# INFORMATION IN THE FINANCIAL NEWS: EFFECT OF MARKET COMMENTARY ON STOCK MARKET PERFORMANCE

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF SOCIAL SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

AYNUR GİRAY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF BUSINESS ADMINISTRATION
IN
THE DEPARTMENT OF BUSINESS ADMINISTRATION

SEPTEMBER 2012

Approval of the Graduate School of Social Sciences

Prof. Dr. Meliha Altunışık
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Business Administration.

Assoc. Prof. Dr. Engin Küçükkaya
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Business Administration.

Assist. Prof. Dr. Seza Danışoğlu
Supervisor

**Examining Committee Members**

| | | |
|---|---|---|
| Prof. Dr. Z. Nuray Güner | (METU, BA) | |
| Assist. Prof. Dr. Seza Danışoğlu | (METU, BA) | |
| Assoc. Prof. Dr. Süheyla Özyıldırım | (Bilkent U., MAN) | |

**I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.**

Name, Last name: Aynur Giray

Signature:

# ABSTRACT

INFORMATION IN THE FINANCIAL NEWS:
EFFECT OF MARKET COMMENTARY ON STOCK MARKET PERFORMANCE

GIRAY, Aynur
MBA, Department of Business Administration
Supervisor: Dr. Seza Danışoğlu

September 2012, 73 pages

This paper studies the effect of investment sentiment on asset prices. A sentiment proxy is calculated by performing content analysis on the Wall Street Journal's 'Heard on the Street' columns. This proxy is extracted by the principal component analysis of the word tags from the Harvard psychological dictionary that is used by the content analysis software General Inquirer. The relationship between stock prices, trading volume and the media sentiment proxy is estimated within the VAR context. Results suggest that stock price and trading volume are affected by the media sentiment factor. Findings also imply that stock prices and trading volume in the current time period are mainly affected by the past returns and volume.

Keywords: Behavioral Finance, Investor Sentiment, Media Factor

# ÖZ

FİNANS HABERLERİNİN DEĞERİ: FİNANSAL YORUMLARIN
PİYASA DEĞERİ
ÜZERİNDEKİ ETKİSİ

GIRAY, Aynur
Y. Lisans, İşletme Bölümü
Tez Yöneticisi: Dr. Seza Danışoğlu

Eylül 2012, 73 sayfa

Bu çalışma yatırımcı duyarlılığının piyasa değeri ve piyasa yoğunluğu üzerindeki etkisini incelemektedir. Yatırımcı duyarlılığı medya değişkeni üzerinden tanımlanmıştır. Medya değişkeni Amerika ve Avrupa'da yayınlanan Wall Street Journal gazetesinde bulunan 'Heard on the Street' sütununun içeriği ile oluşturulmuştur. Medya değişkeni için Harvard Psikoloji sözlüğü ile oluşturulan metin analizi programı kullanılmıştır. Devamında yapılan temel birleşen analizi ile duyarlılık değişkeni oluşturulmuştur. Piyasa değeri ve piyasa yoğunluğu ile medya değişkeninin ilişkisi VAR analiz yönetimi ile incelenmiştir. Çalışma sonuçları, medya değişkeninin piyasa değeri ve piyasa yoğunluğu üzerinde etkisi olduğunu göstermiştir. Ancak hisse senedi fiyatlarının ve piyasa hacminin güncel değerleri en çok kendi geçmiş değerlerinden etkilenmektedir.

Anahtar Kelimeler: Davranışsal Finans, Yatırımcı Eğilimi, Medya Etkisi

To my parents

# TABLE OF CONTENTS

## LIST OF TABLES

TABLES

## LIST OF FIGURES

FIGURES

# CHAPTER 1

# INTRODUCTION

Traditionally, a formal education in finance has dismissed the idea that one's personal psychology can be a deficiency in making good investment decisions. From the rational perspective of finance, investors should make only decisions that are in their best interest. They should able to discern among all the options facing them and accurately compute their value including short term and long-term value and choose the options that maximize their best interest.

In behavioral finance, it is not assumed that people are perfectly sensible; instead this field studies people's irrational behavior and its affect on financial decisions. Behavioral finance researchers want to understand human frailty and to find more compassionate and realistic ways for people to avoid temptation, exert more self-control and ultimately reach their long-term financial goals.

As it is well-known, social media, news media and conversations with other investors are the main sources of learning while making investment decisions in today's financial markets. Recent empirical studies have heightened the debate about the importance of public information arrival in asset pricing. Beginning with Fama (1970), an efficient capital market is characterized as one in which security prices fully reflect all available information. It has been shown that prices may not respond quickly to new information and many studies measure the efficiency of a market with the speed with which prices change upon news arrival. The weakness of the efficient markets hypothesis is that more often than not one cannot identify what news has caused the asset price to change. The price seems to fluctuate up or down even when there is no news.

This study examines the relationship between the content of financial news in the media and stock market activity by focusing on WSJ Europe's *"Heard On the Street"* articles. The sample consists of approximately 6,982 articles that were published in the WSJ Europe and US between 2000 and 2011. 5,353 articles are from the WSJ US, and, 1629 texts are from the WSJ Europe. A textual-analysis software program is used to analyze the contents of each article and evaluate the reflections of language usage in the movement of financial markets.

The significance of soft information in the language is increasing in value when it is researched in the market, a fact that is evidenced by the increase in the linguistic algorithms used by practitioners in the market.

The following variables will be examined in the course of this research:

(1) The intensity of language usage in the daily newspaper column

(2) Market response, in the form of movements in a predetermined stock index that represents the largest market capitalization stocks

This study reveals media's linkage to financial markets. Stock returns in the US market seem to be more sensitive to news compared to the European markets. . Also, there is evidence that for both the US and European market, the main driver of stock market prices are the past price trends of the market.

In this study background literature is reviewed in Chapter 2. Information about the sample, data collection and statistical methodology is discussed in Chapter 3. Chapter 4 presents the study's findings, and Chapter 5 concludes the study.

# CHAPTER 2

# LITERATURE REVIEW AND THEORETICAL BACKGROUND

## 2.1. Efficient Markets Hypothesis

The efficient markets hypothesis is a widely researched topic by financial economists. It is generally believed that security markets are efficient in reflecting information about individual stocks and the stock market as a whole. The belief is that when information arises, the news spreads very quickly and it is incorporated into the prices of securities without any delay. In an efficient market, arbitrage opportunities do not exist; hence, investors cannot earn abnormal returns through technical or fundamental analysis.

The efficient markets hypothesis was -and still is- seen as a cornerstone of the modern finance theories. Proponents of the hypothesis claim that the participants in a market act rationally given what they know about a company's position and prospects. At the heart of the concept there is the market, and, being the accumulated activities of its many participants, it rapidly assimilates any information about a company as soon as the information becomes available. Hence, stock prices generally reflect all available information about the firms.

Prof. Burton Malkiel of Princeton University whose book “A Random Walk down Wall Street” was first published in 1973 popularized the theory. The efficient markets hypothesis is associated with the idea of a *“random walk”,* which is a term loosely used in the finance literature to characterize a price series where all subsequent price changes represent random departures from previous prices. The logic of the random walk idea is that if the flow of information is unimpeded and information is immediately reflected in stock prices, then tomorrow’s price change will reflect only tomorrow’s news and will be independent of the price changes today.  However, news is, by definition, unpredictable. Thus, price changes are expected to be unpredictable and random. One conclusion of the theory is that professional stock market analysts should not be able to pick stocks that do any better on average than the market as a whole since the market has already incorporated all available information into the price of a security. Efficient markets hypothesis is sometimes expressed by the proposition that the recommendations of stock market professionals should do no better than throwing darts at pages of the Wall Street Journal. This vivid popularization of the theory has given rise to a regular feature in newspapers like Wall Street Journal.

Most of the finance theories are based on two main assumptions: people make (1) rational decisions, and, (2) unbiased predictions about the future. Various pricing models have been developed in the literature based on these assumptions. Models such as the capital asset pricing model, the arbitrage pricing model, the option pricing model are all based on these assumptions of rationality. The empirical testing of these models has provided evidence that investors do not always make rational decisions. Following these findings, the influence of human psychology and emotions on financial decisions has started to be studied. The major contribution in this area comes from the 2002 Nobel Prize Laureates in Economics, psychologists Daniel Kahnemann and Vernon Smith. This study investigates the influence of human psychology around external world. People are making decisions in social their environment. Based on this fact, several studies have been conducted  a link between soft information like news media, TV programmes, radio, analyst's recommendation to stock returns, investor forums.

Recent literature has focused on the market reaction to the cognitive psychology of investors. Rational models are developed where heuristics of naïve investors are identified. Meanwhile, behavioral finance researchers suggest that investors are dramatically influenced by other people who raise their opinion about the market conditions (Hong, Kubik, and Stein, 2004), and also market commentators in newspapers (Huberman and Ragev, 2001; Liang, 1999), and the opinion as it is exposed on television (Busse and Green, 2002). In these studies, it is argued that in order to predict the stock market price changes, it is crucial to understand the psychology of the influencer together with the conjecture of investors.

Several prior studies mention that abnormalities in pricing and trading volume could be viewed as a proxy to the investor sentiment factor. As a result, researchers started using sentiment indicators for analyzing the change in market returns. Baker and Wurgler (2006) have documented that mispricing is a result of both an uninformed demand shock and a limit on arbitrage, where investor sentiment stands as the main reason behind the mispricing

The main line of literature on investor behavior has started with Barberis, Shleifer and Vishny (BSV) in 1998. They present the model of a representative investor who has noisy signals of reliable information. They mention in their work that:

> *"Recent empirical research in finance has uncovered two families of pervasive regularities: under-reaction of stock prices to news such as earnings announcements, and overreaction of stock prices to a series of good or bad news. In this paper, we present a parsimonious model of investor sentiment, or of how investors form beliefs, which is consistent with the empirical findings. The model is based on psychological evidence and produces both under-reaction and overreaction for a wide range of parameter values."*

Rational investors, who are aware of these abnormalities, can earn superior returns without bearing extra risk. Superior returns in this context can be gained through buying the losing stocks and selling the earning ones. BSV find that noisy information causes prices to underreact to reliable information and overreact to unreliable information. The finding that investors do not always react in a corresponding manner to the new information is evidence of the departure from efficient market theories. Based on their research where they utilize experimental psychology, De Bondt and Thaler (1985) suggest that overreaction occurs mainly when unexpected and dramatic news events occur. Shefrin (2000) has collected previous research and raised the following argument:

> *"What we seem to have is overreaction at very short horizons, say less than one month (Lehmann, 1990), momentum possibly due to underreaction for horizons between three and twelve months (Jegadeesh and Titman 1993)*

*and overreaction for periods longer than one year (De Bondt and Thaler 1985, 1987, 1990)."*

The conceptual study of Cutler et al (1989) can be considered to be the first one to identify a link between news coverage and stock prices. They investigate the non-economic and macroeconomic events which are listed as "New York Times Lead Story". Study demonstrates that news is important to account for important market moves; however it is still difficult to explain the variance on the prices basis of publicly available information.

### 2.2. Impact of Media on Investor Behavior

Shiller (2000) argues that the news media plays an important role in setting the stage for market moves and provoking them. His conjecture is that investors follow the printed word even though much of it is pure hype, suggesting that market sentiment is driven by the content of the news.

In his work 'Irrational *Exuberance',* Shiller states the following:

> *"Every other business, news media need to perpetually gain attention of their customers in order to survive and be successful. For capturing ongoing interest of their customers, news media should create interesting news stories that will be effective to wide majority of public. Hence, financial market news contains a valuable source of information. Firstly, this is a provider to ongoing and daily new information source. Also, it has a star quality; financial markets capture great public attention as investors may gain fortunes."*

Another study conducted by Merton (1987) also demonstrates that the media impact facilitates investors' awareness. He argues that investors will buy and hold only those securities about which they have information. Barber and Odean (2003) provide direct evidence that individual investors tend to buy stocks that are in the news. In his study Klibano (1998) shows that country-specific news reports on the front page of New York Times affect the pricing of closed-end country funds. Later, Huberman and Regev (2001) also find that an article in the Financial Times on a biochemical firm caused prices of that company to soar. Antweiler and Frank

(2004) consider the influence of internet stock message boards. They investigate the 1.5 million messages posted on Yahoo Finance and Dow Jones internet indexes and find that stock messages predict market volatility, but their effect on returns is small.

DeLong et al (1990) are among the first researchers to find that investment decisions are affected from investors' sentiment. In their model, two sets of traders exist: *professional arbitrageurs* and *unsophisticated traders,* i.e. noise traders. According to their findings, the prevailing risk in the market is created by the unpredictability of the noise traders. Professional arbitrageurs respond to the behavior of noise traders rather than acting on fundamentals. By doing so, professional arbitrageurs do not only consider pseudo signals such as volume and price patterns but also the sentiment indices.

Fischer Black first posited the idea of noise trading in 1986. He defines the noise traders to be the ones that *"trade on noise as if it were information… Noise makes financial markets possible, but it also makes them imperfect. If there is no noise trading, there will be very little trading in individual assets"*. According to Black, most of the noise traders behave as a group. Shleifer (2000) notes *"investor sentiment reflects the judgment errors made by a substantial number of investors, rather than uncorrelated random mistakes."* According to Shleifer, for mass movements to happen two conditions should be present. First, there must be limits for informed investors to trade. This case is the condition that is argued by Shleifer and Summers (1990) by giving the reason as informed traders who wish to profit from their information face risk that eventually limits their trading volume. Second, there must be systematic trading by individuals. Barber (2009) finds strong evidence of systematic trading by analyzing data obtained from two large discount and retail brokers. Likewise, De Long, Shleifer, Summers, and Waldmann (1990) explain systematic movements of investors by noise trading and they argue that certain uninformed traders tend to strategically act on noisy signals, and therefore, their trading can affect prices in a systematic way. Mentioned occurrences are considered to cause asset prices to deviate from fundamental values when they are unusually bullish or bearish.

Interaction and impact of investor sentiment on stock price volatility are usually considered to create four effects: trade by bullish or bearish investors, price deviations from fundamentals, deviation of noise traders’ misperceptions about risk, and, adjustment in the market risk due to changes in noise traders’ demand of stocks based on their sentiment. Barberis et al (1998) model investor sentiment and show that news can cause both over- and under-reaction to stock prices.

Baker and Wurgler (2007) argue that the key issue nowadays for researchers is to find out how to measure investor sentiment and quantify its effects. They describe various possible proxies to measure investor sentiment. These can be listed as retail investor trades, mutual fund flows, trading volume, dividend premium, closed-end fund discount, option implied volatility, IPO first-day returns, IPO volume, equity issues over total new issues, and insider trading. As authors point out, it is no longer questionable whether sentiment affects investors, thus, stock returns; but the appropriate measure of investor sentiment must be developed.

Chan (2003) finds evidence of a post-news drift. Author proposes that investors underreact to new information and there is a persistent effect of Reuters Sentiment on asset prices that seems to be the strongest after bad news is released. In another study, Tetlock (2011) tests whether investors distinguish between old (stale) and new information about firms. A firm's return on the day of stale news negatively predicts its return in the following week, which implies that individual investors overreact to stale information, leading to temporary movements in firms' stock prices.

### 2.3. Behavioral Finance

The standard finance model, in which unemotional investors always force capital market prices to equal the rational present value of expected future cash flows, has considerable difficulties for fitting these patterns. Researchers in behavioral finance have therefore been attempting to augment the standard model with an alternative model built on two basic assumptions.

