poeschel Verlag, Stuttgart.

Renault, T., 2017. Intraday online investor sentiment and return patterns in the U.S. stock market. *Journal of Banking & Finance* 84, 25-40.

Roll, R., 1977. A critique of the asset pricing theory's tests Part I: On past and potential testability of the theory. *Journal of Financial Economics* 4(2), 129-176.

Roll, R., 1986. The hubris hypothesis of corporate takeovers. *Journal of Business* 59(2), 197-216.

Roll, R. and Ross, S. A., 1980. An empirical investigation of the arbitrage pricing theory. *Journal of Finance* 35(5), 1073-1103.

Rosenberg, B., Reid, K. and Lanstein, R., 1985. Efficient capital markets II. *Persuasive Evidence of Market Inefficiency* 11(3), 9-16.

Sabherwal, S., Sarkar, S. K. and Zhang, Y., 2011. Do internet stock message boards influence trading? Evidence from heavily discussed stocks with no fundamental news. *Journal of Business Finance & Accounting* 38(9-10), 1209-1237.

Sanders, R. W. and Zdanowicz, J. S., 1992. Target firm abnormal returns and trading volume around the initiation of change in control transactions. *Journal of Financial and Quantitative Analysis* 27(1), 109-129.

Scharfstein, D. S. and Stein, J. C., 1990. Herd behavior and investment. *American Economic Review*, 465-479.

Scheinkman, J. A. and Xiong, W., 2003. Overconfidence and speculative bubbles. *Journal of Political Economy* 111(6), 1183-1220.

Scheufele, B. and Haas, A., 2008. Die Rolle der Unternehmensberichterstattung am Aktienmarkt. Eine Zeitreihenanalyse des Zusammenhangs zwischen der Print-, Online-und Fernsehberichterstattung sowie den Handelsvolumina und Kursen ausgewählter deutscher Aktien. *M&K Medien & Kommunikationswissenschaft* 56(3-4), 347-367.

Scheufele, B. and Haas, A.; 2008. Medien und Aktien. Theoretische und empirische Modellierung der Rolle der Berichterstattung für das Börsengeschehen. 1. Auflage, Verlag für Sozialwissenschaften, Wiesbaden.

Scholes, M. S., 1972. The market for securities: Substitution versus price pressure and the effects of information on share prices. *Journal of Business* 45(2), 179-211.

Schwert, G. W., 1996. Markup pricing in mergers and acquisitions. *Journal of Financial Economics* 41(2), 153-192.

Schwert, G. W., 2000. Hostility in takeovers: in the eyes of the beholder? *Journal of Finance* 55(6), 2599-2640.

Servaes, H., 1991. Tobin's Q and the gains from takeovers. *Journal of Finance* 46(1), 409-419.

Sharpe, W. F., 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* 19(3), 425-442.

Sharpe, W. F. and Alexander, G. J.; 1990. Investments. 4th ed., Englewood Cliffs, New Jersey.

Shefrin, H. and Statman, M., 1985. The disposition to sell winners too early and ride losers too long: Theory and evidence. *Journal of Finance* 40(3), 777-790.

Shiller, R. J., 1981. The use of volatility measures in assessing market efficiency. *Journal of Finance* 36(2), 291-304.

Shiller, R. J., 2000. Irrational exuberance. *Philosophy & Public Policy Quarterly* 20(1), 18-23.

Shiller, R. J., Fischer, S. and Friedman, B. M., 1984. Stock prices and social dynamics. *Brookings Papers on Economic Activity* 2, 457-510.

Shleifer, A.; 2000. Inefficient markets: An introduction to behavioral finance. Oxford University Press, Oxford.

Shleifer, A. and Summers, L. H., 1990. The noise trader approach to finance. *Journal of Economic Perspectives* 4(2), 19-33.

Shleifer, A. and Vishny, R. W., 1997. The limits of arbitrage. *Journal of Finance* 52(1), 35-55.

Siganos, A., 2013. Google attention and target price run-ups. *International Review of Financial Analysis* 29, 219-226.

Siganos, A. and Papa, M., 2012. FT coverage and UK target price run-ups. *The European Journal of Finance* 21(12),

Sims, C. A., 1980. Macroeconomics and reality. *Econometrica* 48(1), 1-48.

Slovic, P., Fischhoff, B. and Lichtenstein, S., 1977. Behavioral decision theory. *Annual Review of Psychology* 28(1), 1-39.

Sprenger, T. O., Tumasjan, A., Sandner, P. G. and Welpe, I. M., 2014. Tweets and trades: The information content of stock microblogs. *European Financial Management* 20(5), 926-957.

Stanzel, M.; 2007. Qualität des Aktienresearchs von Finanzanalysten. Eine theoretische und empirische Untersuchung der Gewinnprognosen und Aktienempfehlungen am deutschen Kapitalmarkt. Deutscher Universitätsverlag, Wiesbaden.

Stattman, D., 1980. Book values and stock returns. *The Chicago MBA: A journal of selected papers* 4(1), 25-45.

Stulz, R. M., Walkling, R. A. and Song, M. H., 1990. The distribution of target ownership and the division of gains in successful takeovers. *Journal of Finance* 45(3), 817-833.

Tetlock, P. C., 2007. Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance* 62(3), 1139-1168.

Tetlock, P. C., Saar-Tsechansky, M. and Macskassy, S., 2008. More than words: Quantifying language to measure firms' fundamentals. *Journal of Finance* 63(3), 1437-1467.

Trueman, B., 1994. Analyst forecasts and herding behavior. *Review of Financial Studies* 7(1), 97-124.

Tumarkin, R. and Whitelaw, R. F., 2001. News or noise? Internet postings and stock prices. *Financial Analysts Journal* 57(3), 41-51.