The first assumption, laid out in Delong, Shleifer, Summers, and Waldmann (1990), is that investors are subject to sentiment. Investor sentiment, defined broadly, is a belief about future cash flows and investment risks that is not justified by the facts at hand. The second assumption, emphasized by Shleifer and Vishny (1997), is that betting against sentimental investors is costly and risky. As a result, rational investors, or arbitrageurs as they are often called, are not as aggressive in forcing prices to fundamentals, as the standard model would suggest

For many years behavior finance was considered as heretics. Today, psychology and emotions influencing finance decisions is becoming an increasingly convincing argument. The 2002 Nobel Prize in Economics provides a confirmation to this argument as it went to the psychologist Daniel Kahnemann and well-known economist Veron Smith. Over the last decade, the debate between *'efficient market'* argument and propositions of behavioral finance captured great attention and a big body of literature has emerged. Recent finance literature exhibits clues about market behavior in the cognitive psychology of investors. Basic assumptions and theories are summarized as follows:

- *Mental accounting (Thaler, 1985) and risk seeking in the domain of losses (Kahneman and Tversky, 1979) may lead investors to hold onto losing investments and sell winners (see Shefrin and Statman, 1985; Odean, 1998).*
- *The representativeness heuristic (Tversky and Kahneman, 1974) may lead investors to buy securities with strong recent returns (see DeBondt and Thaler, 1987; DeLong et al., 1990b; DeBondt, 1993; and Barberis, Shleifer, and Vishny, 1998).*
- *Overconfidence may cause investors to trade too aggressively and, in combination with self-attribution bias, could contribute to momentum in stock returns (Kyle and Wang, 1997; Odean, 1998).*
- *Limited attention may constrain the set of stocks investors consider buying thus concentrating purchases in attention grabbing stocks. Moreover anticipated regret may dissuade investors from purchasing stocks that have risen since they were previously sold or purchased (Odean, Strahilevitz, and Barber, 2004).*

Fama in 1991 defines efficient market where market participants act rationally given what they know about a company's position and prospects, which became the cornerstone of the financial markets. The market, being the accumulated activities of its many participants rapidly assimilates any information about a company as soon as it becomes available. Therefore, all instruments in the market generally reflect all available information. Pricing models such as *Capital Asset Pricing Model*, *Arbitrage Pricing Model,* and *Option Pricing Model* has all been evolved basing on rationality assumption.

Assertions such as '*overreaction*' or '*under-reaction*' of investors are acknowledged to be crucially important for the tension between efficient market theory and behavioral finance. One argument is that aggregate markets are not entirely efficient after all. The fact that investors make various systematic mistakes all the time is proposed as one reason to such claim. Moreover, historical evidences pose confirmative power to the non-efficient market arguments. Incidents such as '*Tulipmania'* in Holland during 1630s, '*Mississippi Scheme' in* 1720 or '*South Sea Bubble'* in the very same year in England are some of the various examples where efficient market theory turned out to be insufficient to explain the dynamics behind the stock market booms and crashes via traditional theories. Also rational stock price behavior departs from the fundamentals by widely documented patterns on stock returns over weekends, holidays and different calendar periods can also be difficult to reason through upcoming news (Thaler 1987).

Regarding to the history of stock markets, one cannot be fully a supporter of the traditional financial theory. As it is experienced in the history that unrational market moves have been continuously occurring. The history of the stock market is full of events that are striking enough to earn their own names: the Great Crash of 1929, the 'Tronics Boom of the early 1960s, the Go-Go Years of the late 1960s, the Nifty Fifty bubble of the early 1970s, the Black Monday crash of October 1987, and the Internet or Dot.com bubble. Despite the rationality arguments in the efficient market literature, financial markets are already acknowledged to have certain market abnormalities, which cannot be examined via rational choice presumptions. People's behaviors' routinely violate these assumptions. Empirical studies have been demonstrating that stock returns are showing two basic abnormalities.

### 2.4. The Overreaction and Underreaction Abnormality

In the finance literature, focus occurs on two abnormalities regarding the returns to stock exchange returns. One of them is *over-reaction abnormality*, meaning the over-reaction of an investor due to continuous negative trends. One reason is the lacking information for accurate reaction. Often this situation is observed, as investors don't show interest for 1 to 12 months period due to lacking information on stock prices, or continuous negative news. In short, over-reaction abnormality leads a price increase on the stocks that are corresponding to good news and in the following years a decrease in the average return.

A simple definition to '*overreaction'* is that when stock returns following good news is higher than the average return in the period following bad news and observed that these stocks have been known as it has a good reputation and high past returns. Hence, they are generally overvalued. Moreover, it is the natural outcome of overreaction where securities are overpriced during a long-term good news trend and they have low average returns afterwards. Stocks that have persistent bad news record seen as undervalued and it has a high potential to earn superior returns. Meanwhile *'under reaction'* is defined as investors not seeing the real value of current good news. Two basic abnormalities about price changes have been proven by empirical studies.

In efficient markets theory, overreaction and under reaction patterns can be seen as an arbitrage opportunity. Betting against unbalanced price movements due to overreaction and under reaction of markets exposes a contradiction to market efficiency theory. A rational investor may catch the opportunity to build a strategy against these abnormalities without bearing extra risk. By buying losers stocks and selling value stocks, investors can gain superior returns.

Two studies by DeBondt and Thaler (1985, 1987) motivates researchers about overreaction in the stock exchange market. DeBondt and Thaler show that stocks experiencing a poor performance over a 3-5 year period subsequently tend to outperform stocks that are had previously performed relatively well. On average, stocks which are *'losers'* in terms of their returns subsequently become *'winners'* and vice versa. Clara and Thomas propose two explanations about market overreactions:

- *That the 'overreaction effect' is just another manifestation of the 'size effect'. The size effect is the tendency of small firms to generate on average, superior returns to large firms. The argument would follow that the losers were small firms and that these small firms would subsequently outperform the larger firms. DeBondt and Thaler did not believe this a sufficient explanation, but Zarowin (1990) found that allowing for firm size did reduce the subsequent return on the losers*
- *That the reversals of the fortune reflect changes in equilibrium required returns. The losers are argued to be likely to have considerably higher CAPM betas, reflecting investors' perceptions that they are more risky. Of course, betas can change over time, and a substantial fall in the firms' share prices (for the losers) would lead to a rise in their leverage ratios, leading in all likelihood to an increase in their perceived riskiness. Therefore, the required rate of return on the losers will be larger, and their ex post performance better*

Fama and French (1992) incorporate overreaction evidence into their three factor model. The basic idea of CAPM is that the expected return of a stock is solely dependent on the beta of a stock. Fama and French add two additional risk factors, size and book to market equity ratio to do a much better examination of expected return of stocks.

Paper of Barberis, Shleifer and Vishny (1997) structure the investor expectation formation of future earnings in their model. The model is based on heuristics of representativeness and conservatism when people use in assessing the probabilities of outcomes. They show that investor belief formation process may be categorized into two distinct empirical regulatory; namely underreaction to news and overreaction to consistent good or bad news.

Paper of Odeon (1998) presents a model to reason how investors form beliefs within a context of empirical findings. This model is aligned with the results of Tversky and Kahneman (1974) on the important heuristic named as representativeness and another known as conservatism defined by Edwards (1968). Underreaction behavior of the market is consistent with conservatism and overreaction is consistent with representativeness. These papers are not in a position to explain why arbitrage fails to eliminate mispricing. They assume the findings of DeLong, Shleifer and Vishny (1998) by which they show that deviations from the rationality can persist due to investor sentiment is not predictable and arbitrageurs that are betting against mispricing is risky in the short run where extreme movements from the fundamental value can be seen on the prices.

One other example to the counter-arguments that are against the efficient market theory addresses *'up crashes'* due to macroeconomic events in the world. Popular view is that markets rise slowly and crash suddenly is overblown. In January 3rd 2001, Nasdaq went up 14% in one day following a rate cut. October 6th 1931, Dow went up 14.87% following President Hoover's plan for economic recovery. The biggest day crash in the finance history occurred on October 19th 1987 when Dow fell 22.6%, which is much larger than the largest up-crash, but as well twice as big as the next largest down-crash.

Cutler, Poterba, and Summers (1988) find a linkage between aggregate stock returns and various type of news. Macroeconomic news is considered and majority is about political and world events that cover large population through years of 1871 to 1925. They attempt to identify unexpected component, *"the power of macro-economic events",* on monthly returns of S&P index by including dividend payments, industrial production, money supply, inflation rate, market volatility, long and short term interest rate. Cutler et al (1988) suggest that for further understanding of asset price movements requires two sorts of research. The first should attempt to model price movements as functions of evolving consensus opinions about the implications of given pieces of information. The second should formulate and test theories of "propagation mechanisms" that can explain why shocks with small effects on discount rates or cash flows may have large effects on prices.

### 2.5. Stock Prices and Ex-Post Returns

Stock price returns reflect something more other than the news about fundamentals, which is consistent with evidence on correlates of ex-post returns. According to the efficient market hypothesis market abnormalities as mentioned above should not exist. Opposing this standpoint many market abnormalities have been experienced over the years, such as the 'January Effect' (Rozeff and Kinney, 1976; Thaler, 1987), the 'Monday Effect' (French, 1980;Wang, Li, and Erickson, 1997), the 'Post-earnings announcement drift' (Jones and Litzenberger, 1970;Bernard and Thomas, 1990), and the 'Size and book-to-market effects' (Banz, 1981). Thaler (1987) has argued that widely documented patterns on stock returns over weekends, holidays and different calendar periods can also be difficult to reason through assumptions of market efficiency.

In addition, *'Prospect Theory'* of Kahneman and Tversky (1979) suggest that individuals behave as if they regard extremely improbable events as impossible and extremely probable events as certain. *'Value Function'* lies in the core of the theory. For wealth levels under a given reference point investors are considered to be risk seekers. In other words, they are prepared to make riskier bets in order to stay above their preferred target level of wealth. Meanwhile for wealth levels above this reference point, the value function is downward sloping which is in line with the conventional theories. Thus, investors are risk averse.

Following the respective debates, behavioral finance has evoked as an alternative way of price formation to include and reason abnormalities that are persistently observed in the markets. Rational behavior assumption of traditional market theories has been overdrawn by irrelevant behavior of human beings. While individuals are considered to act "rational" in standard finance, they are accounted

as “normal” in behavioral finance. In short behavioral finance is a study of how psychology affects financial decisions, corporations and the financial markets.
On the paper of Vissing-Jorgensen in 2003, they have summarized the contributions of behavioral finance as follows:

- *Documents price patterns that seem inconsistent with traditional finance models of efficient markets and rational investors*
- *Documents behaviors by investors that seem inconsistent with the advice of traditional finance theory*
- *Provides new theories for these patterns and behaviors, often based on behaviors documented in the psychology literature or observed in experiments*
- *Argues that if prices deviate from fundamentals due to the behavior of irrational investors, arbitrage by rational investors may not be able to force prices back to fundamentals*

Behavioral finance related discussions have attracted notable attention as the substantial movements in the firms’ stock prices do not seem to correspond to the changes in quantitative measures. Therefore qualitative variables started to be taken into account as they may help to explain the stock returns.

There is a vast amount of literature searching the impact of emotions on individual behavior in financial markets. Hirshleifer and Shumway (2003) show how stock returns are affected by the weather across the world. Edmans, Garcia, and Norli (2007) associate the outcomes of sporting events, such as a country losing a game in the World Cup, to the following drops in the stock markets which is also consistent with the prospect theory of Kahneman and Tversky (1979).

## 2.6. The Impact of Media on the Financial Markets

### 2.6.1. New Media and Finance World

The idea proposed by the famous economist Shiller is that the news media effect on stock market is firstly recognized in Dutch Tulip Mania crisis in 1630s. Shiller (2000) suggests that the news media is a determinant of financial decision making in on financial markets because of many valid reasons. To begin with, they are attracted to financial markets because at the very least the markets provide constant news in the form of daily price changes. Financial news may have a great human-interest potential to the extent that it deals with the making or breaking of fortunes. Secondly, it is observed that the news repeatedly try to present discussions about issues on the public mind. In this context, media often disseminate and reinforce ideas that are not supported by real evidence. Schiller suggests that at that point, the influence of noise traders becomes apparent. Noise makes financial markets possible, but it also makes them imperfect. According to Black (1986), another famous economist whose findings were similar to those of Shiller, noise trader’s influence has been comprehensive.

World-renowned economists Shleifer and Summers (1990) support and enhance the theory as they propose that the idea of group movement and introduce investor sentiment in financial markets. According to their propositions, significant market events generally occur only if there is similar thinking among large groups of people (investor sentiment) and the new media are essential vehicles for the spreading ideas.

Also, even if the conventional media—print media, television, and radio—have a profound capability for spreading ideas, their ability to generate active behaviors is still limited. Interpersonal and interactive communications, particularly face-to-face or word-of-mouth communications, still have the most powerful impact on our behavior

Journalistic texts are still dominant in investors’ world. Uskali (2009) also argues that even if information and communication technology advanced dramatically from the 1920s to 2000, the flaws of business journalism in writing about stock markets have remained almost the same.

Significant market events generally occur only if there is similar thinking among large groups of people (investor sentiment), the new media are essential vehicles for the spread of ideas. News media do play an important role both in setting the stage for market moves and in instigating the moves themselves. The financial market efficiency literature has devoted great attention to the stock market reaction to the printed media. In an efficient market, these publications should consider to be as conveying second-hand information, with no consequences on prices and volumes.

One of the sources that evoke investor sentiment to force people to have group thinking is the news they follow via different media mediums. These mediums intend to attract the audience attention occasionally by trying to present debate about issues on the public and often disseminate and reinforce ideas that are not supported by real evidence. Many people seem to think that it is the reporting of specific news events and the content of news that affects financial markets.

Starting with Lloyd- Daves & Canes (1978), who examines the performance of second-hand information on published in the Wall Street Journal newspaper, find that WSJ articles affect stock prices on publication. Tetlock (2007) and Tetlock, Saar-Tsechansky, and Macskassy (2008), show in several papers that news stories in national newspapers are associated with substantial price responses. In these papers identification usually focuses on what kind of information a story conveys—for example, a firm's cash flows, risk, or sentiment.

The news media specifically is dependent on the financial markets likewise. Financial markets in this sense act like content providers as the news regards daily price changes continuously. Financial news may have a great human-interest potential to the extent that it deals with the making or breaking of fortunes.

Media cultivation of debate is one important issue to touch upon regarding the impact of media on financial markets. In attempt to attract audiences the news try to present debates on various political and economic issues, mostly reflecting the common interest of the public. These debates other than the actual news reports are often considered to be disseminating and reinforcing ideas that are not necessarily always supported by real evidence. This is one particular characteristics of media in which one can observe the manipulative power of media most. Reports on the market outlook as they provide insight to the news content play an important role in this manipulative power. Many news stories in fact seem to have been written under a deadline to produce something to go along with the numbers from the market. It generally states groups of stocks have risen more than others in recent months.

The question to be raised in an analysis on the impact of media over the financial markets is whether big stock markets in deed changes following the news days where important events occur. It is a general perception that it is the content used for the reporting of specific news events and more important the way these contents are framed in the news affect the changes in the financial markets.