Tversky, A. and Kahneman, D., 1974. Judgment under uncertainty: Heuristics and biases. *Science* 185(4157), 1124-1131.

Tversky, A. and Kahneman, D., 1981. The framing of decisions and the psychology of choice. *Science* 211(4481), 453-458.

Vlastakis, N. and Markellos, R. N., 2012. Information demand and stock market volatility. *Journal of Banking & Finance* 36(6), 1808-1821.

Von Neumann, J. and Morgenstern, O.; 2007. Theory of games and economic behavior. Princeton university press, Princeton.

Vozlyublennaia, N., 2014. Investor attention, index performance, and return predictability. *Journal of Banking & Finance* 41, 17-35.

Welagedara, V., Deb, S. S. and Singh, H., 2017. Investor attention, analyst recommendation revisions, and stock prices. *Pacific-Basin Finance Journal* 45, 211-223.

Whaley, R. E., 2000. The investor fear gauge. *Journal of Portfolio Management* 26(3), 12-17.

Womack, K. L., 1996. Do brokerage analysts' recommendations have investment value? *Journal of Finance* 51(1), 137-167.

Wurgler, J. and Zhuravskaya, E., 2002. Does arbitrage flatten demand curves for stocks? *Journal of Business* 75(4), 583-608.

Zhang, X., Fuehres, H. and Gloor, P. A., 2011. Predicting stock market indicators through twitter "I hope it is not as bad as I fear". *Procedia-Social and Behavioral Sciences* 26, 55-62.

Zivney, T. L., Bertin, W. J. and Torabzadeh, K. M., 1996. Overreaction to takeover speculation. *Quarterly Review of Economics and Finance* 36(1), 89-115.

Zweig, M. E., 1973. An investor expectations stock price predictive model using closed-end fund premiums. *Journal of Finance* 28(1), 67-78.

Appendix

Table A-1. Summary Statistics on Firm/Trading Level - 5 Days Formation Period

This table reports the summary statistics of the main internet message board (based on a 5 days formation period) and financial control variables. The observations are on a firm-day level. $LogMessages_{i,t}$ is the log transformation $(1+M_t)$, $Bullishness_{i,t}$ is the standardized bullishness index defined in formula (10), $PosSentiment_{i,t}$ and $NegSentiment_{i,t}$ describe the positive and negative sentiment denoted in formula (11) and (12), Agreement is the agreement index described in formula (13), $AbRet_{i,t}$ describes the firm's abnormal return, calculated as the difference of compound raw returns and value-weighted market return, $Volatility_{i,t-30,t-1}$ is the 30-day-standard deviation of returns prior to day *t*, $Upgrade_{i,t}$/$Downgrade_{i,t}$ describe the number of analyst upgrade/downgrade recommendation on day *t*, $PosMeanES_{i,t}$/$NegMeanES_{i,t}$ denote dummy variables for positive/negative mean earnings surprise on day *t*.

	n	Mean	Median	Std. Dev.	10th Pctl.	25th Pctl.	75th Pctl.	90th Pctl.
$AbRet_{i,t-1}$	344,523	-0.003	-0.002	0.062	-0.054	-0.024	0.016	0.047
$AbRet_{i,t-2}$	344,523	-0.002	-0.002	0.062	-0.056	-0.024	0.017	0.050
$LogMessages_{i,t}$	344,523	1.297	1.099	0.703	0.693	0.693	1.609	2.303
$Bullishness_{i,t}$	344,523	1.101	1.099	0.839	0.693	0.693	1.609	2.197
$PosSentiment_{i,t}$	344,523	1.223	1.099	0.733	0.693	0.693	1.609	2.197
$NegSentiment_{i,t}$	344,523	0.136	0.000	0.380	0.000	0.000	0.000	0.693
$Agreement_{i,t}$	344,523	0.925	1.000	0.247	1.000	1.000	1.000	1.000
$Volatility_{i,t-30,t-1}$	344,523	0.046	0.038	0.035	0.017	0.025	0.055	0.079
$Upgrade_{i,t}$	344,523	0.048	0.000	0.435	0.000	0.000	0.000	0.000
$Downgrade_{i,t}$	344,523	0.082	0.000	0.661	0.000	0.000	0.000	0.000
$PosMeanES_{i,t}$	344,523	0.008	0.000	0.087	0.000	0.000	0.000	0.000
$NegMeanES_{i,t}$	344,523	0.002	0.000	0.047	0.000	0.000	0.000	0.000

Table A-2. Summary Statistics on Firm/Trading Level - 10 Days Formation Period

This table reports the summary statistics of the main internet message board (based on a 10 days formation period) and financial control variables. The observations are on a firm-day level. $LogMessages_{i,t}$ is the log transformation ($1+M_t$), $Bullishness_{i,t}$ is the standardized bullishness index defined in formula (10), $PosSentiment_{i,t}$ and $NegSentiment_{i,t}$ describe the positive and negative sentiment denoted in formula (11) and (12), Agreement is the agreement index described in formula (13), $AbRet_{i,t}$ describes the firm's abnormal return, calculated as the difference of compound raw returns and value-weighted market return, $Volatility_{i,t-30,t-1}$ is the 30-day-standard deviation of returns prior to day *t*, $Upgrade_{i,t}$/$Downgrade_{i,t}$ describe the number of analyst upgrade/downgrade recommendation on day *t*, $PosMeanES_{i,t}$/$NegMeanES_{i,t}$ denote dummy variables for positive/negative mean earnings surprise on day *t*.

	n	Mean	Median	Std. Dev.	10th Pctl.	25th Pctl.	75th Pctl.	90th Pctl.
$AbRet_{i,t-1}$	331,655	-0.003	-0.002	0.061	-0.054	-0.024	0.016	0.047
$AbRet_{i,t-2}$	331,655	-0.003	-0.002	0.062	-0.055	-0.024	0.017	0.050
$LogMessages_{i,t}$	331,655	1.295	1.099	0.700	0.693	0.693	1.609	2.303
$Bullishness_{i,t}$	331,655	1.099	1.099	0.837	0.693	0.693	1.609	2.197
$PosSentiment_{i,t}$	331,655	1.220	1.099	0.730	0.693	0.693	1.609	2.197
$NegSentiment_{i,t}$	331,655	0.136	0.000	0.379	0.000	0.000	0.000	0.693
$Agreement_{i,t}$	331,655	0.924	1.000	0.247	1.000	1.000	1.000	1.000
$Volatility_{i,t-30,t-1}$	331,655	0.045	0.038	0.034	0.017	0.025	0.055	0.078
$Upgrade_{i,t}$	331,655	0.049	0.000	0.441	0.000	0.000	0.000	0.000
$Downgrade_{i,t}$	331,655	0.083	0.000	0.661	0.000	0.000	0.000	0.000
$PosMeanES_{i,t}$	331,655	0.008	0.000	0.088	0.000	0.000	0.000	0.000
$NegMeanES_{i,t}$	331,655	0.002	0.000	0.047	0.000	0.000	0.000	0.000

Table A-3. Summary Statistics on Firm/Trading Level - 30 Days Formation Period

This table reports the summary statistics of the main internet message board (based on a 30 days formation period) and financial control variables. The observations are on a firm-day level. $LogMessages_{i,t}$ is the log transformation ($1+M_t$), $Bullishness_{i,t}$ is the standardized bullishness index defined in formula (10), $PosSentiment_{i,t}$ and $NegSentiment_{i,t}$ describe the positive and negative sentiment denoted in formula (11) and (12), Agreement is the agreement index described in formula (13), $AbRet_{i,t}$ describes the firm's abnormal return, calculated as the difference of compound raw returns and value-weighted market return, $Volatility_{i,t-30,t-1}$ is the 30-day-standard deviation of returns prior to day *t*, $Upgrade_{i,t}$/$Downgrade_{i,t}$ describe the number of analyst upgrade/downgrade recommendation on day *t*, $PosMeanES_{i,t}$/$NegMeanES_{i,t}$ denote dummy variables for positive/negative mean earnings surprise on day *t*.

	N	Mean	Median	Std. Dev.	10th Pctl.	25th Pctl.	75th Pctl.	90th Pctl.
$AbRet_{i,t-1}$	283,585	-0.003	-0.002	0.060	-0.053	-0.023	0.015	0.046
$AbRet_{i,t-2}$	283,585	-0.002	-0.002	0.061	-0.054	-0.023	0.017	0.049
$LogMessages_{i,t}$	283,585	1.280	1.099	0.685	0.693	0.693	1.609	2.303
$Bullishness_{i,t}$	283,585	1.082	1.099	0.826	0.693	0.693	1.581	2.079
$PosSentiment_{i,t}$	283,585	1.204	1.099	0.715	0.693	0.693	1.609	2.197
$NegSentiment_{i,t}$	283,585	0.136	0.000	0.378	0.000	0.000	0.000	0.693
$Agreement_{i,t}$	283,585	0.924	1.000	0.247	1.000	1.000	1.000	1.000
$Volatility_{i,t-30,t-1}$	283,585	0.044	0.037	0.032	0.017	0.025	0.054	0.076
$Upgrade_{i,t}$	283,585	0.056	0.000	0.474	0.000	0.000	0.000	0.000
$Downgrade_{i,t}$	283,585	0.096	0.000	0.717	0.000	0.000	0.000	0.000
$PosMeanES_{i,t}$	283,585	0.008	0.000	0.089	0.000	0.000	0.000	0.000
$NegMeanES_{i,t}$	283,585	0.002	0.000	0.048	0.000	0.000	0.000	0.000

Table A-4. Summary Statistics on Firm/Trading Level - 60 Days Formation Period

This table reports the summary statistics of the main internet message board (based on a 60 days formation period) and financial control variables. The observations are on a firm-day level. $LogMessages_{i,t}$ is the log transformation (1+M_t), $Bullishness_{i,t}$ is the standardized bullishness index defined in formula (10), $PosSentiment_{i,t}$ and $NegSentiment_{i,t}$ describe the positive and negative sentiment denoted in formula (11) and (12), Agreement is the agreement index described in formula (13), $AbRet_{i,t}$ describes the firm's abnormal return, calculated as the difference of compound raw returns and value-weighted market return, $Volatility_{i,t-30,t-1}$ is the 30-day-standard deviation of returns prior to day *t*, $Upgrade_{i,t}$/$Downgrade_{i,t}$ describe the number of analyst upgrade/downgrade recommendation on day *t*, $PosMeanES_{i,t}$/$NegMeanES_{i,t}$ denote dummy variables for positive/negative mean earnings surprise on day *t*.

	N	Mean	Median	Std. Dev.	10th Pctl.	25th Pctl.	75th Pctl.	90th Pctl.
$AbRet_{i,t-1}$	224,285	-0.003	-0.002	0.059	-0.053	-0.023	0.015	0.046
$AbRet_{i,t-2}$	224,285	-0.002	-0.002	0.061	-0.054	-0.023	0.017	0.049
$LogMessages_{i,t}$	224,285	1.259	1.099	0.663	0.693	0.693	1.609	2.197
$Bullishness_{i,t}$	224,285	1.059	1.099	0.810	0.649	0.693	1.386	2.079
$PosSentiment_{i,t}$	224,285	1.182	1.099	0.695	0.693	0.693	1.609	2.197
$NegSentiment_{i,t}$	224,285	0.136	0.000	0.374	0.000	0.000	0.000	0.693
$Agreement_{i,t}$	224,285	0.925	1.000	0.248	1.000	1.000	1.000	1.000
$Volatility_{i,t-30,t-1}$	224,285	0.044	0.037	0.032	0.016	0.025	0.053	0.075
$Upgrade_{i,t}$	224,285	0.051	0.000	0.435	0.000	0.000	0.000	0.000
$Downgrade_{i,t}$	224,285	0.091	0.000	0.700	0.000	0.000	0.000	0.000
$PosMeanES_{i,t}$	224,285	0.007	0.000	0.082	0.000	0.000	0.000	0.000
$NegMeanES_{i,t}$	224,285	0.002	0.000	0.046	0.000	0.000	0.000	0.000