Many researchers discuss the idea that investor sentiment plays a significant role in shaping stock prices. With respect to investor sentiment finance news is crucial to be taken in to account as it provides simultaneous feedback. Shiller (2000) in his popular book reasons various claims to demonstrate the importance of media on finance. He claims that the news media is naturally attracted to financial markets because of its continuous and simultaneous feedback providing nature. Moreover he points out that the financial news in specific have a great human-interest potential to the extent that it deals with the making or breaking of fortunes. One other point discussed in the paper is that the financial news often has a template where they point out that certain groups of stocks have risen more than others in recent periods.

Schiller continues his discussion as he points out that news media like any other business needs to perpetually gain attention of their customers in order to survive and be successful. He states *"For capturing ongoing interest of their customers, news media should create interesting news stories that will be effective to wide majority of public."* Hence, financial market news contains a valuable source of information.

A study of Ojala and Uskali (2009) has traced the possible weak signals in the articles of *The New York Times* before the stock crashes of 1929, 1987 and 2000. By investigating some key words that are believed to associate with financial crisis, they find that there were only a few weak signals in *The New York Times* before the 1929 stock crash. The most valuable signals were usually hidden inside the reports and were not published on the front pages.

Followers of the behavioral finance concept have conducted research on the media effect, which is considered to be one of the reasons for stock market price changes that are not correlated with rational market dynamics.

Tetlock (2007) investigates the relation to see if news media induces, amplifies or simply reflects investors' interpretation of stock market performance. He shows that the number of negative words in the *"Abreast of the Market"* column of the Wall Street Journal predicts stock returns at the daily frequency from 1984 to 1999. Author first describes the relation between the content of financial news media and stock market activity by focusing on WSJ's *"Heard On the Street"*. He bases his research mainly on persuasive literature. In the research as well it is argued that high media is associated with low investor sentiment. Thus this is found to result in downward pressure on prices.

In another study Garcia (2011) has studied the effect of sentiment on asset prices during the 20th century (1905-2005). For the proxy of the sentiment the fraction of positive and negative words in two columns of financial news from the New York Times is used in the research. Main finding of the paper is that given news content helps to predict stock returns at the daily frequency, particularly during in times of hardship. The economic recessions correspond with the times that heightened the sensitivity to news. One standard deviation change in our media factor moves the DJIA by 12 basis points in recessions; the marginal effect during expansions is only 3.5 basis points. Garcia also focused on the positive content effect of the news and found that positive dominance in news also has a significant impact on the stock market. In this paper, he investigated the relation between hard data contained in the news toward market up and downs and resulted that qualitative information does not interact with hard facts. Paper also argues that the effect of news reverses over the following four trading days.

In another paper of Tetlock (2007), which is an application of his previous arguments on a more micro level, has researched whether the different ways of utilizing language in news media has an impact on individual firms' accounting earnings and stock return predictions or not. Tetlock (2007) shows that the number of negative words in the "Abreast of the Market" column of the Wall Street Journal predicts stock returns at the daily frequency from 1984 to 1999. Main findings of this paper are:

- *The fraction of negative words in firm-specific news stories forecasts low firm earnings*
- *Firms' stock prices briefly underreact to the information embedded in negative words*
- *Earnings and return predictability from negative words is largest for the stories that focus on fundamentals*

The conclusion drawn out of the study suggests that linguistic media content captures hard-to-quantify aspects of firms' fundamentals, which are incorporated in stock prices.

Demers and Vega (2010) examine whether the *"soft*" information in management's quarterly earnings press release has informative over company-issued "hard" information. Media expressed negativity is used to evaluate upcoming index and firm performance in their research. One particular point distinguishes their work from Tetlock's (2007) is the way media-expressed negativity is treated. In their seminal work, Demers and Vega (2010) mainly focuses on management-expressed sentiment and the study utilizes various price metrics to examine the stock price performance.

They are the first that use several textual-analysis programs to extract various dimensions of managerial net optimism and certainty from more than 20,000 corporate earnings announcements over the period 1998 to 2006. Their finding is that unanticipated net optimism in manager language affects announcement period abnormal returns and predicts post earnings announcement drift supported by additional market responsive to soft information when; (i) information is more verifiable due to the simultaneous release of more quantitative data; (ii) there are multiple informed experts (analysts and the media) following the firm; and (iii) the managers 'earnings forecasting reputation is better.

Liu, Smith and Syed (1990) studied the impact of "Heard on the Street' column on company stock prices and found out that it has a statistically significant impact on prices on the publication day and its impact diminish within 2 days preceding the publication. They conduct their analysis based on an event study, which investigates the relation between predefined company stocks and return changes. The cumulative abnormal return of all company stocks that are identified over the three day period from two days before publication to publication is 3.09 percent. They have evaluated the column content based on its *"buy/sell"* recommendations and additional to its impact on prices they realized that the reaction of stock prices is symmetric with respect to recommendation type. Before the previously mentioned study, Lloyd Davies and Canes (1978) also examine the performance of information on *"Heard on the Street"* for the period of 1970-1971. They document a statistically significant impact of the articles on stock prices in relation with the publication date.

A similar study by Barber and Loeffler (1993) analyzes security returns and trading volume around the publication date of a monthly column "Dartboard" published in Wall Street Journal. Study results also contribute to abnormal returns on exchanges and high trading volumes in the financial markets. However due to the content of this column it contains the recommendations of specified and well-known investment advisory agencies and also concludes that certain investors buy stocks based on analyst's recommendations. This pressure causes a temporary price return and volume changes on the recommended stocks, and this pressure diminishes within few days.

An event study by Palmon, Sudit and Yezegel (2009) is about columnists' recommendations of several business leading magazines. They find that recommendations on mergers and acquisitions generate significantly greater market reactions. In overall the results also indicate the media news impact on prices on the short term.

Paper of Engelberg and Parsons find that the presence or absence of local media coverage is strongly related to the probability and magnitude of local trading. The strongest causal evidence comes from examining exogenous shocks to the transmission of media coverage to local investors is that on days when extreme weather events (hailstorms and/or blizzards) are likely to disrupt the normal delivery of daily newspapers, the relation between media coverage and trading is broken. This is an important and evidence as weather changes cannot be correlated with either underlying content or unobservable determinants of investor demand.

A study of Palmon, Ephraim and Sudit (2009) also emphasizes the columnists' effect on stock prices. In this study, three leading business magazines; *Business Week, Forbes* and *Fortune* stock recommendation columns are investigated. They conduct short term and long-term return behavior to analyze the recommendation effect on the market. Their results suggest similarities with the previous researches. However, they document that recommendations that contain references to management or provide merger & acquisition related rumor trigger significantly greater market reaction. Finally they find that long term effects of recommendations would not have been consistently create abnormal market behavior with controlling for market risk, size, book-to-market and momentum effects. Galphin, Bhattacharya, Ray and Yu (2004) examine the role of the media in the internet IPO bubble. They document that the media overestimate the good news for internet IPOs in the bubble period and underestimate the bad news for internet IPOs in the post-bubble period. Market, on the other hand, discounts this in equilibrium in the long run.

A substantial amount of research has emerged that investigate the impact of second hand information on behavior in relation to the stock prices. One study from Smith and Sayed (1992) examines the impact of scandals that are published in the "Heard on the Street" column of the Wall Street Journal. They claim that authors and editors became more cautious against the information leaks over the time. Moreover one particular scandal they investigated appeared to not create an impact.

Another paper from Barber and Loefller (1993) examines the effect of second hand information on the behavior of security process and volume using analyst's recommendation published in the monthly "Dartboard" column of the "Wall Street Journal". They find that for the two weeks following the recommendation average abnormal returns nearly twice the level of average abnormal returns and average volume double normal levels on the two days following recommendation publication. This trend is reversed within 25 trading days. It is concluded that abnormal return and volume is the result of naïve buying pressure and information content of the recommendations of the "Dartboard" column.

A noise and soft information related study by Demers and Vega (2010) utilizes the textual data likewise to capture the impact of news on the market. Their findings signal that unanticipated net optimism in manager language affects announcement period abnormal returns and predicts post earnings announcement drift. Net optimism and certainty dimensions have been analyzed on average , soft information affects asset prices both within the earnings announcements event window and during 60 trading day post announcement period even after controlling contemporaneously released hard copy.

One other line of literature takes social media into consideration. The idea is that the social media channels are also becoming important on stock market returns. For example, Twitter can predict the ups and downs of the stock market, a new study finds. Measuring how calm the Twitterverse is on a given day can foretell the direction of changes to the Dow Jones Industrial Average three days later with an accuracy of 86.7 percent. This research uses a standard psychology tool called the Profile of Mood States, and asks people to rate how closely their feelings match 72 different adjectives, including *"friendly," "peeved," "active," "on edge"* and *"panicky,"* and uses the responses to measure mood along six dimensions: calmness, alertness, sureness, vitality, kindness and happiness. Then takes 9.8 million tweets from 2.7 million tweeters between February and December 2008, compared the national mood to the Dow Jones Industrial Average. The result is that one emotion, calmness, lined up surprisingly well with the rises and falls of the stock market — but three or four days in advance.

One other example to the studies on the relation between social media impact on financial markets is done by Nofsinger (2005) Paper argues that the general level of optimism/media in society affects the emotions of most financial decision-makers at the same time. This has basically three outcomes. First, social media affects the decisions of consumers, investors, and corporate managers alike. Second, the stock market itself can be seen as a social mood measure. Thirdly, like the consequences that are mentioned above, social mood may be used to forecast future financial and economic activity. Empirical analysis shows that high social mood causes an increase optimistic decision, whereas low social mood drives the increase of pessimistic decisions.

# CHAPTER 3

# DATA AND METHODOLOGY

## 3.1. Media Factor Proxy

### 3.1.1. Media Factor Source

This study uses several sources of data. The main source is the regular coverage of business news articles in the Wall Street Journal (WSJ). The Wall Street Journal is chosen as a media factor source because it is widely considered to be one of the most prestigious business newspapers in the world and it has been referred to as an important source of business and finance-related news. The WSJ is owned by Dow Jones and Company and has repeatedly ranked as the number one daily newspaper in the US with a daily circulation of more than 2 million. The newspaper's European arm, WSJ Europe also reaches a large number of readers with almost 200 thousand daily circulations that is estimated to cover an average of 79 percent of the European population.

In this study, the articles obtained from the *'Heard on the Street'* column published daily in both the American and the European versions of the WSJ are analyzed. The statistics provided by the WSJ suggest that the paper has a wide reach among investors and is one of the essential readings for finance professionals, corporate decision-makers and retail investors. The 'Heard on the Street' column provides international business analysis in both printed and online versions of the newspaper and has its own page on the WSJ website. It has been a staple business column for Wall Street Journal readers since the 1960s and is widely followed by the business and finance community.

### 3.1.2. Media Factor Derivation

In this study, the media content measures are derived from the WSJ archives that go back to year 2000 for the US edition and 2002 for the European edition. The archives for the last 2 years are available to any subscriber of the Wall Street Journal Europe and US. The previous 10 years' archives are available from Factiva.

The final data set for the study contains 5,353 extractions from the WSJ US and 1,629 extractions from WSJ Europe archives. European dataset covers the period from January 2, 2002 to June 3, 2011, and the US dataset covers the period from January 1, 2000 to August 10, 2011.

## 3.2. Stock Market Proxy

### 3.2.1. Index Choice and Collection

Since the Heard on the Street column focuses on the market as a whole, it is crucial to select an index that also represents a significant portion of all the movements in the markets. In order to analyze the relationship between aggregate stock market returns and media content, this study uses two proprietary sets of market indexes (stock return information).

The total return index for the Dow Jones Industrial Average and the Dow Jones Global Indexes are collected from the Dow Jones Indexes website. Total return and volume values are available free of charge on the company from its website.

The index averages are calculated as the simple arithmetic mean prices of the stocks that comprise the indicated index at that time. The procedure is that the sum of the prices of all stocks in the index is divided by a divisor. The divisor is adjusted in case of stock splits, spinoffs or similar structural changes in order to ensure that such events do not in themselves alter the numerical value of the index.

The result gives the price of the index at that time. Dow Jones Indexes website includes the average daily price of the indexes, along with other detailed information.

**Index 1: American Market Indicator: Dow Jones Industrial Average**

The Dow Jones Industrial Average, also called the Industrial Average, is one of the several indices created by Dow Jones and Company. This average is computed from the prices of 30 stocks specifically chosen to represent the trends in the industrial sector. In the last years, the index is compiled so that two thirds of the component companies are manufacturers of industrial and consumer goods, and the remaining companies represent industries of financial services, entertainment and information technology. It is the most-quoted market indicator in newspapers, on television and on the Internet. Index 1 is used as a market proxy when analyzing the relationship between media and US market returns.

**Index 2: European Market Indicator: Dow Jones Global**

The 150-stock index, the Global Dow, is compiled to cover the most innovative, vibrant and influential corporations from around the world. It is a family of international equity indexes, including world, region, and country indexes and economic sector, market sector, industry-group, and subgroup indexes. Its components are selected among the global stocks that have an excellent reputation that demonstrate sustained growth, are of interest to a large number of investors and accurately represent the market sectors covered by the average. This index follows the stocks that are not just the leading global companies based on their size and reputation but also the ones that have a promise for the future.

## 3.3. Methodology

### 3.3.1. The Research Question

This study aims to contribute to the existing literature by analyzing the impact of media content on asset returns while focusing on a data set that has not been previously studied. The previous studies of Engelberg and Parsons (2011) and Dougal, Engelberg, Garcia, and Parsons (2011) helped to build the link between media and asset pricing variables and to demonstrate that investors seem to communicate with their external environment while making their investment decisions. This study furthers the analysis by utilizing qualitative research methods as suggested by Tetlock in his 2007 study.

In order to examine the correlation between the news provided to the market through various media channels and the reaction of the market prices to these news, the proxy for the media usage is obtained from the journalistic texts of Wall Street Journal's *'Heard on the Street'* column. Following the literature discussed in Chapter 2, this study aims to analyze the interaction between news and stock market returns using the daily content from the Wall Street Journal US and Europe. For this purpose the research model is based on answering the following research question:

Does the financial news presented in the Heard on the Street column of the WSJ have any impact on the stock market returns? In answering this question, the study will shed light on the behavior of the sample stock market indexes following the publication of the WSJ daily column. As it is mentioned in the literature reviewed, temporary buying pressure created by naïve investors may contribute to the abnormal returns and volume observed in the markets from time to time. This study also examines the influence of the lagged market volume in order to understand the timing of market information effects. In addition, the effect of changes in market volume is analyzed as a proxy for the immediate market response to the available information, regardless of whether the information is true or not. This last analysis is carried out to test the existence of noise trading. Based on the research question, related hypotheses are built to clarify the media factor effect on stock returns.

*$H_{0,1}$: Media proxy forecasts market sentiment*

If this hypothesis cannot be rejected, then high or low media proxy values should predict abnormal returns in short time horizons and a reversion to fundamentals at longer time horizons.

*$H_{0,2}$: Media proxy reflects past market sentiment*

If this hypothesis cannot be rejected, then negative media proxy values should follow after low returns and positive media proxy values should follow after high returns.

$H_{0,3}$: For each case, if media proxy either reflects the past sentiment or predicts the future sentiment, unusually high or low levels of media should be associated with increases in trading volume

### 3.3.2. Research Methodology

This section is devoted to the description of the methods employed in the analysis of the relationship between media news (as a proxy for investor sentiment) stock market returns and trading volumes. The methodology that is used in this study can be categorized into two groups:

- The first group contains the analyses related with the calculation of the market sentiment factor based on media content. Explanations for the statistical methodology utilized to the derive media factor are presented below.
- The second group is comprised of the time series analysis that is employed to examine the impact of news on stock market returns.

### 3.3.3. Media Factor Extraction

The parameter "investor sentiment" is at the focal point of this study. Since this parameter is an unquantifiable sentiment, a proxy measure needs to be developed. The methodology of the study aims to "extract" the investor sentiment from the contents of the "Heard on the Street" column of the WSJ. The so-called "dictionary approach" is used in order to develop a quantifiable measure of sentiment based on the textual content of the column. For each column *i* published on day *t*, content analysis is conducted using the dictionaries provided by the Harvard University's Psychology Department. The next three subsections describe the details of the content analysis, the specific tool used in this study and the calculation of the investor sentiment proxy, respectively.

**Content Analysis**

Content analysis has been defined as a systematic, replicable technique for compressing many words of text into fewer content categories based on explicit rules of coding (Krippendorff, 1980; and Weber, 1990).

Text is one of the basic and most important sources of material in social sciences. Although this is a notable fact, a vast amount of statistical work proceeds without utilizing it for analysis. Even in the cases where text is acknowledged as a source, it is not common to find a systematic way of analyzing it. 'Computer Content Analysis' (CCA) is a technique used by various branches of social and political sciences. Typically, the use of CCA involves constructing a category system and creating a 'dictionary' that associates a set of words with a theoretically relevant concept. The selected tool for the CCA summarizes the content of the documents that are plugged in and suggests a vector of categories based on the occurrence frequencies.

CCA is a qualitative method that summarizes the presence of certain words, concepts, themes, phrases, characters, or sentences within texts by counting various aspects of the content. The results are often presented in numbers or percentages. General procedure involves coding the text and later categorizing the material in order to build manageable levels for feeding the input content. At the end, it provides enough numeric information that can be analyzed further with various quantitative methods. Bernard Berelson, who is the American leading scientist on the subject, defines content analysis as *"a research technique for the objective, systematic, and quantitative description of manifest content of communications"*.

The CCA can employ methods that either look for keywords for a conceptual analysis or construct a detailed graph structure for a relational analysis. In the conceptual analysis, a concept is chosen and the number of its occurrences within the text is recorded. This version of the CCA makes no assumptions about how the text is actually generated since it is based on a theory of keyword content. It is an attempt to describe measures of content that do not assume a developed theory, and it concentrates on word usage as a guide to the content. The importance of word usage for understanding the meaning was pursued in the philosophy of language studies (Wittgenstein, 1958; Quine, 1960, 1961) and linguistics (Harris, 1954, 1963b; Cruse, 1986). Since "terms" in a text may be implicit as well as explicit, it is important to clearly define implicit terms before the beginning of the counting process. In order to limit subjectivity in the definitions of concepts, specialized dictionaries are used. In dictionary-based approaches, the data-generating model is one where there is a set of categories that are assumed to be expressed in the text with a known distribution of vocabulary words given the category. For instance, if the researcher is looking for the category of "media" in the text, then a set of words that are representative of media (based on the dictionary used) are counted for their frequency within the text and then these frequencies are used to determine the level of media. A dictionary-based content analysis is appropriate when the categories that are constructed by the researcher coincide well with those constructed by the author.

Relational analysis, on the other hand, builds on conceptual analysis by examining the relationships among concepts in a selected text. It is based on a highly developed theory of how causal relationships are expressed in text. Once context vectors are constructed, using multidimensional scaling or cluster analysis can visualize them. If a distance measure between vectors is chosen, then the distances themselves can be analyzed directly. These distances are then used as measures of contextual similarity between any two terms in a document.

In this study, the “Heard on the Street” columns are examined by using content analysis based on the argument in the literature that this column’s contents have an influence on stock returns. . For example, analysis of the column by Lloyd-Davies and Canes (1978) and Liu, Smith, and Syed (1990) indicate that the contents of the column seem to have an impact on stock prices. This study analyzes the column’s contents by using the General Inquirer software and the specifics of the software are described in the next section.

General Inquirer Software

The General Inquirer is web-based collaborative software developed by the joint efforts of researchers at the Harvard University and the Massachusetts Institute of Technology. It “…has been supported by grants from the USA National Science Foundation and Research Grant Councils of Great Britain and Australia…” The General Inquirer merges the Harvard and Lasswell dictionaries and it contains 182 tag categories in total. The General Inquirer is a mapping tool and it maps each text file with counts on dictionary-supplied categories. The program takes text transcripts and processes them to give word counts and frequency percentages of various word lists based on two separate dictionaries. The first dictionary is the Harvard dictionary that was developed with the software itself and consists of general word categories such as environment, positivity, negativity and emotions. The second dictionary, the Lasswell dictionary, was created by social scientists Namenwirth and Weber as part of their work detailed in the book “Dynamics of Culture.”

Its online server provides the access to the General Inquirer software. In order to access the server, Aleksander Wawer and Roger Hurwitz (server developers) provide a user code and password to eligible users. It is important to note that this server does not have publicly available access.

For analyzing the content of the Heard on the Street columns, two categories of media content proxies have been formed from the 182 word categories: The first level categorization is done by using the largest possible classification tags: each word in the text is categorized as either positive or negative. Using the Harvard dictionary's tag architecture of the English language does the second level categorization. These are dictionary schemes based on commonsense and general categories of meaning. The details of the categorization scheme are described below:

First categorization scheme tags:

- Positive tag group: Contains 1,915 words for positive outlook
- Negative tag group: Contains 2,291 words for negative outlook

Second categorization scheme tags:

- Positive tag group contains 1045 positive words
- Negative tag group contains 1160 negative words
- Strong tag group contains 1902 words implying strength
- Weak tag group contains 755 words implying weakness
- Active tag group contains 2045 words implying an active orientation
- Passive tag group contains 911 words indicating a passive orientation
- Pain tag group contains 254 words indicating suffering, lack of confidence, or commitment
- Pleasure tag group contains 168 words indicating the enjoyment of a feeling, including words indicating confidence, interest and commitment
- Feel tag group contains 49 words describing particular feelings, including gratitude, apathy, and optimism, not those of pain or pleasure
- Arousal tag group contains 166 words indicating excitation, aside from pleasures or pains, but including arousal of affiliation and hostility

- EMOT tag group contains 311 words related to emotion that is used as a disambiguation category, but also available for general use
- Virtue tag group contains 719 words indicating an assessment of moral approval or good fortune, especially from the perspective of middle-class society
- Vice tag group contains 685 words indicating an assessment of moral disapproval or misfortune
- Ovrst tag group contains 696 words indicating emphasis in realms of speed, frequency, causality, inclusiveness, quantity or quasi-quantity, accuracy, validity, scope, size, clarity, exceptionality, intensity, likelihood, certainty and extremity
- Undrst tag group contains 319 words indicating de-emphasis and caution in these realms.

Ultimately, three different media proxies are formed based on the analysis of the above tags. The first is the Positive proxy, the second is the Negative proxy, and the third is a "Media" proxy that is constructed by applying a principal component analysis (PCA) on the second categorization of the scheme tags in the GI software. These three proxies are included in the vector autoregressive models in the later steps of the analysis. The details of the PCA and VAR are provided in other subsections below.

After the selection of the category tags, the correct preparation of textual data for the program is crucial. The General Inquirer's Java® version processes all the text within each of the files contained in a specific folder. An output record of tag counts is generated for each file. A sample output can be found in the appendix. The steps taken in order to prepare the input for the software are explained below:

1. Edit each file of "Heard on the Street" to have any content removed that should not be part of the analysis, such as headings.
2. Copy the relevant text from the Word document to an Excel file since the server accepts Excel files as an input. In each file, the first column

must include an identification code for each row (and each row represents a date on which the Heard on the Street column is published).

3. Formulate each row in the Excel file to represent one body of text that is coded with its publication day (as described in step 2). This step generates 5,353 spreadsheets for the US case and 1,629 spreadsheets for the European case (a total of 6,982 data points).
4. Before feeding the Excel files to the server, check the number of rows since the server works best with a maximum of 250 rows.
5. Consolidate the results in a final Excel spreadsheet and rename each file to include information about the date and the source.

The flow diagram is also inserted in Appendix B at the end of the chapters. Flow diagram could be an aid for the researcher to analyze the input texts in an appropriate order. Content analysis methodology should be carried out with minimizing errors.

**General Inquirer Output**

The output is a matrix of the "tag counts" for each category with separate rows of counts for each file (each daily text) processed. A sample can be found in the appendix. Once the output matrix is generated by the General Inquirer, the next step is the calculation of the “market sentiment” proxy based on the counts in the table. For this purpose, the percentage of each tag category is normalized so that the sentiment measures have a zero mean and a unit variance. This allows interpreting the regression coefficients in the next step in terms of one standard deviation shocks to the sentiment measures, thus making it easier to gauge the magnitude of the results.

Since the GI output provides the “counts” for each tag, the first step in the normalization process is calculating the “percentage” values for these count figures. The percentages are calculated by dividing the count for each tag in each day by the total word count for that day. For instance, on Jan. 2, 2002, the count for the positive tag in the WSJ Europe text is 43. On the other hand, the total word count for that day is 959. Out of these 959 words, 84 are “leftover” meaning there were no matches between these 84 words and the words that the GI software counts for each of its tags based on dictionary definitions. This means, out of the 875 words that GI counts, 43 were tagged as positive; in other words, 4.91% of the words in the Heard on the Street column on Jan. 2, 2002 were tagged as positive by the GI.

After calculating the daily percentage figures for the tags, the next step is the calculation of the z-scores based on sample averages. For instance, the 4.91% calculated for the positive tag on Jan. 2, 2002 has a z-score of 0.0952. This normalized score is calculated by subtracting the positive tag’s mean percentage score of 0.0479 over the entire sample and dividing the difference by the positive tag’s standard deviation of 0.0121 again over the entire sample.

In the following step, principal components analysis (PCA) is used to derive the sentiment proxy from the normalized percentages. It should be noted that several different approaches for the calculation of the market sentiment indexes are suggested by researchers, e.g. Brown and Cli (2004, 2005), Qiu and Welch (2006), Lemmon and Portniaguina (2006), Tetlock (2007). In this study, the GI software that is used to analyze the text content generates rather general categories of tags (positive versus negative). With the help of the PCA, it becomes possible to construct a sentiment proxy that is made up of more specific sentiment tags that are listed above as the second-level categorization.

In general, PCA is a method that reduces a set of observations from variables that are possibly correlated into a smaller set of, linearly uncorrelated, artificial variables called principal components. In the transformation process, the first principal component will have the largest possible variance so that it represents as much of the variability in the original data as possible. The subsequent components will also have as high a variance as possible while satisfying the condition that they are uncorrelated with the components that precede them. Typically, the components generated serve as an intermediate step and they may be used as inputs to a multiple regression. PCA generates a simple media proxy by using general category tags from the Harvard dictionary. This process reduces the 16 categories from the dictionary into a single media factor that captures the maximum variance in the GI categories. It is designed to detect the complex structure in the WSJ texts and to eliminate the redundant categories in the dictionary.

Time-Series (VAR) Analysis

The vector autoregressive model was initially introduced by Sims (1980). It is used to analyze the linear interdependencies among multiple times series observations. In this study, the VAR is used to analyze the relationship between the media/sentiment proxy generated in the previous step and returns and volume observed on the sample stock market indexes. Previously, Brooks and Tsolacos (1999) have employed a VAR methodology for investigating the interaction between the UK property market and various macro-economic variables. Likewise, Tetlock (2007) and Garcia (2011) have used the VAR method in their studies where the relationship between a similarly constructed media factor and the stock returns are analyzed.

A univariate autoregression model is a single-equation, single-variable linear model in which the current value of a variable is explained by its own lagged values. VAR is a system regression model where there is more than one dependent variable and so it can be considered as a hybrid between univariate time series models and simultaneous equations. VAR is used to analyze the interrelation of time series under consideration as well as the dynamic impacts of random disturbances (or innovations) on the system of variables. In a system of two variables, each variable is modeled as a function of the lagged values of itself and the other variable and individual regressions are estimated together as a system of equations. For estimations, the ordinary least squares method is appropriate only if the disturbances in the model can be shown to be serially uncorrelated with a constant variance. These pre-conditions are tested and results are presented in Chapter 4. The equation below represents the construction of a VAR model with variables x and y being estimated with k number of lags.

$$x_t = b_1 y_{t-1} + ... + b_k y_{t-k} + b_{k+1} x_{t-1} + ... b_{k+1+n} x_{t-n} + e_t^y \quad \text{(Equation 3.1)}$$

In the Equation 3.1, the two endogenous variables y and x are also the explanatory variables in lagged form. Error terms are called impulses or innovations or shocks in the language of VAR.

In this study, there are two classes of variables that are included in the VAR models. The first class includes the endogeneous variables: daily stock index return, daily stock index volume, three different media factors generated from the GI analysis (positive, negative, media). The second class includes the exogeneous variables. The first exogeneous variable is the standard deviation of stock index returns over the past 60 days. This variable is included in the model since previous studies in the literature have demonstrated a significant effect of this variable on the current stock returns. Garcia (2011) and Tetlock (2007) include these exogeneous variable to fixed volatility adjustments, neglect outliers period for crisis period and other known determinants of variations. For the rest of the paper we normalize our sentiment measures so they have zero mean and unit variance. This will allow us to interpret the regression coefficients in terms of one standard deviation shocks to the sentiment measures, thus making it easier to gauge the economic magnitude of our results.

By including the variable, this study will provide comparable results. The second exogeneous variable takes the values of 1 through 5 for each day of the week. Likewise, a third exogeneous variable takes the values 1 through 12 for each month of the year. These two variables are included in the model to account for the anomalies in return generation previously demonstrated in the literature. Finally, a dummy variable that takes the value of 1 for the 2008 inputs is included in order to account for the possible effects of the financial crisis.

# CHAPTER 4

# RESULTS AND ANALYSIS

## 4.1. Principal Component Analysis

The purpose of this step is to construct a “media” media proxy by conducting a PCA on the second-level categorizations of Harvard dictionary tags. Since the sample includes texts from the WSJ US and WSJ Europe, these two sources are analyzed separately.

### 4.1.1. PCA Analysis Europe

Dictionary tags for the WSJ Europe texts were obtained from the General Inquirer in the manner that was described in the previous section. As a second step, varimax rotated principal components were estimated based on these tags. Table 4.1 presents the results of this estimation. As explained before, the goal in the PCA is to choose the component that explains the largest percentage of the variation in the data. Two alternative components are estimated with 36.3% variability explained by the first and 27.6% variability explained by the second component. As a result of its higher explanatory power, the first component is chosen as the media proxy for use in WSJ Europe analyses. In this component, the negative, weak, passive, EMOT and vice tags have loadings that are larger than 0.30 and these are the only tags that will be used in the construction of the media factor. For instance, on Jan. 2, 2002, the media factor is -13.4%. This number is calculated by taking a weighted average of the normalized percentages of the negative, weak, passive, EMOT and vice tags based on their factor loadings from Table 4.1.

### 4.1.2. PCA Analysis US

Table 4.2 presents the results of the component estimation for the WSJ US texts. Once again, two alternative components are estimated with 38.3% variability explained by the first and 39.3% variability explained by the second component. As a result of its higher explanatory power, the second component is chosen as the media proxy for use in WSJ US analyses. In this component, the positive, strong, and virtue tags have loadings that are larger than 0.30 and these are the only tags that will be used in the 77.7%. This number is calculated by taking a weighted average of the normalized percentages of the positive, strong, and virtue tags based on their factor loadings from Table 4.2.