Table A-5. Lag-Order Selection Statistics for VAR – Individual Level

This table shows the results of the lag-order selection tests for the VAR analysis and on the individual firm-level. * indicates the lag-order selected by each criterion, where LH-Ratio = Likelihood-Ratio, FPE = Final prediction error, AIC = Akaike information criterion, SBIC = Schwarz information criterion, and HQIC = Hannan-Quinn information criterion.

Lag	CD	J-stats	J p-Value	MBIC	MAIC	MQIC
1	0.5703	309.7144	0.0000	193.7194	285.7144	255.2670
2	0.6340	186.6693	0.0000	109.3393	170.6693	150.3710
3	0.6423*	114.3946*	0.0000	75.7296*	106.3946*	96.2455*

Table A-6. Summary Statistics

This table summarizes the descriptive statistics for section 3.5.2. $SUEAF_{i,t}$ is the difference in actual EPS and forecasted EPS using analyst forecasts 90 days prior to the earnings announcement date scaled by the stock price of the end of the year, $LogMessages_{i,7}$ is the log transformation of (1 + M_t) for the event window [-2,-1], $Bullishness_{i,7}$ is the cumulated bullishness index using formula (12), $Agreement_{i,7}$ is the cumulated agreement index using formula (13). $Return_{50}$ is the compounded return over the period of [-61,-12] and $Return_5$ for the five-day return period [-6,-2] prior to the earnings forecasted EPS using analyst forecasts 90 days prior to the earnings announcement date scaled by the stock price of the end of the year. $LogTurnover_{50}$ is the log-transformed average turnover over the time window [-61,-12] prior to the earnings announcement date, $Volatility_{i,10}$ is the standard deviation of daily returns in the time window [-11,-2] and $LogSize_{i,t}$ is the log-transformation of the market capitalization at the end of the financial year.

Variables	N	Mean	Median	Std. Dev.	10^{th} Pctl.	25^{th} Pctl.	75^{th} Pctl.	90^{th} Pctl.
SUEAF	382	-0.010	0.000	0.198	-0.090	-0.011	0.010	0.060
$LogMessages_{i,t-1,t-2}$	382	1.631	1.386	0.803	0.693	1.099	2.197	2.833
$Bullishness_{i,t-1, t-2}$	382	1.351	1.386	0.944	0.462	0.693	1.946	2.565
$Agreement_{i,t-1, t-2}$	382	0.856	1.000	0.319	0.200	1.000	1.000	1.000
$Return_{50}$	382	-0.032	-0.024	0.370	-0.376	-0.182	0.121	0.298
$Return_5$	382	-0.003	0.000	0.088	-0.100	-0.045	0.038	0.095
$LogTurnover_{50}$	382	13.870	13.550	2.133	11.240	12.270	15.490	16.840
$Volatility_{t-2, t-11}$	382	0.030	0.025	0.018	0.012	0.017	0.037	0.052
$LogSize_{i,t}$	382	24.450	24.110	1.764	22.530	23.310	25.460	26.970

Table A-7. Summary Statistics

This table summarizes the descriptive statistics for section 3.5.2. $SUEAF_{i,t}$ is the difference in actual EPS and forecasted EPS using analyst forecasts 90 days prior to the earnings announcement date scaled by the stock price of the end of the year, $LogMessages_{i,7}$ is the log transformation of (1 + M_t) for the event window [-30,-1], $Bullishness_{i,7}$ is the cumulated bullishness index using formula (12), $Agreement_{i,7}$ is the cumulated agreement index using formula (13). $Return_{50}$ is the compounded return over the period of [-61,-12] and $Return_5$ for the five-day return period [-6,-2] prior to the earnings forecasted EPS using analyst forecasts 90 days prior to the earnings announcement date scaled by the stock price of the end of the year. $LogTurnover_{50}$ is the log-transformed average turnover over the time window [-61,-12] prior to the earnings announcement date, $Volatility_{i,10}$ is the standard deviation of daily returns in the time window [-11,-2] and $LogSize_{i,t}$ is the log-transformation of the market capitalization at the end of the financial year.

Variables	N	Mean	Median	Std. Dev.	10^{th} Pctl.	25^{th} Pctl.	75^{th} Pctl.	90^{th} Pctl.
SUEAF	568	-0.014	0.000	0.179	-0.090	-0.011	0.008	0.045
$LogMessages_{i,t-1, t-30}$	568	2.976	3.045	1.318	1.099	1.946	3.998	4.663
$Bullishness_{i,t-1, t-30}$	568	2.450	2.436	1.515	0.693	1.386	3.496	4.362
$Agreement_{i,t-1, t-30}$	568	0.727	1.000	0.356	0.109	0.425	1.000	1.000
$Return_{50}$	568	-0.017	-0.013	0.332	-0.327	-0.157	0.133	0.291
$Return_5$	568	-0.002	0.000	0.079	-0.088	-0.038	0.035	0.087
$LogTurnover_{50}$	568	13.830	13.550	2.064	11.330	12.240	15.310	16.780
$Volatility_{t-2 ,t-11}$	568	0.028	0.024	0.016	0.012	0.016	0.034	0.050
$Size_{i,t}$	568	24.440	24.190	1.690	22.530	23.340	25.430	26.870