Table 4.1: Varimax Rotated Factor Loadings, Dictionary Tags for WSJ Europe

| Dictionary Tags | 1 | 2 |
|---|---|---|
| Pstv | -0,17 | 0,73* |
| Ngtv | 0,73* | -0,10 |
| Strong | -0,13 | 0,44* |
| Weak | 0,74* | -0,16 |
| Active | -0,25 | 0,25 |
| Passive | 0,54* | 0,14 |
| Pleasur | 0,07 | 0,52 |
| Pain | 0,63* | 0,19 |
| Feel | 0,03 | 0,17 |
| Arousal | 0,28 | 0,24 |
| EMOT | 0,44* | 0,45 |
| Virtue | -0,00 | 0,53* |
| Vice | 0,66* | -0,16 |
| Ovrst | 0,21 | 0,16 |
| Undrst | 0,27 | -0,10 |
| Eigen Value | 5,45 | 4,15 |
| Variance Explained | 36,3% | 27,6% |

Table 4.2: Varimax Rotated Factor Loadings, Dictionary Tags for WSJ US

| Dictionary Tags | 1 | 2 |
|---|---|---|
| Pstv | -0.18 | 0.83* |
| Ngtv | 0.81* | -0.30 |
| Strong | -0.20 | 0.66* |
| Weak | 0.73* | -0.31 |
| Active | 0.25 | 0.48 |
| Passive | 0.59* | -0.10 |
| Pleasur | 0.16 | 0.35 |
| Pain | 0.54* | -0.12 |
| Feel | 0.14 | 0.04 |
| Arousal | 0.25 | 0.10 |
| EMOT | 0.61* | 0.17 |
| Virtue | -0.19 | 0.75* |
| Vice | 0.63* | -0.33 |
| Ovrst | 0.44 | 0.21 |
| Undrst | 0.58 | -0.41 |
| Eigen Value | 5.74 | 5.90 |
| Variance Explained | 38.3% | 39.3% |

## 4.2. VAR Analysis

### 4.2.1. Assumptions of VAR Analysis

Before a VAR analysis can be conducted. The variables for the VAR must be checked for their stationary and homoscedasticity.

A stationary series can be defined as one with a constant mean. Constant variance and a constant auto-covariance are attributed for each given lag. A series is considered as stationary if the distributional property of its values remains the same over time. Determining the stationary is important since the use of standard regression techniques requires homoscedatic and serially uncorrelated disturbances. Otherwise, non-stationary data could lead to spurious regression results. Where the regression estimates have desirable properties but they are actually useless.

Moreover, if the series turn out to be non-stationary, then further steps must be taken to transform the series into a stationary one before the observations can be used in the VAR analysis. In this study, the Augmented Dickey-Fuller (ADF) test (Dickey and Fuller. 1979), and the Phillips-Perron test (Phillips and Perron. 1988) -which builds on the ADF but makes a non-parametric correction for autocorrelation- are used to test for the existence of a unit root in the series. . Results are shown in Tables 4.3 and 4.4.

Table 4.3: Stationarity Tests for WSJ Europe

| | Augmented Dickey-Fuller | | | | Philips-Perron | | | |
|---|---|---|---|---|---|---|---|---|
| | t value | t (%5) | t (%1) | p | t value | t (%5) | t (%1) | p |
| Media Factor | -12.47 | -2.86 | -3.43 | 0.00** | -39.99 | -2.86 | -3.43 | 0.00** |
| Volume | -17.90 | -2.86 | -3.43 | 0.00** | -58.58 | -2.86 | -3.43 | 0.00** |
| Return | -29.50 | -2.86 | -3.43 | 0.00** | -35.92 | -2.86 | -3.43 | 0.00** |

(Note: *p<.05, **p<.01)

The null hypotheses for both the ADF and PP tests are the existence of a unit root, which would indicate non-stationarity in the series. Results in Table 4.3 show that all variables that will be included in the VAR analysis are stationary since the null hypotheses are rejected at the 1% significance level.

Table 4.4: Stationarity Tests for WSJ US

| | Augmented Dickey-Fuller | | | | Philips-Perron | | | |
|---|---|---|---|---|---|---|---|---|
| | t value | t (%5) | t (%1) | p | t value | t (%5) | t (%1) | P |
| Media Factor | -17.60 | -3.41 | -3.96 | 0.00** | -55.01 | -3.41 | -3.967 | 0.00** |
| Volume | -21.61 | -3.41 | -3.96 | 0.00** | -90.67 | -3.42 | -3.96 | 0.00** |
| Return | -15.32 | -3.412 | -3.97 | 0.00** | -60.03 | -3.41 | -3.95 | 0.00** |

(Note: *p<.05, **p<.01)

Once again, for the WSJ US case, the null hypotheses are rejected at the 1% significance level for all variables implying that all series are stationary.

The second assumption that needs to be tested is the homoscedasticity of the error terms where a constant variance is implied. If the errors do not have a constant variance, they are said to be heteroskedastic. The concept of heteroskedasticity is very important when it comes to interpreting the results of a regression. If an ordinary least squares regression is performed in the presence of heteroskedasticity. The calculated standard errors for the coefficients are meaningless and cannot be interpreted. In this study, the White's general test for heteroskedasticity (White. 1980) is used to test for this assumption. The results are presented in Tables 4.5 and 4.6.

Table 4.5 Heteroskedasticity Test for WSJ Europe (White Test)

| | | |
|---|---|---|
| CHI-SQ = | 34.935 | (p=0.566) |

Table 4.6 Heteroskedasticity Test for WSJ US (White Test)

| | | |
|---|---|---|
| CHI-SQ = | 38.035 | (p=0.961) |

The null hypothesis for White's test is homoscedasticity. Since the p-values for both the WSJ Europe and WSJ US samples are both larger than 1%. the null hypothesis is failed to be rejected and therefore both of the series are shown to have a constant variance.

In the VAR analysis, 3 endogeneous variables have been used. These are return of the stock indexes, volume of the stock indexes and media factor. All of these variable have two version for the US and European market. Related analyses are carried out for both of the markets.

### 4.2.2. Choosing the Lag Length for the VAR Analysis

One of the first decisions that need to be made when conducting a VAR analysis is choosing the lag lengths for the endogeneous variables. This is a critical decision since if the lag length is kept too long. this will lead to a loss of degrees of freedom and possible multicollinearity and if the lag length is kept too short. this may lead to specification errors.

There are a number of tests that can be performed as an aid in this decision. In this study, the Akaike Information Criteria (AIC) and the Final Prediction Error Criteria (FPE) are employed. The appropriate lag length must be determined by allowing a different lag length for each equation at each time and choosing the model with the lowest AIC and FPE values. The results of these tests are presented in Tables 4.7 and 4.8 and imply that a lag of 5 days would be the optimum choice.

Table 4.7: Lag Length Test for WSJ Europe

| Lag | FPE | AIC |
|---|---|---|
| 0 | 0.54 | 7.90 |
| 1 | 0.47 | 7.76 |
| 2 | 0.46 | 7.74 |
| 3 | 0.46 | 7.74 |
| 4 | 0.45 | 7.71 |
| 5 | 0.44* | 7.69* |

*Indicates most appropriate lag length

Table 4.8: Lag Length Test for WSJ US

| Lag | FPE | AIC |
|---|---|---|
| 0 | 0.79 | 8.28 |
| 1 | 0.68 | 8.13 |
| 2 | 0.65 | 8.08 |
| 3 | 0.63 | 8.05 |
| 4 | 0.62 | 8.04 |
| 5 | 0.62 | 8.03 |

Table 4.8 (continued)

| | | |
|---|---|---|
| 6 | 0.62 | 8.03 |
| 7 | 0.62 | 8.03 |
| 8 | 0.61* | 8.02* |

*Indicates most appropriate lag length

Based on the AIC and FPE results. Lag length of 5 days is chosen for the WSJ Europe sample and a lag length of 8 days is chosen for the WSJ US sample.

### 4.2.1. Different Representations of the VAR Model

#### 4.2.1.1. Testing for Granger Causality

Causality means that a variable x would be causal to a variable y if x could be interpreted as the cause of y and/or y as the effect of x. Testing causality in the Granger sense involves using *F*-tests to determine whether lagged information on a variable *Y* provides any statistically significant information about a variable *X* in the presence of lagged values of *X*. In this study, a Granger causality test is conducted to identify any causality that may exist between the endogeneous variables of stock return, trading volume and media factor.

In the VAR analysis, 3 endogeneous variables have been used. These are return of the stock indexes, volume of the stock indexes and media factor. All of these variable have two version for the US and European market.

Based on the results from the previous section, a lag of 5 days is used for the WSJ Europe sample and a lag of 8 days is used for theWSJ US sample. The results are presented in Tables 4.9 and 4.10.

Table 4.9: Granger Causality Test Results for WSJ Europe

| | F-value | P-value |
|---|---|---|
| **$H_0$: Media Factor does not Granger-cause Stock Return** | 2.27 | 0.05* |
| **$H_0$: Media Factor does not Granger-cause Trading Volume** | 0.78 | 0.57 |
| **$H_0$: Trading Volume does not Granger-cause Stock Return** | **0.54** | **0.75** |
| **$H_0$: Stock Return does not Granger-cause Trading Volume** | 0.40 | 0.85 |
| **$H_0$: Return Does not Granger-cause Media** | 0,61 | 0,69 |
| **$H_0$: Trading Volume Does not Granger-cause Media** | 1,79 | 0,11 |

By looking at the results in Table 4.9 it is seen that for the WSJ Europe sample, the only significant p-value is for the relationship between the media factor and the stock returns. The null hypothesis that no such relation exists can be rejected at the 5 percent significance level. This result implies that there is a causal relationship between these two variables that goes from the direction of the media factor to the stock returns. The tests do not detect any other causal relationships.

Table 4.10: Granger Causality Test Results for WSJ US

| | F-value | P-value |
|---|---|---|
| **$H_0$: Media Factor does not Granger-cause Stock Return** | 2.45 | 0.04* |
| **$H_0$: Media Factor does not Granger-cause Trading Volume** | 2.58 | 0.02* |
| **$H_0$: Trading Volume does not Granger-cause Stock Return** | 0.65 | 0.74 |
| **$H_0$: Stock Return does not Granger-cause Trading Volume** | 4.05 | 0.00* |
| **$H_0$: Return Does not Granger-cause Media** | 0,22 | 0,95 |
| **$H_0$: Trading Volume Does not Granger-cause Media** | 0,15 | 0,98 |

The p-values presented in Table 4.10 suggest that there is a very pronounced instantaneous causal relationship between the US stock market returns and trading volume. This implies that the trading volume is triggered by stock returns. Also, there is the null hypothesis that no such relation exists can be rejected at the 5 percent significance level. It can be stated that the implied causal relationship goes from the direction of the media factor for both to stock returns and the trading volume.

#### 4.2.3. Impulse Response Functions

The F-tests conducted for Granger causality cannot reveal whether changes in the value of a given variable have a positive or negative effect on the other variables in the system or how long it would take for the effect of that variable to work through the VAR system. The nature of this type of a relationship between the variables can be examined by estimating the impulse responses for the VAR model.

An impulse response function traces the response of the endogenous variables to one standard deviation shock or change to one of the disturbance terms in the system. For each variable from each equation, a unit shock is applied to the error and the effects through time periods are observed. A shock to a variable is transmitted to all of the endogenous variables through the dynamic structure of the VAR. Therefore, an impulse response function shows the interaction between/among the endogenous variable sequences. For identifying the long run effects of structural shocks on the output fluctuations, impulse-response graphs are constructed. More generally, an impulse response refers to the reaction of any dynamic system in response to some external change. Figures 4.1 and 4.2 present the impulse response functions estimated for the VAR system.

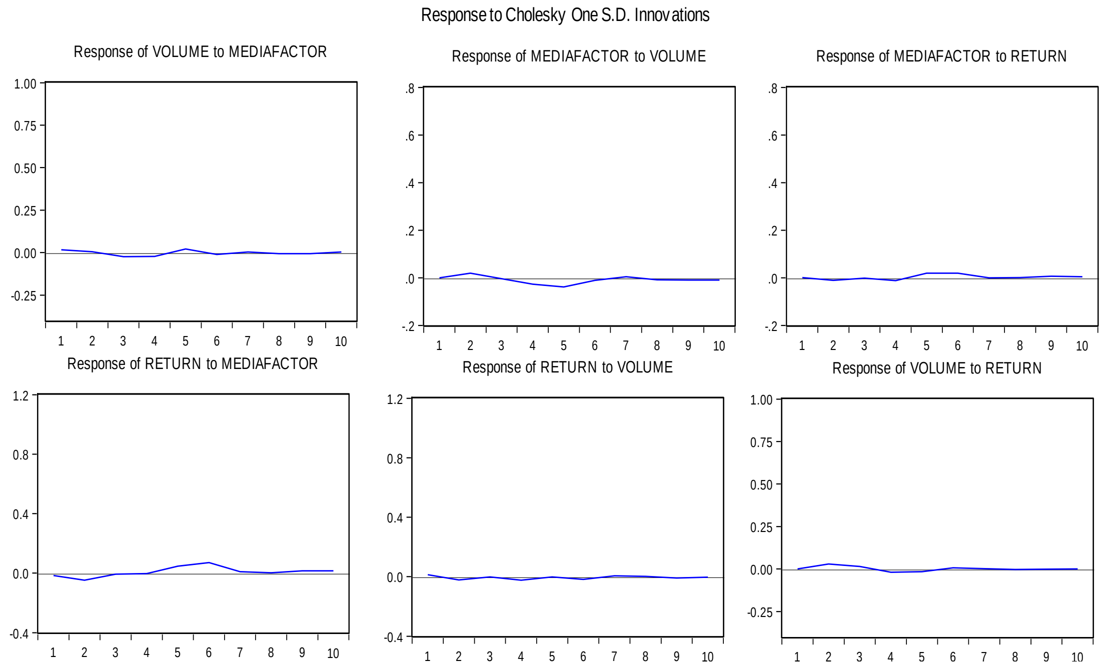


Figure 4.1: Impulse Response Graphs for WSJ Europe

Figure 4.1 shows the response of media, return and volume variables to a one standard deviation shock that occurs in the return, volume and media factor variables. The impulse functions in Figure 4.1 do not demonstrate a very strong response from the variables in the system. The stock return variable seems to have a slight response in days 2 and 6 when there is a one standard deviation shock to the media factor. Same slight response in days from 2 to 6 also absorbed to a one standard deviation shock to the volume variable on media factor variable.

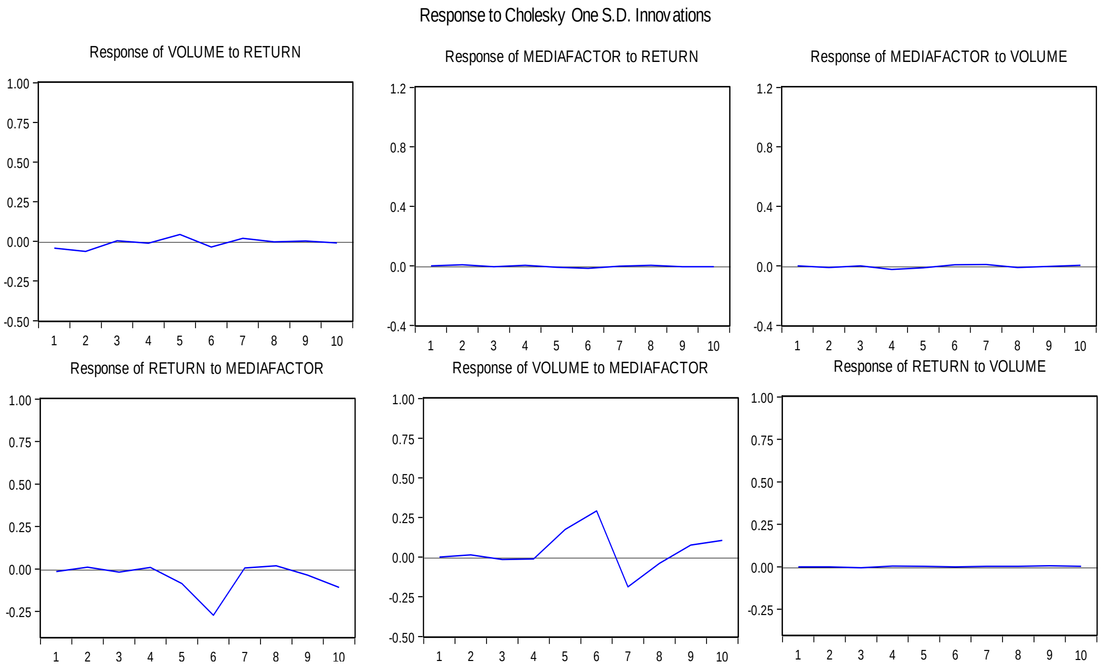


Figure 4.2: Impulse Response Graphs for WSJ US

Contrary to the WSJ Europe sample, the impulse functions estimated for the WSJ US sample detect stronger responses. When there is a one standard deviation shock to the media factor, there is a response from the stock return variable in days 6. 8 and 9 and there is a response from the volume variable in day 6. There is also a slight response from the volume variable to the shock in the stock return variable but the magnitude of this response is a lot smaller compared to the response to shocks in the media factor.

### 4.2.4. VAR Estimates

For the VAR model, all lags up to 8 days for the US market and 5 days for the European market are included in the estimations. As stated in the Methodology chapter, the endogenous variables are the media factor, the stock return and the trading volume. The exogenous variables are the past 60-day volatility of stock returns, dummy variables for day of-the-week, the month of the year and the 2008 crisis.

All exogeneous factors are shown by the symbol Exog and lags are shown by the symbol LX where X defines the lag value. For instance, the estimated VAR equation for the stock return variable can be written as follows:

$Return_t$ = Constant + Coeff 1*LX (Media Factor)+ Coeff 2*LX (Volume)+ Coeff 3*LX (Return)+Coeff5 (Exog) $_{(t-1)}$ (Equation 4.1)

Table 4.11 presents the VAR estimation results for the WSJ Europe sample. For each lagged value of the right-hand-side variables. The top row presents the VAR coefficient estimate and the bottom row presents the coefficient' s t-value.

Table 4.11: VAR Estimates of Return and Volume Variables. WSJ Europe

| | Return | Volume | Media |
|---|---|---|---|
| Media Factor (-1) | -0.064 | 0.017 | 0,089 |
| | -1.629 | 0.561 | 0,497 |
| Media Factor (-2) | 0.001 | -0.027 | 0,062 |
| | 0.006 | -0.918 | 0,445 |
| Media Factor (-3) | -0.006 | -0.038 | 0,069* |
| | -0.152 | -1.283 | 2,730* |
| Media Factor (-4) | 0.072 | 0.022 | 0,128 |
| | 1.861 | 0.751 | 0,028 |
| Media Factor (-5) | 0.096* | -0.011 | 0,084 |
| | 2.469* | -0.359 | 0,296 |
| Return (-1) | 0.095* | 0.027 | -0,010 |
| | 3.729* | 1.370 | -0,580 |
| Return (-2) | -0.106* | 0.021 | -0,001 |
| | -4.147* | 1.078 | -0,033 |
| Return (-3) | 0.033 | -0.009 | -0,011 |
| | 1.281 | -0.478 | -0,668 |

Table 4.11 (continued)

| | | | |
|---|---|---|---|
| Return (-4) | 0.036 | -0.016 | 0,021 |
| | 1.388 | -0.825 | 1,275 |
| Return (-5) | -0.061* | 0.001 | 0,017 |
| | -2.413* | 0.021 | 1,050 |
| Volume (-1) | -0.027 | -0.336* | 0,024 |
| | -0.877 | -14.279* | 1,164 |
| Volume (-2) | -0.005 | -0.078* | 0,001 |
| | -0.152 | -3.074* | 0,062 |
| Volume (-3) | -0.032 | -0.001 | -0,033 |
| | -0.972 | -0.030 | -1,545 |
| Volume (-4) | -0.010 | -0.058* | -0,057* |
| | -0.318 | -2.329* | -2,688* |
| Volume (-5) | -0.032 | 0.106* | -0,026 |
| | -1.027 | 4.459* | -1,309 |
| Volatility | -0.072 | -0.016 | -0,096 |
| | -0.334 | -0.099 | -0,684 |
| Crisis Dummy | -0.110 | 0.005 | 0,081 |
| | -1.546 | 0.100 | 1,735 |
| Weekday | -0.013 | 0.241* | 0,010 |
| | -0.661 | 15.590* | 0,777 |
| Month | 0.005 | -0.001 | 0,001 |
| | 0.626 | -0.157 | 0,172 |
| Constant | 0.111 | -0.706* | 0,042 |
| | 0.522 | -4.331* | 0,305 |

(*) represent the related coefficient are significant at %95 significance level

The bold figures represent those lagged values that have a significant effect on the dependent variables. Table 4.11 shows a number of significant influences on the current stock returns. The media factor from five days ago and the stock returns from one day ago, two days ago and five days ago all have a statistically significant influence on today's stock returns in the WSJ Europe sample. It looks like stock returns take about five days to adjust to information that arrives in the market.

When the dependent variable is the trading volume, it is observed that volume is only affected significantly from its own lagged values and the weekday variables. This can be interpreted as trading volume triggers the market place to trade more or less.

When the dependent variable is the media factor, it is observed that media factor is only affected significantly from its own lagged value and volume lagged value. This also means that trading volume can be an important variable for the news media to produce more pronounced headlines for the investor world. Other investor sentiments that are built via the first level of categorization of GI program are also analyzed using VAR estimation methodology. When 'Media Factor' is taken from Positiv or Negativ tag structure, VAR structures do not catch a valid relation between media to trading volume or market return. Related equation and VAR tables can be found in Appendix F at the end of chapters.

Table 4.12 presents the VAR estimation results for the WSJ US sample. For each lagged value of the right-hand-side variables. The top row presents the VAR coefficient estimate and the bottom row presents the coefficient's t-value.

Table 4.12: VAR Estimates of Return and Volume Variables. WSJ US Sample

| | Return | Volume | Media |
|---|---|---|---|
| Media Factor (-1) | -0.034 | 0.019 | -0,003 |
| | -1.080 | 0.633 | -0,104 |
| Media Factor (-2) | 0.010 | -0.024 | -0,001 |
| | 0.085 | -0.818 | -0,050 |
| Media Factor (-3) | -0.027 | 0.013 | -0,003 |

Table 4.12 (continued)

| | | | |
|---|---|---|---|
| | -0.955 | 0.306 | -0,101 |
| Media Factor (-4) | 0.022 | 0.017 | 0,460 |
| | 0.723 | 0.485 | 1,947 |
| Media Factor (-5) | 0.076* | -0.049 | -0,213 |
| | 1.930* | -1.584 | -0,912 |
| Media Factor (-6) | 0.045 | -0.002 | -0,192 |
| | 1.486 | -0.106 | -0,821 |
| Media Factor (-7) | 0.030 | 0.095* | 0,119 |
| | 1.000 | 2.619* | 0,510 |
| Media Factor (-8) | 0.021 | 0.045 | 0,259 |
| | 0.722 | 1.490 | 1,102 |
| Return (-1) | 0.104* | 0.058* | 0,004 |
| | 4.357* | 2.071* | 0,124 |
| Return (-2) | 0.075* | 0.008 | -0,010 |
| | 2.874* | 0.360 | -0,279 |
| Return (-3) | 0.067* | -0.002 | 0,002 |
| | 2.610* | -0.137 | 0,069 |
| Return (-4) | 0.002 | -0.008 | -0,011 |
| | 0.144 | -0.414 | -0,311 |
| Return (-5) | 0.069* | 0.062* | -0,021 |
| | 2.691* | 2.313* | -0,591 |
| Return (-6) | 0.039 | 0.008 | -0,002 |
| | 1.369 | 0.399 | -0,065 |
| Return (-7) | 0.051* | 0.057 | 0,016 |
| | 1.927* | 1.668 | 0,454 |
| Return (-8) | 0.027 | 0.016 | -0,012 |
| | 0.804 | 0.791 | -0,339 |
| Volume (-1) | -0.027 | -0.536* | -0,001 |
| | -0.299 | -22.797* | -0,017 |
| Volume (-2) | -0.005 | -0.334* | 0,020 |

| | | | |
|---|---|---|---|
| | -0.485 | -14.968* | 0,475 |

Table 4.12 (continued)

| | | | |
|---|---|---|---|
| Volume (-3) | -0.001 | -0.214* | 0,003 |
| | -0.055 | -11.000* | 0,062 |
| Volume (-4) | -0.004 | -0.158* | 0,016 |
| | -0.372 | -8.242* | 0,360 |
| Volume (-5) | -0.002 | -0.098* | 0,070 |
| | -0.157 | -5.593* | 1,552 |
| Volume (-6) | -0.010 | -0.078* | 0,140* |
| | -0.759 | -4.469* | 3,163* |
| Volume (-7) | -0.009 | -0.087* | 0,055 |
| | -0.669 | -4.914* | 1,317 |
| Volume (-8) | -0.035 | -0.071* | 0,018 |
| | -1.789 | -3.807* | 0,493 |
| Volatility | 0.028 | -0.008 | 0,068 |
| | 1.210 | -0.569 | 0,414 |
| Crisis dummy | -0.036 | 0.055* | -0,015 |
| | -1.705 | 2.221* | -0,180 |
| Weekday | -0.019 | 0.016 | -0,028 |
| | -0.968 | 0.799 | -1,296 |
| Month | 0.085 | -0.014 | -0,015 |
| | 4.692 | -0.857 | -1,672 |
| Constant | 0.045 | 0.013 | 0,140 |
| | 2.519 | 0.719 | 0,786 |

(*) represent the related coefficient are significant at %95 significance level

Results in Table 4.12 suggest that stock returns in the US market are mostly affected by lagged stock returns. In addition, the media factor from five days prior also seems to have a significant impact on today’s stock returns. Similar to the WSJ Europe sample, returns take about five days to adjust to relevant information that arrives in the market. Looking at the volume equations, it is observed that, unlike the WSJ Europe results, the current volume in the US market seems to be significantly affected by the stock returns generated one day and five days ago. News that arrived in the market as early as seven days ago also has a significant effect on the trading volume. Finally, the US market volume seems to have been affected from the 2008 financial crisis.

When the dependent variable is the media factor, it is observed that media factor is only affected significantly from its volume-lagged value. This also means that trading volume can be an important variable for the news media to produce more pronounced headlines for the investor world.

Other investor sentiments that are built via the first level of categorization of GI program are also analyzed using VAR estimation methodology. When ‘Media Factor’ is taken from Positiv or Negativ tag structure, VAR structures do not catch a valid relation between media to trading volume or market return. Related equation and VAR tables can be found in Appendix F at the end of chapters.

# CHAPTER 5

# CONCLUSION

Stock exchanges are important engines of modern financial life. This is also true of news media. Journalistic products, especially the real-time news, are of key importance in modern business life. Seconds are vital in selling or buying stocks or other financial instruments. Stock exchanges and business media have been linked to each other for a long time, but these relationships are now being reexamined in light of advancements in the behavioral finance literature.

In this study, the aim is to reveal the impact of the media factor on market returns. The "Heard on the Street" columns are used to develop a proxy for the market sentiment. Readers of this column are supposed to be informed as well as uninformed individual investors who rely on analyst advice to make investment decisions and are subject to estimation biases or susceptible to the impact of sentiment. At first, considering the US and Europe market, evidence is presented which suggests that the media factor has an impact on stock returns. A VAR model upon the media factor, stock returns and volume is estimated and the findings imply that the return generating process in the current time period is mainly affected by the past returns, while media factor also has a significant effect.

## REFERENCES


Antweiler, W., Frank, M (2005), ‘Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards’, *Journal of Finance,* 59, 1259-1294.

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

Barber, B, Loeffler D. (1993), ‘The ‘Dartboard’ Column: Second-Hand Information and Price Pressure’, *Journal of Financial and Quantitative Analysis,* 28, 273-284.

Barber, B., 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., Vishny R. (1998), “A model of Investor Sentiment”, *Journal of Finance,* 49, 307-343.

Banz, R.W. (1981), "The Relationship Between Return and Market Value of Common Stocks", *Journal of FinancialEconomics.*

Berelson. B. (1952), Content Analysis in Communication Research.

Bernard, V., Thomas, J. (1990), ‘Evidence that Stock Prices do not Fully Reflect the Implications of Current Earnings for Future Earnings’, *Journal of Accounting and Economics*, 13, 305-340.

Bhattacharya, Utpal, Yu, X. (2008), ‘The Causes and Consequences of Recent Financial Market Bubbles: An Introduction’, *Review of Financial Studies*, 21, 3-10.

Black, F. (1986), ‘Noise’, *Journal of Finance,* 41, 529-543.

Buse J. (2002), ‘Market efficiency in real time’, *Journal of Economics*, 65, 415-437.

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

Cutler, D. M., J. M. Poterba, and L. H. Summers (1989), ‘What Moves Stock Prices?’, *Journal of Portfolio Management,* 15, 4-12.

De Long, Bradford J., Shleifer, A., Summers, H., Waldmann, R. (1990), ‘Noise Trader Risk in Financial Markets’, *Journal of Political Economy,* 43, 106-120.

Demers, E., Vega, C. (2011), ‘Linguistic Tone in Earnings Announcements: New sor Noise’, *FRB International Finance Discussion Paper*, 951, from web sited lastly on 10 September 2012 http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1152326.

Dougal, C., Engelberg, J., Garcia, D., A. C. Parsons, C. (2011), ‘Journalits and the Stock Market’, last sited at 10 September 2012 from http://ssrn.com/abstract=1784431.

Edmans, A., Garcia, D., Norli (2007), ‘Sport Sentiment and Stock Returns’, *The Journal of Finance*, 62, 4, 1967–1998.

Engelberg, J., C. A. Parsons (2009), ‘The Causal impact of Financial Markets’, last sited at 10 September 2012 from http://ssrn.com/abstract=1462416.

De Bondt, W., Thaler, R. (1985), ‘Does the Stock Market Overreact?’, *Journal of Finance,* 40, 793-805.

De Bondt, Thaler, R. (1987), ‘Further Evidence on Investor Overreaction and Stock Market Seasonality’, *Journal of Finance,* 42, 3, 557-581.

DeLong, J., Shleifer, A., Summers, L., Waldmann, R., (1990), ‘Noise trader risk in financial markets’, *Journal of Political Economy,* 98, 703–738.

Fama, E. (1970), ‘Efficient capital markets: a review of theory and empirical work’, *Journal of Finance,* 25, 383–417.

Fama E.F. and French K.R. (1992), ‘The cross section of expected stock returns’, *The Journal of Finance,* 47, 427–65.

Hirshleifer, D., Shumway, T. (2003), ‘Stock returns and the Weather’, *Journal of Finance,* 58-3, 1009–1032.