Table A-8. Message Board Activity as Predictor of Earnings Surprise (SUEAF)

Firm- and year-fixed regressions were conducted. T-statistics computed are based on standard errors clustered by firm and year and are denoted in parentheses. $SUEAF_{i,t}$ is the difference in actual EPS and forecasted EPS using analyst forecasts 90 days prior to the earnings announcement date scaled by the stock price of the end of the year, $LogMessages_{i,7}$ is the log transformation of (1 + M_t) for the event window [-30,-1], $Bullishness_{i,7}$ is the cumulated bullishness index using formula (12), $Agreement_{i,7}$ is the cumulated agreement index using formula (13), High/LowAgree$D_{i,7}$ is a dummy variable indicating the cumulated agreement index to be in the top/bottom 20-percentile, $Return_{50}$ is the compounded return over the period of [-61,-12] and $Return_5$ for the five-day return period [-6,-2] prior to the earnings announcement date. $LogTurnover_{50}$ is the log-transformed average turnover over the time window [-61,-12] prior to the earnings announcement date, $Volatility_{i,10}$ is the standard deviation of daily returns in the time window [-11,-2] and $LogSize_{i,t}$ is the log-transformation of the market capitalization at the end of the financial year. ***, **, and * describe significance at 1%, 5% and 10% level, respectively.

	(1)	(2)	(3)	(4)	(5)
	$SUEAF_{t0}$	**$SUEAF_{t0}$**	**$SUEAF_{t0}$**	**$SUEAF_{t0}$**	**$SUEAF_{t0}$**
$LogMes_{i,t-30,t-1}$	0.001	0.004	-0.005	-0.008	
	(0.008)	(0.008)	(0.011)	(0.013)	
$BullInd_{i,t-30,t-1}$	-0.014**	-0.013**	-0.018**	-0.019**	
	(0.006)	(0.006)	(0.008)	(0.008)	
$PosSentiment_{i,t-30,t-1}$					-0.009
					(0.009)
$NegSentiment_{i,t-30,t-1}$					0.010
					(0.010)
$AgreeInd_{i,t-30,t-1}$	0.051*		0.025		
	(0.029)		(0.043)		
High_Agree_$D_{i,t}$		0.051**		-0.013	0.055*
		(0.024)		(0.050)	(0.030)
Low_Agree_$D_{i,t}$		0.007		-0.025	0.012
		(0.021)		(0.028)	(0.021)
BullInd x $AgreeInd_{i,t-30,t-1}$			0.012		
			(0.012)		
BullInd x High_Agree_$D_{i,t-30,t-1}$				0.022	
				(0.016)	
BullInd x Low_Agree_$D_{i,t-30,t-1}$				0.013	
				(0.014)	
$Return_{50}$	-0.054**	-0.053**	-0.054**	-0.052**	-0.054**
	(0.023)	(0.024)	(0.023)	(0.023)	(0.024)
$Return_5$	0.355*	0.334*	0.350*	0.322*	0.332*
	(0.197)	(0.188)	(0.194)	(0.187)	(0.188)
$LogTurnover_{50}$	0.015	0.017	0.016	0.019	0.015
	(0.016)	(0.017)	(0.016)	(0.017)	(0.016)
$Volatility_{i,t-11,t-2}$	-1.574*	-1.634*	-1.488*	-1.554*	-1.671*
	(0.880)	(0.877)	(0.862)	(0.861)	(0.894)
$Size_{i,t}$	0.018	0.017	0.018	0.014	0.018
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
Constant	-0.626	-0.622	-0.606	-0.515	-0.646
	(0.404)	(0.397)	(0.412)	(0.409)	(0.400)
Observations	568	568	568	568	568
Adjusted R-squared	11.5%	12.3%	11.5%	12.5%	12.0%

Table A-9. Message Board Activity as Predictor of Earnings Surprise (SUEHIST)

Firm- and year-fixed regressions were conducted. T-statistics computed are based on standard errors clustered by firm and year and are denoted in parentheses. $SUEHIST_{i,t}$ is the difference in actual EPS in year *t* and the previous year actual EPS scaled by the stock price of the end of the year, $LogMessages_{i,7}$ is the log transformation of (1 + M_t) for the event window [-30,-1], $Bullishness_{i,7}$ is the cumulated bullishness index using formula (3), $Agreement_{i,7}$ is the cumulated agreement index using formula (4), $High/LowAgreeD_{i,7}$ is a dummy variable indicating the cumulated agreement index to be in the top/bottom 20-percentile, $Return_{50}$ is the compounded return over the period of [-61,-12] and $Return_5$ for the five-day return period [-6,-2] prior to the earnings announcement date. $LogTurnover_{50}$ is the log-transformed average turnover over the time window [-61,-12] prior to the earnings announcement date, $Volatility_{i,10}$ is the standard deviation of daily returns in the time window [-11,-2] and $LogSize_{i,t}$ is the log-transformation of the market capitalization at the end of the financial year. ***, **, and * describe significance at 1%, 5% and 10% level, respectively.

	(1) $SUEHIST_{t0}$	(2) $SUEHIST_{t0}$	(3) $SUEHIST_{t0}$	(4) $SUEHIST_{t0}$	(5) $SUEHIST_{t0}$
$LogMes_{i,t-30,t-1}$	-0.158	-0.122	-0.150	-0.266	
	(0.155)	(0.144)	(0.163)	(0.263)	
$BullInd_{i,t-30,t-1}$	0.017	-0.045	0.022	-0.115*	
	(0.105)	(0.080)	(0.115)	(0.060)	
$PosSentiment_{i,t-30,t-1}$					-0.146
					(0.089)
$NegSentiment_{i,t-30,t-1}$					-0.035
					(0.184)
$AgreeInd_{i,t-30,t-1}$	-0.077		-0.041		
	(0.256)		(0.303)		
$High\_Agree\_D_{i,t}$		-0.239		-0.980	-0.271
		(0.315)		(0.975)	(0.420)
$Low\_Agree\_D_{i,t}$		-0.442		-0.784	-0.391
		(0.300)		(0.724)	(0.275)
BullInd x $AgreeInd_{i,t-30,t-1}$			-0.016		
			(0.103)		
BullInd x $High\_Agree\_D_{i,t-30,t-1}$				0.268	
				(0.253)	
BullInd x $Low\_Agree\_D_{i,t-30,t-1}$				0.144	
				(0.307)	
$Return_{50}$	-0.445*	-0.424*	-0.446*	-0.400*	-0.422*
	(0.232)	(0.232)	(0.229)	(0.227)	(0.236)
$Return_5$	1.226**	1.326**	1.225**	1.213**	1.327**
	(0.511)	(0.571)	(0.512)	(0.503)	(0.592)
$LogTurnover_{50}$	-0.134	-0.130	-0.135	-0.111	-0.134
	(0.135)	(0.136)	(0.137)	(0.126)	(0.131)
Volatilityi,$_{t-11,t-2}$	11.054	12.028	10.973	13.028	12.238
	(8.704)	(9.001)	(8.846)	(9.522)	(9.324)
$Size_{i,t}$	0.220*	0.218*	0.220*	0.205*	0.224*
	(0.120)	(0.123)	(0.120)	(0.117)	(0.122)
Constant	-3.443**	-3.266**	-3.462**	-2.483	-3.379**
	(1.398)	(1.449)	(1.421)	(1.799)	(1.477)
Observations	667	667	667	667	667
Adjusted R-squared	6.1%	7.2%	6.0%	8.2%	7.0%

Table A-10. Average CARs Around Bid Announcements – Takeover Subsample

This table presents the average CARs based on the market model. We apply the standardized cross-sectional Boehmer-test to test for significance. ***, **, and * describe significance at 0.1%, 1% and 5% level, respectively.

Window		*Panel A*	*Panel B*	*Panel C*	*Panel D*	*Panel E*
[-t]	**[+t]**	**Full sample**	**News & HC**	**News only**	**HC only**	**No media**
-40	10	7.46%***	7.75%**	-4.62%	8.23%***	2.42%
-20	10	6.34%***	6.89%***	0.31%	6.69%***	2.62%
-10	10	5.92%***	6.99%***	-1.02%	5.85%***	3.66%
-3	3	3.27%***	3.84%***	-1.98%	3.54%***	-0.17%
-2	2	3.10%***	2.93%***	-2.15%	3.57%***	0.90%
-1	1	2.70%***	2.28%***	-3.04%	3.07%***	2.35%**
0	0	1.49%***	1.73%***	-1.92%	1.52%***	1.03%**
-1	-1	0.56%***	0.47%*	-0.45%	0.60%***	0.74%
-2	-2	0.56%***	0.64%**	0.16%	0.56%**	0.34%
-3	-3	0.31%	0.41%	0.25%	0.27%	0.26%
-4	-4	0.27%	0.66%**	-0.25%	0.16%	-0.18%
-5	-5	0.40%*	0.42%	0.53%	0.44%*	-0.05%
1	1	0.65%**	0.08%	-0.67%	0.95%**	0.58%
2	2	-0.16%	0.01%	0.73%***	-0.05%	-1.79%**
3	3	-0.14%	0.49%	-0.08%	-0.30%	-1.33%**
4	4	0.23%**	0.41%**	0.73%	0.09%	0.57%
5	5	-0.04%	0.07%	0.29%	0.07%	-1.40%**
-10	-1	3.36%***	3.48%***	1.20%	3.06%***	5.73%***
-20	-11	0.42%	-0.11%	1.33%	0.83%	-1.04%
-30	-21	0.50%	0.95%	-5.44%	0.53%	-0.28%
-40	-31	0.62%	-0.08%	0.50%	1.01%	0.08%
-40	-1	4.90%***	4.23%	-2.40%	5.44%***	4.49%
#Events		352	103	5	217	27

Table A-11. Description of Main Variables

This table provides an overview of the variables applied to the empirical analysis conducted in chapter 4.

Category	Variable	Description	Source
Attention	$FullMedia_{-40}$	Indicator variable: one if any news article and messages on the investment platform HotCopper occurred about the target firm on the investment platform HotCopper in the period beginning 40 days until 1 day before the bid announcement date	Thomson Reuters News Analytics; HotCopper
	$News_{-40}$	Indicator variable: one if any news article and no other messages on the investment platform HotCopper occurred about the target firm on the investment platform HotCopper in the period beginning 40 days until 1 day before the bid announcement date	Thomson Reuters News Analytics
	$SocialMedia_{-30}$	Indicator variable: one if at least 10 messages and no other news article occurred about the target firm on the investment platform HotCopper in the period beginning 30 days until 1 day before the bid announcement date	HotCopper
	ASX_{-30}	Indicator variable: one if any official ASX announcement occurred in the period beginning 30 days until 1 day before the bid announcement date	HotCopper
	$AnUp_{-30}$	Indicator variable: one if any analyst upgrade occurred in the period beginning 30 days until 1 day before the bid announcement date	I/B/E/S
	$AnDown_{-30}$	Indicator variable: one if any analyst downgrade occurred in the period beginning 30 days until 1 day before the bid announcement date	I/B/E/S
	$Google1_{i,t}$	Indicator variable: one if any daily change in Google search volume fulfills the outlier criteria in the period beginning 30 days until 1 day before the bid announcement	
	$Google2_{i,t}$	Indicator variable: one if any daily abnormal Google search volume fulfills the outlier criteria in the period beginning 30 days until 1 day before the bid announcement	
Target	$LnSize_{i,t}$	Natural logarithm of the target's market capitalization at the beginning of the month	Compustat
	$Market\text{-}to\text{-}book_{i,t}$	Firms market capitalization at the beginning of the month divided by the firm's book value of equity	Compustat
	$EBITDA_{i,LTM}$	Firm's EBITDA of last twelve months scaled by latest total assets	Compustat
Deal	International	Indicator variable: one if the acquirer's and target's headquarters are located in different countries, zero otherwise	SDC Platinum
	Diversification	Indicator variable: one if acquirer and target do not share the same 2-digit SIC code, zero otherwise	SDC Platinum

	Hostile	Indicator variable:one if the bid is hostile, zero otherwise	SDC Platinum
	Toehold	Indicator variable: one if the acquirer owns a non-zero percentage of the target’s stock prior to announcement date,zero otherwise	SDC Platinum
	Takeover	Indicator variable: one if acquirer's ownership would exceed 50% after the deal , zero otherwise	SDC Platinum
	Tender	Indicator variable: one for tender offers, zero otherwise	SDC Platinum
Acquirer	PublicAcq	Indicator variable: one if the acquirer is publicly listed	SDC Platinum

Table A-12. Pearson Correlation for Factors

This table summarizes the Pearson rank correlations between widely-accepted risk factors. The risk factors include premiums based on market (MkrRf), size (SMB), book-to-market (HML), Carhart momentum (MOM), Pastor-Stambaugh liquidity (PS LIQ), profitability (RMW), investments (CMA) and news sentiment (PNM). ***, **, and * describe significance at 1%, 5% and 10% level, respectively.

	PMN	*MktRF*	*SMB*	*HML*	*Mom*	*Trade-Liq*	*RMW*
MktRF	-0.34***						
SMB	-0.49***	0.26***					
HML	0.08	-0.03	-0.07				
Mom	0.32***	-0.34***	0.04	-0.12**			
TradeLiq	0.00	0.10**	0.18***	-0.11	0.06		
RMW	0.59***	-0.48***	-0.54***	0.42***	0.15*	-0.08	
CMA	0.09	-0.24***	0.05	0.59***	0.16	-0.11*	0.24***

Table A-13. Stock Returns and Sentiment: Predictive Return Differentials

This Table extends Table 5-6 and shows the average monthly raw returns sorted by negative, neutral and positive sentiment. Average raw returns are denoted in percentages. In each month, we divide our sample stocks into negative, neutral or positive sentiment bins. Sentiment is measured by the Ravenpack Composite Sentiment Score (CSS) based on categories developed according to more than 330 market moving events. Return differentials of positive and negative sentiment bins are calculated based on equal-weighted average next month returns of the sentiment portfolios. Return differentials are shown for subsamples of firms sorted by operative profitability and investments in the fiscal year after the formation period.

	Average monthly returns (%)					**Share of sent. in obs. (%)**		
	Sentiment				t-Statistics	Sentiment		
	Neg.	Neut.	Pos.	Pos - Neg.	Pos - Neg.	Neg.	Neut.	Pos.
All stocks	0.51	0.70	1.38	0.87	20.45	0.15	0.55	0.30
Panel F: Profitability in t+1								
1	0.37	0.81	1.64	1.27	8.67	0.19	0.51	0.30
2	0.86	1.25	1.61	0.75	9.07	0.19	0.38	0.43
3	0.75	1.20	1.47	0.72	10.21	0.20	0.35	0.46
RMW(3-1)	0.37	0.38	-0.17					
Panel H: Investment in t+1								
1	0.23	0.48	1.05	0.82	11.92	0.18	0.41	0.40
2	0.63	0.99	1.35	0.72	12.83	0.18	0.43	0.39
3	1.09	1.47	2.10	1.01	10.52	0.20	0.46	0.34
CMA(1-3)	-0.86	-0.98	-1.05					

Table A-14. Factor Time Series Regressions - Only Positive Sentiment

This table reports the results of the factor-based time-series regressions on portfolios that only go long on stocks with average positive sentiment in the previous month. In each month, a stock's sentiment is classified as positive if the average Ravenpack Composite Sentiment Score exceeds 50. The portfolios are formed at the beginning of the month and held for one month. The long-only portfolio returns are regressed on different widely-accepted risk-factors. T-statistics are in parenthesis. ***, **, and * describe significance at 1%, 5% and 10% level, respectively.

	Model 1	*Model 2*	*Model 3*	*Model 4*	*Model 5*	*Model 6*
	CAPM	FF 3-Factor	Carhart 4-Factor	PS Liq. 5-factor	FF 5-Factor	All 7-Factor
	PMN	PMN	PMN	PMN	PMN	PMN
Mkt - rf	1.1036***	1.0085***	0.9176***	0.9078***	0.9840***	0.9067***
	(30.2424)	(38.7141)	(37.8141)	(36.9815)	(31.5824)	(32.2689)
SMB		0.5479***	0.5988***	0.5785***	0.5413***	0.5765***
		(15.2158)	(19.2712)	(18.1655)	(12.7488)	(15.1061)
HML		0.1017***	0.0496*	0.0531*	0.1582***	0.0546
		(2.9591)	(1.6695)	(1.7637)	(3.2483)	(1.2357)
MOM			-0.1736***	-0.1786***		-0.1784***
			(-9.1023)	(-9.3311)		(-9.0341)
TradeLiq				0.0958***		0.0963***
				(3.5982)		(3.5376)
RMW					-0.0456	-0.0046
					(-0.8543)	(-0.0965)
CMA					-0.1015	0.0008
					(-1.4639)	(0.0136)
Intercept	0.0095***	0.0075***	0.0082***	0.0079***	0.0080***	0.0080***
	(5.9378)	(6.7373)	(8.6410)	(8.1279)	(6.8374)	(7.7801)
Observations	215	215	215	204	215	204
R^2	0.81	0.91	0.94	0.94	0.91	0.94

Table A-15. Factor Time Series Regressions - Only Negative Sentiment

This table reports the results of the factor-based time-series regressions on portfolios that only short stocks with average negative sentiment in the previous month. In each month, a stock's sentiment is classified as negative if the average Ravenpack Composite Sentiment Score is lower than 50. The portfolios are formed at the beginning of the month and held for one month. The short-only portfolio returns are regressed on different widely-accepted risk-factors. T-statistics are in parenthesis. ***, **, and * describe significance at 1%, 5% and 10% level, respectively.

	Model 1	*Model 2*	*Model 3*	*Model 4*	*Model 5*	*Model 6*
	CAPM	FF 3-Factor	Carhart 4-Factor	PS Liq. 5-factor	FF 5-Factor	All 7-Factor
	PMN	PMN	PMN	PMN	PMN	PMN
Mkt - rf	1.2692***	1.1255***	0.9773***	0.9787***	1.0021***	0.8996***
	(24.2829)	(30.9898)	(31.2349)	(30.6695)	(24.6769)	(26.3397)
SMB		0.8093***	0.8922***	0.8724***	0.6679***	0.7429***
		(16.1212)	(22.2707)	(21.0703)	(12.0689)	(16.0155)
HML		0.0695	-0.0153	-0.0265	0.2705***	0.1037*
		(1.4509)	(-0.4002)	(-0.6775)	(4.2612)	(1.9308)
MOM			-0.2829***	-0.2833***		-0.2661***
			(-11.5051)	(-11.3849)		(-11.0845)
TradeLiq				0.0432		0.0717**
				(1.2481)		(2.1663)
RMW					-0.3909***	-0.3047***
					(-5.6241)	(-5.3077)
CMA					-0.1481	-0.0189
					(-1.6393)	(-0.2547)
Intercept	0.0011	-0.0016	-0.0004	0.0001	0.0011	0.0016
	(0.4603)	(-1.0116)	(-0.3627)	(0.0601)	(0.7108)	(1.3137)
Observations	215	215	215	204	215	204
R^2	0.73	0.88	0.93	0.93	0.90	0.94

Table A-16. Time Series Regressions by Firm Characteristics

This Table extends Table 5-9 and reports the intercepts (alphas) of the factor-based time-series regressions on equal-weighted average returns of portfolios that goes long on stocks with average positive and shorts stocks with average negative sentiment in the previous month sorted by different firm characteristics. The zero-investment portfolios are regressed on excess market returns (CAPM), the Fama-French (1993) three-factor, Carhart (1997) Momentum-factor, the Pastor-Stambaugh (2003) liquidity-factor, the Fama-French (1995) five-factor and the combination of all factor-models. The equally-weighted portfolios are formed at the beginning of the month and held for one month. T-statistics are in parenthesis. ***, **, and * describe significance at 1%, 5% and 10% level, respectively.

CAPM	FF 3-Factor	Carh. 4-Factor	PS 5-factor	FF 5-Factor	All 7-Factor
Panel F: Trading Volume					
Illiquid					
0.0056***	0.0059***	0.0058***	0.0056***	0.0042***	0.0042***
(4.5741)	(5.0583)	(4.9708)	(4.7914)	(3.7166)	(3.6423)
Medium					
0.0031*	0.0040***	0.0035**	0.0032**	0.0019	0.0017
(1.9646)	(2.7351)	(2.5454)	(2.3102)	(1.3113)	(1.2629)
Liquid					
0.0095***	0.0101***	0.0098***	0.0095***	0.0082***	0.0081***
(6.0716)	(6.8292)	(6.7631)	(6.5752)	(5.5428)	(5.5658)

Table A-17. Stock Returns and Sentiment: Predictive Return Differentials by Quintiles

This table extends Table 5-10 and shows the average monthly raw returns sorted by sentiment quintiles. Average raw returns are denoted in percentages. In each month, we divide our sample stocks according to sentiment quintile bins. Sentiment is measured by the Ravenpack Composite Sentiment Score (CSS) based on categories developed according to more than 330 market moving events. Return differentials between the highest and lowest sentiment quintile bins are calculated based on equal-weighted average next month returns of the sentiment quintile portfolios. Return differentials are shown for subsamples of firms sorted by operative profitability and investments for the fiscal year after the formation period.

	Average monthly returns (%)							**Share of sent. (%)**				
	Sentiment						t-Stat.	Sentiment				
	Q1	Q2	Q3	Q4	Q5	Q5-Q1	Q5-Q1	Q1	Q2	Q3	Q4	Q5
All stocks	0.52	0.17	0.90	1.24	1.46	0.94	19.43	15.08	15.28	40.15	10.02	19.47
Panel F: Profitability in t+1												
1	0.38	-0.32	1.25	1.24	1.89	1.51	8.85	19.01	13.85	37.50	11.20	18.45
2	0.84	0.30	1.61	1.52	1.67	0.83	9.01	19.26	10.41	27.89	15.59	26.84
3	0.74	-0.04	1.62	1.25	1.58	0.85	10.79	19.44	8.77	26.32	15.79	29.67
RMW	0.36	0.27	0.37	0.01	-0.31							
Panel H: Investment in t+1												
1	0.21	-0.20	0.75	1.00	1.07	0.86	11.38	17.89	11.06	30.81	14.61	25.62
2	0.62	0.19	1.33	1.27	1.39	0.77	12.12	17.59	12.57	31.06	14.63	24.15
3	1.08	0.30	1.90	1.78	2.30	1.22	10.95	20.09	12.17	34.18	12.76	20.80
CMA	-0.87	-0.50	-1.15	-0.78	-1.23							