Hong, H., Kubik, J., Stein, J. (2005), ‘Thy Neighbor’s Portfolio: Word-of –Mouth Effects in the Holdings and Trades of Money Managers’, *The Journal of Finance*, LX, 6.

Hurwitz, R., Stone, P., Wawer, A., General Inquirer Online, Cambridge MA: Harvard University.

Jones, C. P., R.H. Litzenberger. (1970), ‘Quarterly Earnings Reports and Intermediate Stock Price Trends’, *Journal of Finance*, 25, 143-148.

Jogerson, Visssing, A. (2002), ‘Towards an Explanation of Household Portfolio Choice Heteregeneity: Nonfinancial Income and Participation Cost Structure’, *NBER Working Paper*, 8884.

Krippendorff, K. (1980), ‘Content Analysis: An Introduction to Its Methodology’, Newbury Park, CA: Sage.

Liu, P., S. D. Smith D., A. A. Syed (1990), ‘Stock Reaction to the Wall Street Journal’s Securities Recommendations’, *Journal of Finance and Quantitative Analysis,* 25, 399-410.

Llyod-Davies, P., Canes, M., ‘Stock price reaction to Wall Street Journal’s Securities Recommendation’, *Journal of Finance and Quantitative Analysis,* 25, 43-56.

Loughran, T., and B. McDonald (2011), ‘When is a liability not a liability?, *Journal of Finance,* 66, 35-65.

Merton, Robert, C. (1987),’ A Simple Model of Capital Market Equilibrium with Incomplete Information’, *Journal of Finance,* 42, 483-510.

Ojala J., Uskali T., (2006), ‘Any Weak Signals?’, The New York Times and the Stock Market Crashes of 1929, 1987 and 2000, *XIV International Economic History Congress,* Helsinki 2006 Session 1.

Parsons C., Engelberg J., (2011), “The Causal Impact of Media in Financial Markets”, *The Journal of Finance,* LXVI, 1.

Pu, Liu, Stanley, D. Smith, Syed, A. (1992), ‘The Impact of the Insider Trading Scandal on the information Content of the Wall Street Journal’s Heart on the Street Column’, The Journal of Financial Research, 15, 2.

Rozeff, M.S., Kinney, W.R. (1976), Capital Market Seasonality: The Case of Stock Returns, *Journal of Financial Economics,* 3, 379-402.

Shefrin, H. (2000), ‘Beyond Greed and Fear: Understanding Behavioral Finance and the Psychology of Investing’, *Financial Management Association Survey and Synthesis Series,* Harvard Business School Press.

Shleifer, A. (2000), Inefficient Markets: A Introduction to Behavioral Finance. Oxford: Oxford University Press.

Shleifer, A., Summers, L. (1990), ‘The noise trader approach to finance’, *Journal of Economic Perspectives,* 4, 19–33.

Shleifer, A., Vishny, R. (1997), ‘The Limits of Arbitrage’, *Journal of Finance,* 52, 1, 35-55.

Sudit, E., Palmon, D., Yezegel, A. (2009), 'The Value of Columnists' Stock Recommendations', *Review of Quantitative Finance and Accounting*, 33, 3, 209-232.

Shiller, R. J. (2000), 'Irrational Exuberance', *Princeton University Press*, Princeton.

Tetlock, P. (2007), 'Giving content to investor sentiment: The role of media in the stock market', *The Journal of Finance, 62,* 3, 1139-1168.

Tversky, A., Kahnemann, D., (1974), 'Judgment under uncertainty: heuristics and biases', *Science*, 185, 1124–1131.

Weber, R. P. (1990), 'Basic Content Analysis', 2nd ed. Newbury Park, CA.

# APPENDICES

## APPENDIX A. GENERAL INQUIRER ONLINE SOFTWARE

The access to the General Inquirer software is provided by its online server. Below Figure A.1 is the initial program input feeding page. To access the program, assigned id and password is needed.

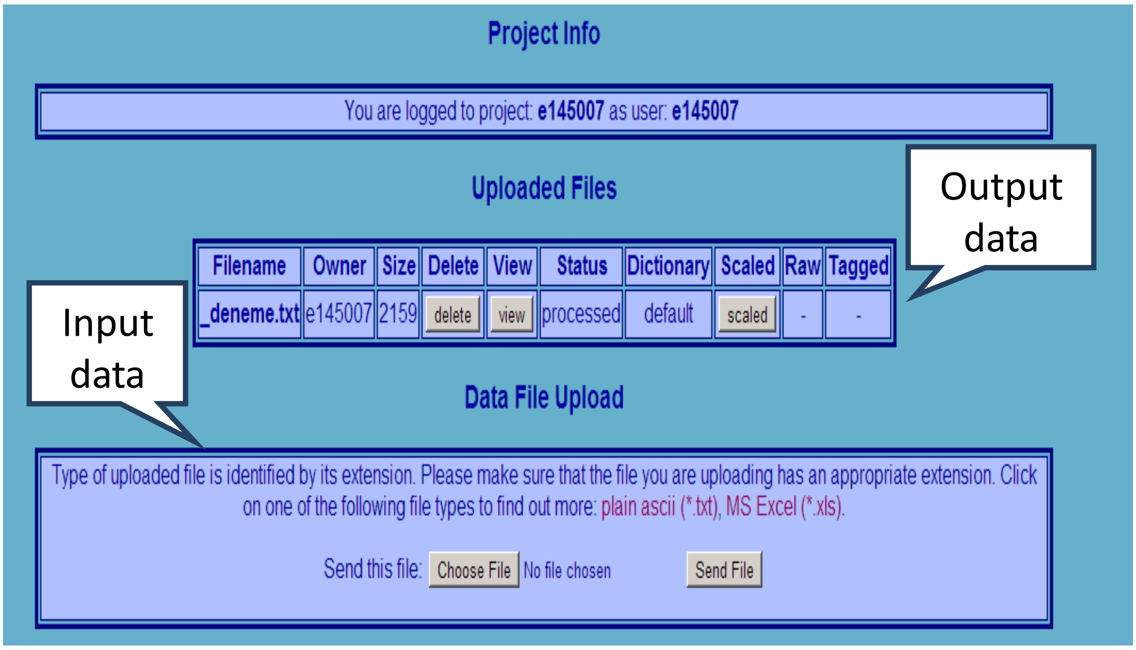


Figure A.1: General Inquirer Software

## APPENDIX B. DATA PREPARATION FLOW FOR GENERAL INQUIRER

Following steps are the data preparation work flow that is highly recommended to be taken to have a consistent GI output. In this study, to construct media sentiment, GI outputs are highly critical; to avert any error this work flow is used.

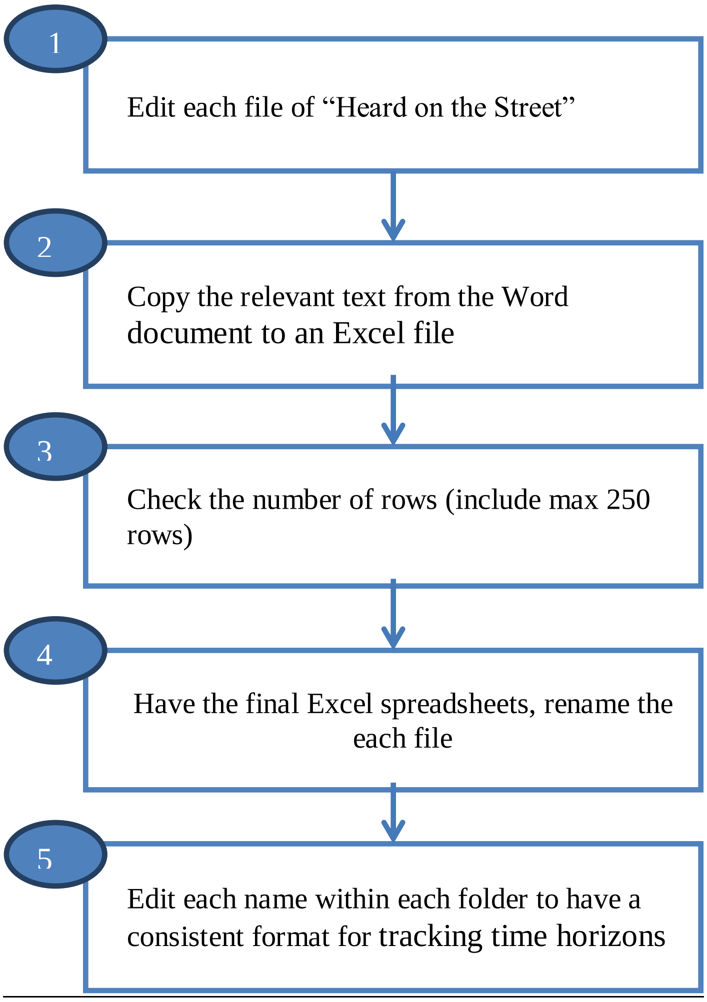


Figure B.1: Data Preparation Work Flow

## APPENDIX C. SAMPLE NEWS ARTICLE OF "HEARD ON THE STREET" FROM FACTIVA DATABASE

Heard on the Street Capital

One's Moves on Probe, Resignation Raise New Queries
By Carrick Mollenkamp 1,247 words
The Wall Street Journal Europe
WSJEM1English (Copyright (c) 2003, Dow Jones & Company, Inc.)

ONE WEEK AGO today, many investors were shocked when Capital One Financial Corp. publicly disclosed its chief financial officer had resigned after U.S. securities regulators told him he could face civil insider-trading charges. Others apparently weren't so surprised. Over the previous weekend, the McLean, Virginia, issuer of credit cards had briefed some investors who buy Capital One bonds about the investigation and the resignation of the executive, David Willey. In exchange for the information, the investors signed confidentiality agreements, according to several people who have discussed the matter with the company. Such confidentiality pacts generally seek to prevent trading on the revealed information, securities lawyers say. Still, it isn't clear how many people were told the news in advance, and the conversations add to questions about how Capital One handled the entire affair, including whether it acted quickly enough in publicly telling investors about the insider-trading probe of Mr. Willey. Securities regulators had notified the company last August that they were conducting an investigation, the company said in its announcement last week and in a subsequent interview. Unlike those with advance information, the great majority of investors had to deal with the revelations along with their first cup of coffee last Monday when the news crossed wire services shortly after 8 a.m. Capital One stock fell sharply when trading opened, and finished the day down nearly 9%. In a written statement late Friday in response to questions about the early heads up, Capital One said that its disclosures were handled appropriately and in compliance with securities laws. "The company had discussions with certain of its lenders under its corporate credit facility, advisors, rating agencies, and regulators shortly before the public announcement," the statement said. "Some of the lenders and advisors are also asset-backed securities investors. The discussions were entirely appropriate and subject to confidentiality obligations. No investor was disadvantaged in any way." A Capital One spokeswoman defended the silence this past August, saying the company didn't consider the U.S. Securities and Exchange Commission notification of an investigation to be a material event. The company's heads-up to some investors surfaced at a lunch last Tuesday at the 21 Club in Manhattan, when Capital One was meeting with investors, including stockholders who weren't in on the early disclosure, according to a person who attended the lunch. There, a Capital One executive said that some holders of its asset-backed securities had been told earlier than other investors. Another person who spoke separately with the company says he understood Capital One had talked to holders of the packaged credit-card debt to avoid any panic in the asset-backed market upon Monday's announcement of the finance official's departure. "I just find that amazing," Kathy

Shanley, an analyst with independent credit-research firm Gimme Credit, says of a heads-up to some investors. "I have never heard of something quite like that." Larry Robbins, a Raleigh, North Carolina, securities attorney at Wyrick, Robbins, Yates & Ponton, says Capital One appears to have avoided violating disclosure laws by using the confidentiality agreements. "Once they bring them over the wall, they can't sell" their Capital One holdings, he says. A focus of the regulatory probe may be a stock transaction, recorded in securities filings, by Mr. Willey this past May involving about 52,075 shares. The transaction was valued at $3.2 million (2.9 million euros), filings show. Richard Morvillo, a Washington attorney for Mr. Willey, said Mr. Willey denies engaging in the purchase or sale of securities based on material nonpublic information and would fight any SEC action vigorously. In May, Capital One shares traded on the New York Stock Exchange in a range of $57 to $64, down slightly from a 52-week high of $66.50 in April but far above the current price. At 4 p.m. Friday, the shares were up 2.1%, or 56 cents, to $27.48 in Big Board composite trading. In mid-July, Capital One's stock fell sharply after the company announced that bank examiners had "identified certain supervisory issues." These included findings that the company needed to bolster allowances for loan losses and to improve the technology it uses to assess credit risk for the borrowers signing up for Capital One's credit cards. On the news, Capital One's stock tumbled $20.12, or about 40%, to $30.48. Like stocks of other credit-card issuers, shares of Capital One also have been hurt in recent months by concerns in general about mounting consumer-credit defaults. An SEC Spokesman declined to comment on what prompted the agency to take a look at trading in Capital One shares and when. But Capital One, in its announcement and a subsequent interview, confirmed that, as of August, the SEC had notified the company that it was undertaking a formal investigation. It isn't known how the notification was worded. Securities attorneys say such notices often are vague. That can make it difficult for a company to decide whether to disclose to investors that it is being investigated, some securities-law specialists say. "This whole concept of when do you get to something that is material, there is not a bright-line test," says Mr. Robbins, the Raleigh attorney. Some attorneys say a formal investigation should be disclosed no matter how vague the SEC notice is. "Once a formal order is received, it does represent a new level of commitment by the SEC to go forward. It's prudent to reconsider whatever earlier disclosure decisions you made," says Randy Eaddy, a securities attorney in the Atlanta office of Kilpatrick Stockton. "In my judgment, there is no good to come from not disclosing the matter." By September, Mr. Willey, who became Capital One chief financial officer in 2001, had hired Mr. Morvillo's firm, Crowell & Moring, Mr. Morvillo, a white-collar defense attorney who formerly had worked at the SEC, confirmed in an interview last week. By late December, Mr. Willey had notified the company that he planned to leave Capital One, according to both the company and Mr. Morvillo last week. By late 2002, the company had hired a search firm to find his replacement, the company said last week. Mr. Morvillo said in the interview that Mr. Willey had been exploring the possibility of resigning for several months.

# APPENDIX D. SAMPLE OUTPUT FROM GENERAL INQUIRER

Table D.1: Sample 1 Output Description

| | |
|---|---|
| unit_id | 1 |
| word_count | 492 |
| Leftover | 33 |

| | | | |
|---|---|---|---|
| Positiv | 44 | Active | 49 |
| Negativ | 33 | Passive | 21 |
| Pstv | 13 | Pleasur | 1 |
| Affil | 10 | Pain | 0 |
| Ngtv | 9 | Feel | 0 |
| Hostile | 5 | Arousal | 1 |
| Strong | 60 | EMOT | 1 |
| Power | 15 | Virtue | 5 |
| Vice | 9 | Ovrst | 31 |
| Weak | 15 | Undrst | 18 |
| Submit | 0 | | |

## APPENDIX E. SAMPLE STATISTICS TABLE FOR MEDIA SENTIMENT

The Table E.1 and E.2 reports the sample statistics for the media content measures used in the paper. These measures are constructed from the columns of "Heard on the Street" from Wall Street Journal in US and Europe in 10-year period. "Positive" and "Negative" measures are constructed by counting the number of positive and negative words and then normalizing by total number of words in the article. Content analysis has evaluated under the dictionary of GI. The "Pessimism" variable is the difference between the "Negative" and "Positive" measures. All numbers are in percentage units.

Table E.1: Sample statistics of media content variable extracted from WSJ US

| | Mean | Median | 25% Percentil e | Standard Deviation |
|---|---|---|---|---|
| Positiv | 7,0% | 6,6% | 20,0% | 0,03 |
| Negativ | 2,7% | 2,2% | 14,6% | 0,02 |
| MediaFactor | 24,0% | 23,3% | 51,9% | 0,07 |

Media Factor has been driven from the PCA analysis

Table E.2: Sample statistics of media content variable extracted WSJ Europe

| | Mean | Median | 25% Percentile | Standard Deviation |
|---|---|---|---|---|
| Positiv | 4,80% | 4,73% | 3,92% | 0,01 |
| Negativ | 3,92% | 3,83% | 2,95% | 0,01 |
| MediaFactor | 10,42% | 10,14% | 22,71% | 0,03 |

Media Factor has been driven from the PCA analysis

## APPENDIX F. VAR REGRESSION OUTPUT

Table F.1 shows the VAR analysis of US results of investigated endogeneous variables when media factor is taken from the first level of categorization as positive from GI software.

Table F.1: VAR analysis of US with positive media factor

| | Return | Volume | Media |
|---|---|---|---|
| Media Factor (-1) | 0.005 | 0.008 | -0.006 |
| | 0.557 | 0.917 | -0.250 |
| Media Factor (-2) | -0.007 | 0.002 | -0.001 |
| | -0.787 | 0.220 | -0.042 |
| Media Factor (-3) | 0.004 | -0.011 | 0.000 |
| | 0.466 | -1.227 | 0.016 |
| Media Factor (-4) | -2.364 | 2.784 | -2.057 |
| | -0.742 | 0.888 | -0.243 |
| Media Factor (-5) | 1.562 | 2.341 | -2.716 |
| | 0.487 | 0.743 | -0.318 |
| Media Factor (-6) | -2.149 | 9.496 | -1.417 |
| | -0.674 | 0.303 | -1.670 |
| Media Factor (-7) | 2.320 | -3.680 | 1.062 |
| | 0.721 | -1.163 | 1.240 |
| Media Factor (-8) | 4.719 | 2.544 | 6.917 |
| | 0.148 | 0.809 | 0.813 |
| Return (-1) | -0.052 | -0.108 | 0.012 |
| | -2.092 | -4.445 | 0.178 |
| Return (-2) | -0.011 | -0.072 | -0.012 |
| | -0.440 | -2.920 | -0.185 |
| Return (-3) | -0.003 | -0.067 | 0.011 |
| | -0.113 | -2.729 | 0.159 |
| Return (-4) | 0.011 | -0.003 | -0.016 |
| | 0.429 | -0.152 | -0.240 |
| Return (-5) | -0.056 | -0.063 | -0.042 |
| | -2.273 | -2.623 | -0.644 |
| Return (-6) | -0.012 | -0.031 | -0.012 |
| | -0.485 | -1.287 | -0.185 |
| Return (-7) | -0.038 | -0.048 | 0.022 |
| | -1.549 | -1.951 | 0.328 |
| Return (-8) | 0.019 | 0.019 | -0.012 |
| | 0.793 | 0.791 | -0.180 |

Table F.1 (continued)

| | | | |
|---|---|---|---|
| Volume (-1) | -0.008 | -0.568 | 0.004 |
| | -0.296 | -22.644 | 0.059 |
| Volume (-2) | -0.015 | -0.423 | 0.025 |
| | -0.519 | -14.807 | 0.328 |
| Volume (-3) | 0.000 | -0.327 | 0.002 |
| | -0.025 | -10.933 | 0.019 |
| Volume (-4) | -0.009 | -0.252 | 0.022 |
| | -0.291 | -8.151 | 0.264 |
| Volume (-5) | -0.012 | -0.171 | 0.135 |
| | -0.372 | -5.526 | 1.611 |
| Volume (-6) | -0.023 | -0.142 | 0.259 |
| | -0.744 | -4.664 | 3.133 |
| Volume (-7) | -0.022 | -0.142 | 0.116 |
| | -0.751 | -4.884 | 1.473 |
| Volume (-8) | -0.047 | -0.094 | 0.040 |
| | -0.008 | -0.568 | 0.004 |
| Volatility | -0.067 | 0.139 | 0.114 |
| | -0.577 | 1.213 | 0.366 |
| Crisis dummy | -0.132 | -0.070 | -0.062 |
| | -2.347 | -1.262 | -0.414 |
| Weekday | -0.013 | 0.068 | -0.050 |
| | -0.890 | 4.621 | -1.247 |
| Month | 0.004 | -0.008 | -0.028 |
| | 0.687 | -1.343 | -1.680 |
| Constant | 0.066 | 0.610 | 0.038 |
| | 0.094 | 0.878 | 0.020 |

Table F.2 shows the VAR analysis of Europe results of investigated endogeneous variables when media factor is taken from the first level of categorization as negative from GI software.

Table F.2: VAR analysis of US with negative media factor

| | Return | Volume | Media |
|---|---|---|---|
| Media Factor (-1) | -0.064 | 0.017 | 0.089 |
| | -1.629 | 0.561 | 3.497 |
| Media Factor (-2) | 0.000 | -0.027 | 0.062 |
| | 0.006 | -0.918 | 2.445 |
| Media Factor (-3) | -0.006 | -0.038 | 0.069 |
| | -0.152 | -1.283 | 2.730 |

Table F.2 (continued)

| | | | |
|---|---|---|---|
| Media Factor (-4) | 0.072 | 0.022 | 0.128 |
| | 1.861 | 0.751 | 5.028 |
| Media Factor (-5) | 0.096 | -0.011 | 0.084 |
| | 2.469 | -0.359 | 3.296 |
| Return (-1) | 0.095 | 0.027 | -0.010 |
| | 3.729 | 1.370 | -0.580 |
| Return (-2) | -0.106 | 0.021 | -0.001 |
| | -4.147 | 1.078 | -0.033 |
| Return (-3) | 0.033 | -0.009 | -0.011 |
| | 1.281 | -0.478 | -0.668 |
| Return (-4) | 0.036 | -0.017 | 0.021 |
| | 1.388 | -0.825 | 1.275 |
| Return (-5) | -0.062 | 0.001 | 0.018 |
| | -2.413 | 0.021 | 1.050 |
| Volume (-1) | -0.027 | -0.336 | 0.023 |
| | -0.877 | -14.279 | 1.164 |
| Volume (-2) | -0.005 | -0.078 | 0.001 |
| | -0.152 | -3.074 | 0.062 |
| Volume (-3) | -0.032 | -0.001 | -0.033 |
| | -0.972 | -0.030 | -1.545 |
| Volume (-4) | -0.010 | -0.059 | -0.057 |
| | -0.318 | -2.329 | -2.689 |
| Volume (-5) | -0.032 | 0.106 | -0.027 |
| | -1.027 | 4.459 | -1.309 |
| Volatility | -0.072 | -0.012 | -0.096 |
| | -0.334 | -0.099 | -0.684 |
| Crisis dummy | -0.110 | 0.005 | 0.081 |
| | -1.546 | 0.100 | 1.735 |
| Weekday | -0.013 | 0.241 | 0.010 |
| | -0.661 | 15.590 | 0.777 |
| Month | 0.005 | -0.001 | 0.001 |
| | 0.626 | -0.157 | 0.172 |
| Constant | 0.111 | -0.706 | 0.042 |
| | 0.522 | -4.331 | 0.305 |

Table F.3 shows the VAR analysis of Europe results of investigated endogeneous variables when media factor is taken from the first level of categorization as positive from GI software.

Table F.3: VAR analysis of Europe with positive media factor

Table F.3 (continued)

| | Return | Volume | Media |
|---|---|---|---|
| Media Factor (-1) | - 0.014 | - 0.031 | 0.002 |
| | - 0.482 | - 1.432 | 0.059 |
| Media Factor (-2) | - 0.022 | 0.031 | 0.002 |
| | - 0.785 | 1.429 | 0.088 |
| Media Factor (-3) | - 0.003 | 0.008 | 0.029 |
| | - 0.093 | 0.379 | 1.142 |
| Media Factor (-4) | - 0.027 | - 0.004 | 0.023 |
| | - 0.952 | - 0.175 | 0.888 |
| Media Factor (-5) | 0.003 | 0.003 | 0.008 |
| | 0.118 | 0.117 | 0.313 |
| Return (-1) | 0.099 | 0.027 | - 0.019 |
| | 3.870 | 1.402 | - 0.804 |
| Return (-2) | - 0.106 | 0.021 | - 0.012 |
| | - 4.116 | 1.071 | - 0.519 |
| Return (-3) | 0.032 | - 0.010 | - 0.025 |
| | 1.236 | - 0.483 | - 1.057 |
| Return (-4) | 0.034 | - 0.016 | - 0.016 |
| | 1.330 | - 0.821 | - 0.669 |
| Return (-5) | - 0.065 | 0.001 | - 0.004 |
| | - 2.535 | 0.067 | - 0.159 |
| Volume (-1) | - 0.029 | - 0.334 | - 0.026 |
| | - 0.945 | - 14.176 | - 0.939 |
| Volume (-2) | - 0.011 | - 0.078 | - 0.041 |
| | - 0.331 | - 3.107 | - 1.374 |
| Volume (-3) | - 0.039 | - 0.003 | - 0.055 |
| | - 1.171 | - 0.111 | - 1.813 |
| Volume (-4) | - 0.010 | - 0.060 | - 0.030 |
| | - 0.317 | - 2.398 | - 0.994 |
| Volume (-5) | - 0.029 | 0.106 | 0.047 |
| | - 0.936 | 4.471 | 1.667 |
| Volatility | - 0.086 | - 0.015 | - 0.286 |
| | - 0.396 | - 0.089 | - 1.442 |
| Crisis dummy | - 0.093 | 0.001 | 0.043 |
| | - 1.317 | 0.009 | 0.672 |
| Weekday | - 0.013 | 0.240 | - 0.003 |
| | - 0.661 | 15.521 | - 0.186 |
| Month | 0.006 | - 0.001 | 0.009 |
| | 0.664 | - 0.143 | 1.156 |
| Constant | 0.120 | - 0.705 | 0.229 |
| | 0.559 | - 4.313 | 1.170 |

Table F.4 shows the VAR analysis of US results of investigated endogeneous variables when media factor is taken from the first level of categorization as negative from GI software.

Table F.4: VAR analysis of Europe with negative media factor

| | Return | Volume | Media |
|---|---|---|---|
| Media Factor (-1) | 0.029 | 0.044 | -0.007 |
| | 0,598 | 0,930 | -0.289 |
| Media Factor (-2) | -0.040 | 0.010 | -0.002 |
| | -0.832 | 0,202 | -0.064 |
| Media Factor (-3) | 0.019 | -0.055 | 0.000 |
| | 0,395 | 0,116 | 0,011 |
| Media Factor (-4) | 2.046 | 5.549 | 4.631 |
| | 0,224 | 0,614 | 0,980 |
| Media Factor (-5) | -2.471 | 4.776 | -4.011 |
| | -0.274 | 0.536 | -0.861 |
| Media Factor (-6) | 1.281 | -1.173 | -3.902 |
| | 0.142 | -0.132 | -0.837 |
| Media Factor (-7) | 1.580 | -7.921 | -2.841 |
| | 0.175 | -0.887 | -0.608 |
| Media Factor (-8) | -1.295 | 4.898 | 0.189 |
| | -0.143 | 0.545 | 0.041 |
| Return (-1) | -0.051 | -0.107 | 0.001 |
| | -0.207 | -0.439 | 0.114 |
| Return (-2) | -0.010 | -0.073 | -0.003 |
| | -0.408 | -0.299 | -0.237 |
| Return (-3) | -0.004 | -0.065 | 0.002 |
| | -0.179 | -0.263 | 0.146 |
| Return (-4) | 0.009 | -0.004 | -0.004 |
| | 0.344 | -0.142 | -0.330 |
| Return (-5) | -0.054 | -0.065 | -0.008 |
| | -0.218 | -0.267 | -0.644 |
| Return (-6) | -0.011 | -0.033 | -0.001 |
| | -0.459 | -0.134 | -0.076 |
| Return (-7) | -0.043 | -0.045 | 0.004 |
| | -0.173 | -0.185 | 0.315 |
| Return (-8) | 0.022 | 0.017 | -0.003 |
| | 0.904 | 0.710 | -0.200 |
| Volume (-1) | -0.009 | -0.567 | 0.000 |
| | -0.344 | -0.226 | 0.000 |
| Volume (-2) | -0.015 | -0.423 | 0.005 |
| | -0.508 | -0.148 | 0.320 |
| Volume (-3) | -0.001 | -0.327 | 0.000 |

Table F.4 (continued)

| | | | |
|---|---|---|---|
| | -0.040 | -0.109 | 0.014 |
| Volume (-4) | -0.007 | -0.253 | 0.006 |
| | -0.226 | -0.810 | 0.375 |
| Volume (-5) | -0.014 | -0.172 | 0.027 |
| | -0.449 | -0.553 | 0.164 |
| Volume (-6) | -0.023 | -0.143 | 0.050 |
| | -0.729 | -0.467 | 0.313 |
| Volume (-7) | -0.022 | -0.143 | 0.022 |
| | -0.750 | -0.490 | 0.143 |
| Volume (-8) | -0.050 | -0.095 | 0.007 |
| | -0.197 | -0.376 | 0.542 |
| Volatility | -0.053 | 0.119 | 0.028 |
| | -0.454 | 0.100 | 0.462 |
| Crisis dummy | -0.132 | -0.063 | -0.012 |
| | -0.230 | -0.114 | -0.432 |
| Weekday | -0.012 | 0.068 | -0.009 |
| | -0.824 | 0.457 | -0.110 |
| Month | 0.004 | -0.009 | -0.006 |
| | 0.628 | -0.145 | -0.172 |
| Constant | -0.317 | -0.134 | -0.078 |
| | -0.105 | -0.449 | -0.500 |

# APPENDIX G

# TEZ FOTOKOPİ İZİN FORMU

**ENSTİTÜ**

Fen Bilimleri Enstitüsü

Sosyal Bilimler Enstitüsü

Uygulamalı Matematik Enstitüsü

Enformatik Enstitüsü

Deniz Bilimleri Enstitüsü

**YAZARIN**

Soyadı: Giray

Adı: Aynur

Bölümü: İşletme Bölümü

**TEZİN ADI** (İngilizce)

INFORMATION IN THE FINANCIAL NEWS: EFFECT OF MARKET COMMENTARY ON STOCK MARKET PERFORMANCE

**TEZİN TÜRÜ**: Yüksek Lisans X Doktora

1. Tezimin tamamı dünya çapında erişime açılsın ve kaynak gösterilmek şartıyla tezimin bir kısmı veya tamamının fotokopisi alınsın.
2. Tezimin tamamı yalnızca Orta Doğu Teknik Üniversitesi kullanıcılarının erişimine açılsın. (Bu seçenekle tezinizin fotokopisi ya da elektronik kopyası Kütüphane aracılığı ile ODTÜ dışına dağıtılmayacaktır.)
3. Tezim bir (1) yıl süreyle erişime kapalı olsun. (Bu seçenekle tezinizin fotokopisi ya da elektronik kopyası Kütüphane aracılığı ile ODTÜ dışına dağıtılmayacaktır.)

Yazarın imzası Tarih: