Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

Ph.D. in General Management

Ciclo XXVII

Settore Concorsuale di afferenza: 13 B4

Settore Scientifico disciplinare: SECS P-09

ESSAYS IN MEDIA AND FINANCE

Presentata da: **Carlo Raimondo**

Coordinatore Dottorato

Prof.ssa Rosa Grimaldi

Relatore

Prof. Emanuele Bajo

Correlatori

Prof. Marco Bigelli

Prof.ssa Malvina Nissim

Esame finale anno 2015

Carlo Raimondo - Ph.D. in General Management
Department of Management, University of Bologna - carlo.raimondo@unibo.it

Ph.D. Thesis

Essays in Media and Finance

A Collection of three Research Papers:

This Ph.D. thesis consists in three research papers focused on the relationship between media industry and the financial sector. The importance of a correct understanding what is the effect of media on financial markets is becoming increasingly important as long as fully informed markets hypothesis has been challenged and disputed. Therefore, if financial markets do not have access to complete information, the importance of information professionals, i.e. the media firms, follows. On the other side, another challenge for economic and finance scholar is to understand to which extent financial features are able to influence media firms and consequentially to condition information disclosure and spread.

The main aim of this Ph.D. dissertation is to contribute to a better comprehension both the phenomena. The first paper, entitled "The Media Ownership Effect", analyzes the effects of owning equity shares in a newspaper-publishing firm. The main findings show how for a firm being part of the ownership structure of a media firm ends to receive more and qualitative better coverage. This confirms the view in which owning a media outlet is a source of conflicts of interest. The second paper focuses on the effect of media-delivered information on financial markets. In the framework of initial public offering in the U.S. stock market, we found empirical evidence of a significant effect of the media role in the IPO pricing. Specifically, increasing the quantity and the quality of the coverage increases the first-day returns (i.e. the underpricing). Finally the third paper tries to summarize what has been done in studying the relationship between media and financial industries, putting together contributes from economic, business, and financial scholars.

The main finding of this dissertation is therefore to have underlined the importance and the effectiveness of the relationship between media industry and the financial sector, contributing to the stream of research that investigates about the media role and media effectiveness in the financial and business sectors.

The Media Ownership Effect

Carlo Raimondo*

Abstract

This paper analyzes the effects of particular financial characteristics in the media industry, with a specific focus on the conflicts of interest that may arise along with the ownership structures of media firms. In a sample of 127.190 articles published by the leading Italian newspapers in the period between 2007 and 2011 we find extensive empirical evidence that firms owing shares in a newspaper-publishing company receive a better "treatment" on the newspaper itself, both in terms of quality and quantity of coverage they receive. This paper also contributes to the textual analysis methodology applied to financial contents providing the first content analysis method to analyze Italian financial texts.

I. Introduction

In which extents is finance able to influence the media industry? Is single media's reputation a sufficient counterweight to make the media resist to external powerful pressures? Answering these questions is the key point of this research paper. To address these issues I analyze a central topic related to the media industry: the media ownership. Firms and individuals owning a media firm may be tempted to use their power to obtain a better or a greater coverage. Being positively covered by media could produce desirable outcomes in terms of reputation and appearance. The usual argument proposed against this framework involves media reputation and competitiveness. A media who decides to publish news in favor of its owner(s) could easily noticed by the public and the public itself could decide to change news dealer in order to avoid the *biased* one. Evidences from everyday life and from scientific literature poorly support this theoretical concept. In the market for news it is difficult to understand how people decide to choose their informational sources: surely they do not decide excluding biased sources. In addition, from time to time being totally biased (or even partisan) about some specific topic provide a sort of benefit for readers they want their beliefs to be confirmed. This applies above all on political topics easily explain the very large number of declared biased media, especially among the newspapers.

Nevertheless this research is not interested in how political facts and arguments are reported, but it is focused on corporate and financial reporting. In this subset we would not expect to

*PhD Student, University of Bologna, Department of Management, Via Capo di lucca 34, 40126 Bologna, Italy. Corresponding author: carlo.raimondo@unibo.it.

find any *a priori* preference for *not unbiased* news. If media are actually delivering important news, readers are expected to strictly prefer those media who report news in the better and most objective way possible. In this research paper I will show how in the daily newspaper industry the most circulated sources are systematically writing better articles for a precise set of subjects, i.e. those in conflict of interests with the media firm. I will analyze all the articles published by top five Italian daily newspapers about all listed firms in a five-year period. The main aim is to prove the relation between holding shares of a media firm and receiving a special treatment from the media firm itself.

This paper is organized as follow: in Section II a literature review will be proposed. Section III presents the hypotheses development. In Section IV I provide detail on data and methodology, with a specific focus on the textual analysis I developed. Section V presents the empirical results, whilst Section VI provides some robustness checks. Finally, Section VII concludes.

II. Related Researches

The role of information in the economy has been stressed since Von Hayek (Hayek 1945) work about knowledge diffusion and availability. Individual and organizations need to be correctly informed and, on the other hand, they need to correctly communicate their contents to the wide public. The role of information is even more important when it regards financial markets and publicly traded instruments, where informative advantage is vital for individuals and organizations.

In the information disclosure process a pivotal role is played by the media firms, as they act as information intermediaries, finding, selecting, commenting, and dealing to the vast public pieces of news. They role is furthermore even more crucial considering the credibility they are able to convey to a story telling, using their reputational capital to create around a fact or an analysis the level of attention requested to transform it in an actual news. Media are then able to (1) select a single event and transform it in a news, to (2) vehicle their opinion about the event, and consequentially to (3) importantly influence firms and organizations. The ways in which media may affect firms are several: they may influence the asset pricing (Dyck and Zingales 2003, Tetlock 2007, Barber and Odean 2008, Bhattacharya et al. 2009, Dougal et al. 2012, Garcia 2013), they may play a watchdog role against frauds (Miller 2006, Dyck, Volchkova, and Zingales 2008, Dyck, Morse, and Zingales 2010), they may generate important firm's reputation changes (Moy and Pfau 2000, Eccles, Newquist, and Schatz 2007). Media importance for the economic sector is therefore prominent especially whenever firms have to deal with the vast public. This characteristic changes following the diversity in economic sector but it is exactly the same going to

firms that issue publicly traded financial instruments. Being the latter the biggest part of several leading economies, media influence on business activities cannot be banished in consumers based industries. Given these facts, for corporate finance scholars analyzing media industry functioning mechanism and the degree of efficiency of the whole sector has become a pivotal challenge for better understand financial markets patterns as well as a complete bouquet of corporate finance hot topics, like for instance the dynamics of an IPO initial returns (Liu, Sherman, and Zhang 2014, Bhattacharya et al. 2009, Pollock and Rindova 2003).

The central point of this research paper is the challenge to correctly understand what drives the mechanisms the media follow in deciding whether a fact or an event has to be covered and how to transform the fact in a piece of news in terms of treatment and embedded judgments. This activity is probably one of the core procedures in the professional journalism together with the ability of understanding the difference between what is actually interesting for the public and what is not. This research does not aim to state if an event is effectively a news or not, but it wants to disentangle which are the determinants that affect (1) the probability to be covered for a firm and (2) the probability of receiving positive or negative comments for the firm itself. The approach I use hereafter is based on a conflict of interest view in which individuals and organizations (firms) that own a media outlet use their influence to force the media to produce news in favor of their private interests, different by the interests of the media itself. Owning a media firm, or being part of its ownership structure, is here viewed as an investment in term of reputation and communication instead of a regular financial participation. This research wants to test in which degree this strong assessment is precise, basing on empirical testing in an international sample using also advanced natal language processing techniques.

This research is very deeply rooted in both economic and financial literature. Starting form Jensen 1979 (Jensen 1979), previous researches show as media can be effective on stock markets, influencing them in a more complex way than being just information dealers. Media may generate stock markets reactions even though they are conveying actual new information through a sort of attention effect. Disentangling the media action has been widely investigated in the financial literature. Several scholars confirm the presence of the informative effect, i.e. the process in which media are able to transmit to the markets and generally speaking to the public. A large body of literature has proved how media are able to effectively influence financial markets in a wide range of diverse situations. Firstly media are able to convey to the public information that is really new and really informative. In their 2003 work Dyck and Zingales (Dyck and Zingales 2003) show how the media reporting of financial news - specifically earning announcement related news - are able to influence investors' behaviours. The authors state the media are really providing new

information to the public, especially because of their ability to access private information directly through the firms involved; the latter ones are willing to concede private information to the media and, indirectly, to the public in order to maintain a *qui pro quod* special relations with the media itself. In this view media and firms take advantage both from their special relation. Fang Peress in their 2009 paper (Fang and Peress 2009) study the cross sections of stock returns as influenced by media coverage and they the breadth of information dissemination affects returns. Dougal et al. (Dougal et al. 2012) differently find out a causal relation between a daily column content on the Wall Street Journal and the stock market average return. Joe et al. (Joe, Louis, and Robinson 2009) show how markets react to bad news of inefficient boards member in the U.S. stock market, generating substitutions in boards' compositions and pushing retail investors to shift their wealth allocations to different firms.

Media often do not generate or deal new information, i.e. information that is totally new and unknown to the markets participants, but they are able to make investors changing their preference, modifying the investor sentiment. Tetlock 2007, Tetlock 2011, Barber and Odean 2008; Solomon 2012; Grullon, Kanatas, and Weston 2004; Frieder and Subrahmanyam 2005; Meschke 2003 all prove that investors react to media exposure independently to which is the actual facts at the base of media reporting. Huberman and Regev(Huberman and Regev 2001) even show as markets react to media outcomes that are not actually conveying any new information. Media are so very powerful opinion makers also in financial markets, generating waves of positive and negative sentiment inside the investors' community, regardless that the underlying situation could be.

Media are then important and effective for markets and, consequentially, potentially harmful if they do not accomplish their role correctly, professionally, and accurately. Media have indeed a right and powerful incentive to produce correct information, as far as they are competing in a market in which customers (readers, audience) are willing to pick the highest quality outlet in term of precise and not-slanted reporting. Unfortunately competition in media industry seems not to be able to produce this desirable outcome. Mullainathan and Shleifer (Mullainathan and Shleifer 2005) show as audience wants to have its beliefs confirmed, generating in this way a sort of demand for biased reporting. George and Waldfogel (George and Waldfogel 2003) and Gentzkow and Shapiro (Gentzkow and J. M. Shapiro 2008; Gentzkow and J. M. Shapiro 2010) also argue similarly, introducing the elements of audience heterogeneity as pivotal concept for beliefs confirmation hypothesis: more audience members have very different beliefs and tastes, more difficult it would become for the media to produce the correspondent biased information useful to confirm their ideas and so to be liked. According to them a media outlet with very wide audience heterogeneity will produce better information.

If heterogeneity in audience tastes produces better informative outcomes for political topics, speaking about economic and financial reporting we should be in presence of a sort of market mechanism for which slanted reporting are actually discarded or not even proposed to the audience because in this setting "tastes" do not actually exist, with the only exception of a general and strong preference for real informative and potentially effective news. Nobody prefers *a priori* firm A to firm B, and then he decides to buy the media who reserve the better service to firm A. Regard to economic and financial news the reputation effect it should work perfectly. In this research we want to study a specific framework in which the reputational mechanism for media firms does not hold and in which media firms, and not audience, has their own tastes in matter of business reporting. We argue that the ownership structure of media firms systematically affects media outcomes in sense of reserving a special and better handling for all the firms linked to the media firm itself, namely the firms holding shares of the media company itself and all the firms linked to these last ones. In this paper's view, for a firm owning important amounts of shares of a media outlet may be a strategic asset to be used in managing the firm reputation, in the sense of increasing the likelihood to receive a good treatment in terms of positive coverage. In this respect we expect to find that firms linked to the firm owning the media firm, or linked to the media firm itself, will gain and will reach a reputational capital that is greater, *ceteris paribus*, than that would have received if not linked to media firm. The decision of holding a certain amount of a media firm is then not determined by the expectation about the economic results of the media and so to the pro quota return on the investment in financial term, but it will be decided by the expected amount of benefits the firm desires to extract using its managing rights. Furthermore has to be noticed how the private benefits of being part of the managing group of a media firm may be not just pro quota like the financial ones but entire. In other words if also some slanted news about firm A, that is part of the ownership structure of the media firm, would lead to a decrease in the total earning of the media firm, it will happen that (1) firm A will receive all the private benefits coming from a good reporting and, at the same time, (2) firm A will bear only the pro quota cost of this operation, generating a strong incentive for firm A to push the media firm in producing benevolently slanted reporting about firm A itself. We argue though that media firm ownership structure could be a critical point for the media sector, if interested in not-slanted reporting.

This will not be the first research studying ownership dynamic in the media industry, several issues have been already investigated in past researches. There are several papers investigating about the consequences of having a state-controlled ownership structure. Djankov et al. (Djankov et al. 2003) conduct a vast research on media ownership in an international context (97 countries). They found that State-ownership of media is not consistent with the idea of public interest according to

which States cure market failures and they also underline the possible threat in terms of conflicts of interest of state-controlled media. Besley and Prat (Besley and Prat 2006) improve a model explaining why organizations and, particularly, governments have strong incentives trying to control the media and consequentially the external opinion on themselves. The authors also underline other two points that are very important: firstly the media ownership structure plays a pivotal role on the media's ability to resist to external pressures and it is strictly tied with the media independence and, secondly, that the lack of independence (like the existence of some conflicts of interests) does affect the outcomes, i.e. it has effects on the news produced. The paper by Houston, Lin, and Ma (Houston, Lin, and Ma 2011) confirm the previous results providing also strong empirical evidence. Gentzkow and Shapiro (Gentzkow and J. Shapiro 2006) put in evidence the relationship between ownership structure independence and competition, where more concentrated ownership in the media industry may lead to less competition and then to undermine the quality control market function. Both the working papers by Della Vigna and Kennedy (Della Vigna and Kennedy 2011) and by Durante and Knight (Durante and Knight 2011) underline the potential pitfalls that the ownership structure of the media firms may generate, using the viewpoint of the conflicts of interest: in the first case the danger is related to the holding linkage between media firms and other firms, while in the second case a complex case of conflict of interest between business and politics is presented. Following this framework, this research wants to systematically investigate the effects of being part of the ownership structure of a media firm in term of coverage directly received by the media itself.

Several considerations could emerge from this work's results. The value of the this research is in fact mainly in the theoretical and practical implications that would arise from the paper's obtained evidence.

From the first perspective the proposed framework arises a very important point in describing the complex relation between media industry and finance. For a firm, being part of the ownership structure of a media firm could be a precise decision related also to its communication strategy and investor relation management. Being constantly positively covered by a media may turn out in decreasing the reputational risk exposure for both the firm and its executives. In other words if my predictions will be confirmed, I might observe the presence in the ownership structure of a media firm as a sort of hedging policy against negative coverage on the media itself. Furthermore, regarding to the listed company owning shares in media firms, I may observe not normal information diffusion to the vast public and consequentially systematic biases in the stock price formation mechanism, where uninformed investors who trust the media - even partially owned by the firm itself - will end to overprice the firm stock price, generating in this way an

equilibrium price for the stock that is artificially higher than in a normal situation with no conflicts of interest due to media ownership patterns.
Summarizing this research results could be important and potentially effective for both the streams of literature related to (1) the connection between financial issues and information spreading pattern and to (2) the media ownership characteristics and their effect on the quality of the news. From the second perspective the results of this research may be relevant for both the viewpoint of practitioners and policy makers and regulators. Practitioners and financial professionals could be very interested in understanding which part of the available information media owners have influenced and which part of information is instead not affected by any conflicts of interest. This pattern may also results in different financial markets' reactions when we are dealing with news about publicly listed firms. Financial literature has shown how media may affect the financial markets both when they delivering effectively new information (for instance Dyck and Zingales 2003) and when they are only putting some topic in the spotlight (for instance Tetlock 2011). Be able to understand which one of the two situations is really happening in a certain moment of time could be crucial to properly assess the correct strategy in terms of investments or disinvestment policies.

The expectations of this paper could be even more important in terms of regulation and policy decision for public regulators (e.g. CONSOB or Bankitalia) and financial markets' coordinators. If information is viewed as a public good that has to be regulated and correctly distributed to markets participants, to have a systematically biased source of news is something not desirable and that has to be regulated. If data will confirm the expectations of this research public regulators could be able to correctly detect and regulate the kind of conflict of interest between media and firm owning the media in order to preserve a correct and possibly unbiased information for all the agents operating on the markets. More generally, the results of this research would confirm the importance of a correct understanding of information generation process and its diffusion mechanism to the markets, as an increasing an innovative stream of literature in the financial economics field is proposing gaining an increasing importance. This paper's evidence would perfectly frame in this innovative literature.

III. Hypotheses Development

As previously observed the importance of media in the information dealing system is indisputable. However, only a little attention has been paid until now to the potential pitfalls that could arise together with this powerful mechanism. If media are introducing their preferences in the news

they are spreading to the public, all the system is working on a false base, i.e. on the base of an informative set that is partially adulterate by media themselves. Despite the potential huge proportion of the phenomenon, only a little attention has been devoted to it by financial and economic literature. Some scholars have instead successfully studied this issue from a political point of view. In the political market, i.e. the general elections, political parties received better treatment from "friend media". Among the others, Della Vigna and Kaplan (DellaVigna and Kaplan 2007) show as receiving better coverage from the media is actually effective in terms of electoral outcomes. Closer to this research is the 2012 paper by Gal-Or et al. (Gal-or, Geylani, and Yildirim 2012) in which the authors present a model according to which media reserve a better treatment in terms of coverage to firms that buy greater amount of advertisement. Furthermore Durante e Knight in their 2012 paper (**durante2012**) studied the effect of owning media outlet and they prove its relation with being successful in a political competition. The aim of this research would be to put together these different contributions that refer to different setting and that they are hardly generalizable in a new framework that explicitly refers to financial markets and their potentially harmful relation with the media sector. I argue the phenomenon is even more important because of the proved effective of media products on the financial markets.

More explicitly, on the basis of the above reported arguments, it is possible to suggest the following hypotheses:

For a firm, being part of the media firm's ownership structure influences the probability to receive coverage by the media itself. (H1)

For a firm, being part of the media firm's ownership structure increases the probability to receive coverage by the media itself. (H1a)

For a firm, not being part of the media firm's ownership structure decreases the probability to receive coverage by the media itself. (H1b)

And then going to the coverage's quality:

For a firm, being part of the media firm's ownership structure influences the probability to receive positive coverage by the media itself. (H2)

For a firm, being part of the media firm's ownership structure increases the probability to receive positive coverage by the media itself. (H2a)

For a firm, not being part of the media firm's ownership structure decreases the probability to receive positive coverage by the media itself. (H2b)

That leads to:

For a firm, being part of the media firm's ownership structure influences the probability to receive negative coverage by the media itself. (H3)

For a firm, being part of the media firm's ownership structure decreases the probability to receive negative coverage by the media itself. (H3a)

For a firm, not being part of the media firm's ownership structure increases the probability to receive negative coverage by the media itself. (H3b)

And to:

For a firm, being part of the media firm's ownership structure influences the probability to receive coverage representing uncertainty by the media itself. (H4)

For a firm, being part of the media firm's ownership structure decreases the probability to receive coverage representing uncertainty by the media itself. (H4a)

For a firm, not being part of the media firm's ownership structure increases the probability to receive coverage representing uncertainty by the media itself. (H4b)

The confirmation of the research hypotheses (i.e. the existence of a systematic bias in the news generation process in favor of firms involved in the ownership structure of the media firm) would confirm the importance of the role played in the information industry by the existence of severe conflicts of interest between media industry and the other ones and would enrich financial

literature of a fundamental contribution able to help the understanding of a debated phenomenon like the relation between media and finance is. At the same time this research's results would help to explain the difference in media treatment that firms received.

IV. Data and Methodology

One of the main aspirations of this research paper is to provide a new methodology to analyze textual data, in order to obtain a new dataset basing on previously not usable and never used relevant textual data. Be able to objectively e systematically analyze daily newspapers' articles in order to study the relationship between ownership patterns and news outcomes is a crucial point of this work I am going to describe in details data and methodology in this section. From this viewpoint, this paper has the ambition to be the first research paper analyzing in a deep and extensive way Italian newspapers' articles referred to the business and financial sector, with a proper procedure I realized *ad hoc* for this kind of contents.

I. Data Collection

Articles, published on daily newspapers, represent the main unit of analysis of this research paper. I restrict the number of articles analyzed following the research design in several ways. First I chose the most important Italian daily newspapers by diffusion, excluding the sportive ones and those with *unconventional* pattern in the ownership structure of the correspondent publishing firm[1]. Data about diffusion comes from an independent agency called *Accertamento e Diffusione Stampa*, a consortium formed by advertising buyers and publishing firms with the mission of monitoring the diffusion of media contents. I also restrict the sample excluding those newspapers not indexed on the main news collection databases (i.e. Lexis-Nexis and Factiva). This process ends with the choice of five high-diffused Italian daily newspapers, namely *Il Corriere della Sera*, *La Repubblica*, *La Stampa*, *Il Giornale*, and *Il Resto del Carlino*. For those newspapers I search and download all the news referred to all Italian listed firms together with a number of unlisted firms that have a role in the Italian newspaper industry[2]. Furthermore, data are referring to a five-year period, from January 2007 to December 2011, allowing to consider time varying fixed effects[3]. Following this

[1]For this reason I do not consider for instance *Il Sole 24 Ore*, since it is owned by an association (i.e. *Confindustria*), or *Libero Quotidiano*, owned by a foundation. This characteristic of not being juridically a company makes impossible to correctly determine ownership structure details.

[2]The sample of news related to unlisted firms refers to the news about all unlisted firms that own any share in at least one of the considered newspapers. For these firms I search for news on all the five newspapers.

[3]Having detailed data about time distribution of news allows me to take in account for all the processes that are time specific, e.g. economic cycle pattern or seasonal phenomena.

research strategy I obtained a sample of 127,190 articles, all the articles about 263 Italian firms[4] published on the five above-cited daily newspapers, in the period spanning from January 2007 to December 2011. These data are then combined with media publishing firms' ownership data, obtained from some Bureau Van Dijk databases, namely Amadeus, Orbis, and Aida datasets.

II. Textual Analysis

A key contribution of this research paper is about the methodology used to process a huge amount of textual data in Italian language, proving a first item that can be used specifically for financial-related text in Italian language. Textual analysis in the financial context has been deeply changed and improved by the 2011 paper by Loughran McDonanld (Loughran and McDonald 2011), where the authors proposed a financial specific set of word lists. Basing on that research, the stream of literature using textual data in financial sector has improved dramatically (Tetlock 2007; Dougal et al. 2012; Loughran and McDonald 2013; Garcia 2013). Curiously, before this paper, a similar set of word lists was not developed for the Italian language. Basing on Loughran McDonald lists for English language, I built a new set of lists for Italian language. The translation process followed few but strict rules: provide the best translation possible to the respect of financial and business sector usage of each single word, be fully comprehensive allowing also a not-biunivocal linkage between English and Italian words, not include words with an ambiguous sense in the Italian language. Furthermore I replicate the logic used in Loughran-Mcdonald lists, providing a list of terms and not a list of lemmas. In order to obtain this property I had to expand the word lists to consider all the inflected forms. In order to do that correctly and properly I used a linguistic tool called *Morph-It* (Zanchetta and Baroni 2005) that maps all Italian terms and their inflections. The result is a new tool that allows ma to analyze the contents of Italian financial texts and retrieve the measures of positive, negative, uncertainty, legal, modal weak, and modal strong tones. The tones are scaled to represent the proportion of *tonal* words on the total number of words in each text, avoiding in this way any possible scale generated biases. For positive and negative tones I also take in account for possible negations of positive and negative words. Since a negated positive (or negative) word, like *not excellent*, has a different meaning than the original ones (in this case *excellent*), but it is difficult to understand properly the real new meaning, I do not include these cases in any lists.

[4]The firms' sample is formed by the union of all Italian listed firms and all unlisted firms owning a share in the considered newspaper publishing firms.

III. Research Design

A crucial point of this research paper is the aim of building strong and consistent analysis basing on a limited but comprehensive set of data. All the research design and the data description are conditioned by the decision of maintaining a simple and easy to understand research design that is, at the same time, able to catch an incredible high number of not-observed and even not-imagined data potentially important in describing the observed phenomenon. As mentioned before, the main unit of analysis of this paper is at the article level. Each article is furthermore associated with a publishing firm (I call it the *Newspaper*), with a firm that is its main topic (I call it simply *Firm*), and with a publishing date (called *Date* and often aggregated to the year level variable *Year* or to the quarter level variable *QDate*). Furthermore a crucial set of variable is those ones linking the publishing firm with the article-topic firm, i.e. the variables representing the ownership patterns in the media industry. The most important and hereafter presented are two: a dummy variable called *Confl* and a continuous variable called *Own*. The dummy variable *Confl* assumes value equal to one when there are some ownership linkage between publishing firm and article-topic firm, i.e. when there is a *Conflict of Interest* between newspaper's owners and newspaper's duty in providing objective news. The continuous variable *Own* represents the size of the conflict of interest, as measured as proportion of shares of the media firm owned by each article-topic firm. The first critique that could be raised is how I can control for other factors that can influence the media outcome for a firm, like for instance advertising expenditures, or investor relation management, or the way in which the single economic sector of the firm is perceived. Furthermore also each single newspaper may have different characteristics in terms of propensity to be friendly or hostile to corporations. The number of this kind of pitfalls could be very large and including new variable to respond in detail to each of that seems not to be very effective. For this reason I decide for a totally different approach that takes advantage of the specific data structure. In the models I am going to estimate I am considering both firm and newspaper fixed effects. The great dimension and the variability present in my data allow me to consider as generated by ownership patterns only that part of variance not explained by the single firm effect or the single newspaper effect[5].

Briefly, this is the framework I am using in the vast part of my empirical analysis. The unit of analysis is at the article level. Each article is an effectively published article having as attributes several characteristics: date of publication, article's publishing newspaper, firm the article refers to, total number of words in the article, and the complete set of *sentiment* indicators (i.e. *Positive, Negative, Uncertainty, Legal, Modal Weak, Modal Strong* tones). Furthermore I merge these data with

[5] Actually also quarter and year fixed effects are taken in account in almost all the proposed models.

the ownership data referred to newspapers' publishing firms. For each year and for each couple newspaper-firm[6] data about the ownership linkage between the newspaper and the firm is used. I use two different measures for the ownership linkage: a simple dummy variable that assumes value equal to one if there are any forms of ownership linkage, and a continuous variable varying from zero to one representing the share of the newspaper publishing firm owned by the firm the article is about. In order to correctly analyze the effects of ownership features on the articles' sentiment as measured by the textual analysis, a binding condition is to be able to disentangle the ownership effect from all the other firms' and newspapers' characteristics that may affect articles' tone. Instead of going collecting a huge amount of data describing all these features, I decide to address systematically this issue exploiting the research design in a not-intuitive way. The following framework is proposed:

$$S_{n,f,t,i} = \beta_0 + \beta_1 f_i + \beta_2 n_i + \beta_3 t_i + \beta_4 f_i n_i + \epsilon_i$$

Where $S_{n,f,t,i}$ is the sentiment of the i article published on the n newspaper, about the f firm, in the t date; β_1 is the effect of each single firm; β_2 is the effect of each single newspaper; β_3 is the effect of each single time period (I use both quarters and year specifications); β_4 is the effect of each each couple *firm-newspaper* not explained by *firm* and *newspaper* fixed effects, i.e. is measuring the effect of the relation between each firm and each newspaper. I am going to use this measure together with the above described measures of ownership, i.e. the *conflict of interest* proxies. In this way I am linking the results in terms of *conflict of interest* effects only to what is not actually described by any characteristics that is either firm specific or newspaper specific. All data about firm size, profitability, industry, governance, glamour degree, that likely could influence the media coverage for a firm are therefore implicitly already taken in account. Following the same way of reasoning all newspaper specific characteristic are already considered too. The strengths of this approach are twofold: first (1) this measure is extremely precise, leaving no space for measurement errors, and then (2) this measure is complete and it prevents from omitted variable biases. In this way I am considering all the variables that vary together with firm or newspaper, even those I am not thinking about, or those for which data and empirical measures are not available, or they do not exist at all. Furthermore, as far as quarter fixed effects are also considered, I am also taking in account for possible variations of the above described effects happening along time.

[6]Hereafter I will simply call *firm* the firm the article is about, and *newspaper* the newspaper that published the article.

IV. Data Summary Statistics

In order to properly analyze the effect of being part of the ownership structure of a media firm, I collect a vast set data, consisting in a huge number of articles published by the daily newspapers I included in the sample. The articles are indeed all the articles about listed Italian firms, both in conflict and not in conflict with the publishing firms, and a subset of unlisted firms formed by all those in conflict with at least one publishing firm. Both for listed and unlisted firms articles are searched between January,1st, 2007 and December, 31st, 2011 (exactly 5 year period). The final output is formed by 127,190 articles referred to 263 firms. In order to clarify as possible the sample composition I am reporting in Table 1 the coverage received by the most covered firms, namely by all the firms that receive at least one hundred articles in at least one newspaper.

INSERT TABLE 1 HERE

Table 1 reports for each firm the total number of articles published in the chosen time window by the five considered newspapers, separately. Looking carefully at these data it is possible to start reasoning about coverage patterns. For instance taking a couple of random examples we may observe how *Cofide*, a firm in the same holding group of *La Repubblica* publishing firm, received the double coverage on the *Repubblica* newspaper compared to the *Corriere della Sera* newspaper. All the coverage details are reported disentangling between the different sources and it also help to understand how the sample is formed.

Also the time patterns of coverage distribution across periods could be interesting. Coverage patterns can be both depending from a time trend evolution or showing some seasonal effects. In the above reported *Figure 1* I present data about total coverage distinguishing by quarter.

INSERT FIGURE 1 HERE

From *Figure 1* it is possible to understand how the total coverage related to the whole sample. The general trend is slightly decreasing, passing from more than eight thousands articles in the first quarter of 2007 to approximately six thousands articles in the last quarter of 2011. Also a tiny seasonal effect could be noticed, since all the first quarter of each year are associated with a

greater number of articles than the other quarters of the same year.

Coverage measures are just a part of this paper data. The second and more important part is that related to the tone indicators. I compute the *tone* variables for all the articles and in the *Figure 2* the average value for all the *tones* are reported by year.

INSERT FIGURE 2 HERE

In the above reported histogram I reported the evolution of all the *tones* by year, in this order: *Positive* tone, *Negative* tone, *Uncertainty* tone, *Legal* tone, *Modal Weak* tone, and *Modal Strong* tone. The tone are expressed as proportion of *tonal* words on the total number of words contained in each article. Consistently with previous studies using similar methodology, the negative tone is on average dimensional greater than the other ones. The average level of *Negative* tone is above 2% of the total number of words, whilst the *Positive* tone is on average close to 1.5%. *Uncertainty* tone and *Legal* tone are instead just above the 1%. Finally both *Modal Weak* and *Modal Strong* are on average smaller than 0.5%. Furthermore, it is also possible to observe how *tone* indicators are not very volatile over time, proving the design validity of the sample selection.

Closer to the focus of this research paper is to observe patterns in coverage and sentiment distribution not over time but following the existence of a *Conflict of Interest*, i.e. following the presence and the magnitude of the ownership linkage between what I define *Newspaper* and what I defined *Firm*[7]. First attempt to present these kinds of data is in the *Table 2*, where data about coverage are presented following the existence of a *Conflict of Interest* and providing details at the year level and at the newspaper level.

INSERT TABLE 2 HERE

I find the above reported table very informative about the size of the central issue of this article in the Italian economy. *Conflict of Interest* could be a very effective and important theoretical framework but the effectiveness and interest of this kind of approach to the financial and business media reporting is strictly related to the size of the *phenomenon*. *Table 2* indisputably shows the dimensional size of the *Conflict of Interest* in the Italian newspapers' sector and how common

[7] emphFirm is the firm the article is about, and *Newspaper* is the newspaper that published the article.

perception of high reputed and *Conflict of Interest* free newspaper could be argued. On the total of news in our sample (i.e. a subset containing all the news related to Italian listed firms with few additions) almost the 16% are suffering a *Conflict of Interest*. Going to each single newspapers, the *Corriere della Sera* published the greater number of news in *Conflict of Interest*: the 39.37%. *La Stampa* follows with 22.67% news related to firm suffering *Conflict of Interest*, then *Il Giornale* (13.90%), *La Repubblica* (4.22%), and *Il Resto del Carlino* (3.21%).

Being classified as suffering *Conflict of Interest* is not the best specification I am providing for measuring the *Conflict of Interest*. Data reporting precise proportion of publishing firm share are computed and used in this paper's empirical analysis. It this then interesting to understand which is the distribution of articles the suffer a *Conflict of Interest*, according to the size of the *Conflict of Interest* itself. *Figure 3* represents exactly this distribution [8].

INSERT FIGURE 3 HERE

From the above reported histogram, some patterns could be detected. First, there is a more than moderate concentration on the left part of the graph, representing firm taking a relevant share in the newspapers' publishing firms without having the majority of the shares (from 5% to 29%). Then relatively few firms hold a share that is able to maintain the total control without totally owing the media firm. Finally a discreet number of firm owns totally the media firms. It is represented by the very right column: more than the 3% of the total number of articles is referred to firms owing the 100% of the media itself[9].

Having described the summary statistics for the observed phenomena at the best of my data allow me, the next step is to present and describe briefly the main variables I am going to use in the empirical models. As mentioned before, one aim of this paper is to be as simple as possible in data analysis and also this part reflects this research design choice. Variables presented in the next table are indeed the core engine of all this research paper.

INSERT TABLE 3 HERE

[8](I am here taking in account only for firm suffering *Conflict of Interest*, firm they do not would be represented by a big column on the 0 value making all the graph more difficult to be understood.)

[9]This data is so high also I considered in the sample of news also the news referred to the publishing firm itself that, by definition, owns the 100% of itself.

Table 3 presents the main variables used in this work in three panels. Panel A presents all the variables related to the textual analysis part. It is possible to observe that in mean (median) each article is formed by 344 (265) words, among those the 1.7% (1.4%) are positive, the 2.2% (2.0%) are negative, the 1.1% (1.0%) are related to uncertainty, the 1.2% (1.0%) are legal or litigious, the 0.4% (0.3%) are modal weak, and the 0.5% (0.4%) are modal strong.

Figure 4 closes the summary statistics section of this paper. In this graph I tried to provide an intuitive representation of the *Conflict of Interest* characteristics, both in terms of associated coverage (the upper part of the graph and the correspondent right y axe metrics) and associated sentiments (the below part of the graph and the correspondent left y axe metrics).

INSERT FIGURE 4 HERE

This graph shows the differences in terms of coverage's quantity and quality between firms suffering *Conflict of Interest* and firms not suffering *Conflict of Interest*. The dotted line in the upper part is representing the share of articles suffering for *Conflict of Interest*, and its variation over time.It reaches the maximum level in 2007, decreasing by almost 2 percentage point in 2008, and then it stays stable on a medium level until 2011. The part below is instead representing the differences in each tone average between in conflict firm and not in conflict firms[10]. First column of each histogram is the *Positive* tone. From the graph it is difficult to see if there are any differences, since all values seem very close to zero (i.e. the average value are the same for in conflict and not in conflict firms). More clear is the representation of the *Negative* tone. Here the columns are consistently for all the five years significantly negative, showing as on average for a firm being in conflict is associated with a lower level of *Negative* tone. Similar is the situation for the extreme right column, that associated with the level of *Legal and Litigious* tone, where firms suffering a *Conflict of Interest* are associated on average with a lower value. More difficult to interpret is the evolution for the *Uncertainty* tone, that seems to be always positive but significantly different than zero only in the more recent years.

[10] A positive value of the histogram means the average is greater for in conflict firms than not in conflict ones.

V. Correlation Tables

Before going to the proper results section, it would be interesting to analyze how the main variables used in this empirical analysis are correlated each others. I am going to present hereafter three tables that would be able to clearly identify interesting patterns and harmful pitfalls contained in the dataset I created. This is particularly important since many of the measures I am using in this empirical analysis are actually brand new and never computed before with this strategy, especially taking in account the fact I am dealing with an almost not explored yet field of research like the financial related text in Italian language.

INSERT TABLE 4 HERE

The first table presented is about the sentiment indicators related variables. In this case I would expect a very strong correlation between variables, as in the English language similar methodology. Looking at the table it is possible to observe that the size of correlation is not as big as expected. I obtained levels similar to those usually found in english text just for the correlation between *Uncertainty* tone and *Modal Weak* tone. Still, since all correlation are always greater than 0.1 and following previous studies, I decided not to use together this variables in order to be the most precise possible.

The next two table are instead presenting the correlations between the main sentiment indicators (*Positive, Negative, Uncertainty, and Legal*) and the five different newspapers we considered (*Corriere della Sera, La Repubblica, La Stampa, Il Giornale, Il Resto del Carlino*) and the correlations between the main sentiment indicators (*Positive, Negative, Uncertainty, and Legal*) and the five different years in the sample (*2007, 2008, 2009, 2010, 2011*).

INSERT TABLE 5 HERE

INSERT TABLE 6 HERE

V. Results

I. Conflict of interest and the quantity of the news

The main results of this paper is to prove two significant and positive relationship: (1) between being a firm in *Conflict of Interest* and the quantity of coverage and (2) between being a firm in *Conflict of Interest* and the quality of coverage. The direction of both the relationships is positive: being in conflict produces a dimensional greater and a qualitatively better coverage. To prove this relationship and the existence of a causal effect I develop a wide set of empirical analysis and I am reporting here the most significant ones. First aim of this research was to investigate about the effect of a coverage effect. In order to do that I collapsed all the observations from the articles level to the firm-newspaper couple unit of analysis, counting how many articles and how many words have been used in total for each couple (e.g. a couple is *Fiat-La Stampa* or *Fiat-La Repubblica*). Results are reported in the table below. INSERT TABLE 7 HERE

The four regressions presented on the *Table 5* coherently show a positive effect of being in *Conflict of Interest* on the two measures of coverage I developed. This table is effective to respond to research hypotheses *H1*, *H1a*, and *H1b*. All of the three found positive answers from the data analysis, with both the specification of coverage and ownership measures I computed. Regression (1) shows how the number of articles is positively and significantly influenced by *in conflict*. This holds also controlling for newspaper fixed effect. On average being part of the ownership structure of a publishing firm produces a coverage significantly greater that not being. The same holds even if we consider as a measure of *Conflict of Interest* not only the simple dummy, but the ownership share (in the table is indicated with *Media Firm Share Owned*). Furthermore and not surprising the analysis performed using as coverage variable the total number of words instead of total number of articles (a more precise measure indeed) returns the same results, even more robust. These regressions are supporting in a very strong way all the hypotheses I made on the relationship between coverage and the *Conflict of Interest*.

II. Conflict of interest and the quality of the news

Probably the main point of this research is instead to look for a qualitative effect on the articles published by newspapers regarding firms taking part in the ownership structure of the newspaper itself - i.e. the firm suffering the *Conflict of Interest*. First attempt in the direction of answering this question is provided in *Table 8*, with the simplest way possible: a t-test on the mean differences.

INSERT TABLE 8 HERE

Table 8 reports the t-test between the two sub-sample: firms in conflict and firms not in conflict. This test proves the differences between the mean of the *Negative* tone in the two subgroups is statistical significant. Since the mean of the in conflict sub-sample is lower than the mean of the not in conflict sub-sample, it is possible to argue that firms in conflict are associated significantly with a lower level of *Negative* tone, i.e. with a qualitatively better coverage. I chose to present just the result for the *Negative* tone in t-test form since the negative tone is the most common and usually reliable sentiment indicator. Nevertheless to be able to argue about hypotheses confirmation is better to look before at the complete model presented hereafter.

Next table is the first representation of the complete model discussed before. Firm specific and newspaper specific fixed effects are taken in account, together with quarters fixed effects[11]

INSERT TABLE 9 HERE

Results of *Table 9* are the very core part of this research paper. In this table consistent, significant, highly supported results are reported. Basing on a sample size of 127,190 observations, I find confirmations for two out of three of the main research hypotheses and for the last one I have no significant effect. The regressions, based on ordinary least squared estimators, show (1) a negative and significant effect of being part of the ownership structure of the media firm on the *Negative* tone (it supports hypotheses H3,H3a, H3b), (2) negative and significant effect of being part of

[11] In these regression models, as well as in all the others will be presented in the paper, only the quarter fixed effects models are shown. Furthermore I obtained the same results using models with year fixed effects, models that I am omitting for brevity.

the ownership structure of the media firm on the *uncertainty* tone, (3) negative and significant effect of being part of the ownership structure of the media firm on the *Legal-Litigious* tone, and (4) negative and significant effect of being part of the ownership structure of the media firm on the *Modal Weak* tone (the last three support hypotheses H4,H4a, H4b). At the contrary there is no evidence supporting the hypotheses H2 (*Positive* tone has not significant impact) or H1 (there is no evidence about coverage's patterns, coverage here only measured as average number of words per article).

Same results but even more significant are those reported in *Table 10*, in which the same regression models are reported but with the continuous specification for the *Conflict of Interest*.

INSERT TABLE 10 HERE

Table 10 supports and strengthens the results of the previous table, when the *Conflict of Interest* is computed as share of the publishing firm owned by the firm each article is about.The regressions, based on ordinary least squared estimators, show (1) a negative and significant effect of being part of the ownership structure of the media firm on the *Negative* tone (it supports hypotheses H3,H3a, H3b), (2) negative and significant effect of being part of the ownership structure of the media firm on the *uncertainty* tone, (3) negative and significant effect of being part of the ownership structure of the media firm on the *Legal-Litigious* tone, and (4) negative and significant effect of being part of the ownership structure of the media firm on the *Modal Weak* tone (the last three support hypotheses H4,H4a, H4b). At the contrary there is no evidence supporting the hypotheses H2 (*Positive* tone has again not significant impact) or H1 (there is no evidence about coverage's patterns[12]).

Something is important to consider in this research framework is the role of the little sample of unlisted firms. Since it was impossible to consider all unlisted firms, I decide to consider all those involved in any ways in at least one of the ownership structure of the publishing firm. What I strongly want to avoid is to generate any sort of biases based on the fact these firms are not listed. Furthermore is pretty interesting to understand is information published on daily newspaper has a different value for listed and unlisted firms. First attempt to answer these question is in *Table 11* and *Table 12*.

[12]Coverage is also here and in all the next regression models intended as number of words per article and not total number of article or total number of words. The best specification for coverage in that provided in *Table 7*

INSERT TABLE 11 HERE

INSERT TABLE 12 HERE

Also after controlling for the presence of unlisted firms, all the results above described for *Table 9* and *Table 10* hold. Magnitudes, sign, and significances of each effect remain the same. We find little support to the effect of being unlisted on the quality of coverage. The little evidence that may be obtained by the second table seem to suggest the unlisted effect as an opposite effect to the *Conflict of Interest* one. A deeper analysis on the combined effect between being unlisted and being in *Conflict of Interest* will be proposed in the *Robustness Checks* section, specifically in *Table 17*.

VI. Robustness Checks

I. TOBIT Specifications

The first check needed to be applied to these regression models derives from pure statistical reasons. For all the *Tone* indicators we use variables that may vary from 0 to 1 by construction, since they are basically the ratio of total number of *tonal* words on the total number of words in each articles. In order to avoid mistakes or inconsistencies a TOBIT specification of the above presented model is hereafter provided. The next two tables (*Table 13* and *Table 14*) show the results for a TOBIT-based specification of the main models, using both the *Conflict of Interest* dummy and the *Conflict of Interest* continuous variable.

INSERT TABLE 13 HERE

INSERT TABLE 14 HERE

The different statistical specification does not change anything in my models. Even in *Table 13* and *Table 14* it is possible to appreciate significant and negative effects of the *Conflict of Interest* variable on the *Negative* tone, as well as negative is the effect on all the uncertainty-related variables (*Uncertainty, Legal-Litigious, and Modal Weak*). Also these models furthermore are supporting the hypotheses H3, H3a, H3b, H4, H4a, and H4b and they are not able neither to support or to deny hypotheses H2, H2a, and H2b, perfectly consistent with the previous OLS based models.

II. Interaction Variables

In order to try to deeper analyze the ownership structure effects some models with interaction variables are hereafter proposed. The first of them is that presented in *Table 15* in which I am going to analyze the combined effect of being in *Conflict of Interest* with each single newspaper, in order to understand if being in conflict is more effective on a newspaper than on another.

INSERT TABLE 15 HERE

Results from this table are not very easy to be understood. From a side and for the first time the coverage effect is confirmed at the article level, providing additional evidence in support of hypotheses H1, H1a, and H1b. From the other side, instead, reading this table is not very straightforward. All the *tonal* effects seem to loose significance, leaving it to the interactions between themselves and the newspapers' dummies. Hypotheses H2, H2a, H2b, H3, H3a, H3b, H4, H4a, H4b are neither supported or denied by these models.

The second set of interaction variables based models considers the interaction between the *Conflict of Interest* dummy with each year of the sample.

INSERT TABLE 16 HERE

Results of these models are confirming the effect of the *Negative* tone but they are not confirming results related to the *uncertainty-related* tones. Here we may have some evidence in support of hypotheses H2, H2a, and H2b and also to the coverage at the article level effect (in support of hypotheses H1, H1a, and H1b). Not easy is to understand and to explaining the time and conflict combined effect. The most informative regressions seem to be the number (3) having the *Negative* and the the number (4) having the *Uncertainty* tone as dependent variable. Here the effect of time on the *Conflict of Interest* seems to be in the direction of increasing the ownership effects in the first years of the considered sample, especially in 2007 and 2008.

Last but maybe the most interesting of the regression taking in account for interactions is that reported in *Table 17*, showing the results for the combined effects of being in *Conflict of Interest* and being an unlisted firm.

INSERT TABLE 17 HERE

Using this specification, all the model described before are still working. We have evidences in support of H3, H3a, H3b, H4, H4a, and H4b, those regarding the *Negative* and the *Uncertainty-related* tones.More interesting is the effect of the interaction variable named *Unlisted*Confl*, especially in the uncertainty-related models. There the interaction variable assumes significance with the same sign as the main variable, leading to a strenght in the basic effect.

III. Propensity Score Matching

The last specification for our data analysis is totally different than the previous one. Here we are considering a *propensity score* between treated (i.e. article in *Conflict of Interest*) and untreated (articles not in *Conflict of Interest*) observations. I compute a propensity score basing on all the variables I have in the dataset and then I am going to pick for each observation in the *treated* sample the nearest possible neighbor in the untreated sample. In this way I formed a control sample based only on statistical property that could be a very powerful check for the entire research design of this paper. The propensity score matching is computed both taking in account and not taking in account for unlisted characteristic. In *Table 18* results are reported.

INSERT TABLE 18 HERE

Table 18, especially in the Spcification 2, is the result closest to confirm all the research hypotheses of this paper. For the only time I also obtained a little but significant evidence in support of H2, i.e. a positive effect of *Conflict of Interest* on *Positive* tone. More importantly, all the previous results are confirmed in both the specifications. These models are indeed supporting the hypotheses H3, H3a, H3b, H4, H4a, and H4b perfectly consistent with the previous OLS and TOBIT models.

VII. Conclusions

Firms owning shares of media publishing firms receive a ”better treatment”. They receive more coverage, especially when measured in terms of total number of articles per year or in the of total space used for them (represented by the total number of words). They receive a better coverage especially in terms of quality of coverage. For a firm in *Conflict of Interest* with the publishing firm a fewer number of negative word is used - even when measure in percentage of total number of word of each article - and also a fewer number of uncertainty related words is used. This paper's data and empirical analyses are strongly supporting the *Conflict of Interest* theory, for which a firm owning a newspaper tries to maximize private benefits through obtaining more coverage than comparable firms and, most importantly, better and qualitatively greater contents of the articles.

References

Barber, Brad M and Terrance Odean (2008). "All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors". In: *Review of Financial Studies* 21.2, pp. 785–818.

Besley, Timothy and Andrea Prat (2006). "Handcuffs for the Grabbing Hand? Media Capture and Governament Accountability". In: *American Economic Review* 96.03, pp. 720–736.

Bhattacharya, Utpal et al. (2009). "The role of the media in the internet IPO bubble". In: *Journal of Financial and Quantitative Analysis* 44.03, pp. 657–682.

Della Vigna, Stefano and Alec Kennedy (2011). "Does Media Concentration Lead to Biased Coverage? Evidence from Movie Reviews".

DellaVigna, Stefano and Ethan Daniel Kaplan (2007). "The Fox News Effect: Media Bias and Voting". In: *The Quarterly Journal of Economics* 122.3, pp. 1187–1234.

Djankov, Simeon et al. (2003). "Who owns the media?" In: *Journal of Law and Economics* 46.

Dougal, Casey et al. (2012). "Journalists and the stock market". In: *Review of Financial Studies* 25.3, pp. 639–679.

Durante, Ruben and Brian Knight (2011). "Partisan Control, Media Bias, and Viewer Responses: Evidence from Berlusconi's Italy".

Dyck, Alexander, Adair Morse, and Luigi Zingales (2010). "Who Blows the Whistle on Corporate Fraud?" In: *The Journal of Finance* 65.6, pp. 2213–2253.

Dyck, Alexander, Natalya Volchkova, and Luigi Zingales (2008). "The Corporate Governance Role of the Media : Evidence from Russia". In: *The Journal of Finance* LXIII.3.

Dyck, Alexander and Luigi Zingales (2003). "The Media and Asset Prices". In:

Eccles, Robert G, Scott C Newquist, and Roland Schatz (2007). "Reputation and its risks." In: *Harvard Business Review* 85.2, pp. 104–14.

Fang, Lily and Joel Peress (2009). "Media Coverage and the Cross-section of Stock Returns". In: *The Journal of Finance* 64.5, pp. 2023–2052.

Frieder, Laura and Avanidhar Subrahmanyam (2005). "Brand perceptions and the market for common stock". In: *Journal of financial and Quantitative Analysis* 40.01, pp. 57–85.

Gal-or, Esther, Tansev Geylani, and Tuba Pinar Yildirim (2012). "The Impact of Advertising on Media Bias". In: *Journal of Marketing Research* XLIX.February, pp. 92–99.

Garcia, Diego (2013). "Sentiment during recessions". In: *The Journal of Finance* 68.3, pp. 1267–1300.

Gentzkow, Matthew and Jesse Shapiro (2006). "Media bias and reputation". In: *Journal of Political Economy* 114.2.

Gentzkow, Matthew and Jesse M Shapiro (2008). "Competition and Truth in the Market for News". In: *Journal of Economic Perspectives* 22.2, pp. 133–154.

— (2010). "What Drives Media Slant? Evidence From U.S. Daily Newspapers". In: *Econometrica* 78.1, pp. 35–71.

George, Lisa and Joel Waldfogel (2003). "Who Affects Whom in Daily Newspaper Markets ?" In: *Journal of Political Economy* 111.4, pp. 765–784.

Grullon, Gustavo, George Kanatas, and James P Weston (2004). "Advertising, breadth of ownership, and liquidity". In: *Review of Financial Studies* 17.2, pp. 439–461.

Hayek, Friedrich August (1945). "The use of knowledge in society". In: *The American economic review*, pp. 519–530.

Houston, Joel F, Chen Lin, and Yue Ma (2011). "Media ownership , concentration and corruption in bank lending $". In: *Journal of Financial Economics* 100.2, pp. 326–350.

Huberman, Gur and Tomer Regev (2001). "Contagious speculation and a cure for cancer: A nonevent that made stock prices soar". In: *The Journal of Finance* 56.1, pp. 387–396.

Jensen, Michael C (1979). "Toward a Theory of the Press". In: *Economics Social Institutions*. Springer, pp. 267–287.

Joe, Jennifer, Henock Louis, and Dahlia Robinson (2009). "Managers and Investors Responses to Media Exposure of Board Ineffactiveness". In: *Journal of Financial and Quantitative Analysis* 44.03, pp. 223–260.

Liu, Laura Xiaolei, Ann E Sherman, and Yong Zhang (2014). "An attention model of IPO underpricing, with evidence on media coverage". In: *unpublished*.

Loughran, Tim and Bill McDonald (2011). "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks". In: *The Journal of Finance* 66.1, pp. 35–65.

— (2013). "IPO first-day returns, offer price revisions, volatility, and form S-1 language". In: *Journal of Financial Economics* 109.2, pp. 307–326.

Meschke, Felix (2003). "CEO interviews on CNBC". In:

Miller, Gregory S (2006). "The press as a watchdog for accounting fraud". In: *Journal of Accounting Research* 44.5, pp. 1001–1033.

Moy, Patricia and Michael Pfau (2000). *With malice toward all?: The media and public confidence in democratic institutions*. Greenwood Publishing Group.

Mullainathan, Sendhil and Andrei Shleifer (2005). "The market for news". In: *American Economic Review*, pp. 1031–1053.

Pollock, Timothy G and Violina P Rindova (2003). "Media legitimation effects in the market for initial public offerings". In: *Academy of Management Journal* 46.5, pp. 631–642.

Solomon, D (2012). "Selective publicity and stock prices". In: *The Journal of Finance* LXVII.2.

Tetlock, Paul C (2007). "Giving content to investor sentiment: The role of media in the stock market". In: *The Journal of Finance* 62.3, pp. 1139–1168.

— (2011). "All the news that's fit to reprint: Do investors react to stale information?" In: *Review of Financial Studies* 24.5, pp. 1481–1512.

Zanchetta, Eros and Marco Baroni (2005). "Morph-it! A free corpus-based morphological resource for the Italian language". In: *Corpus Linguistics* 1.1, pp. 1747–9398.

VIII. Tables and Figures

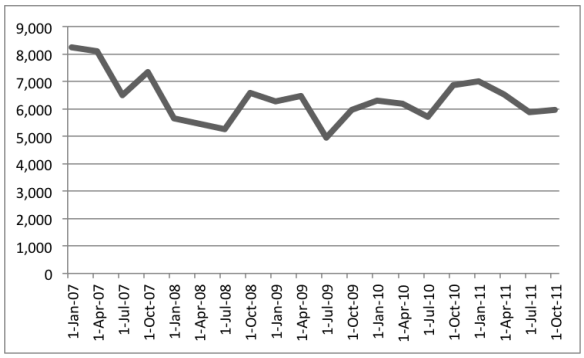

Figure 1: *Coverage per period. In this graph the total number of articles published on the five considered newspapers about the sample firms is reported. Articles are aggregated by quarter and they refer to the totality of listed Italian firms together with the unlisted firms involved in at least one of the ownership structures of the publishing firms.*

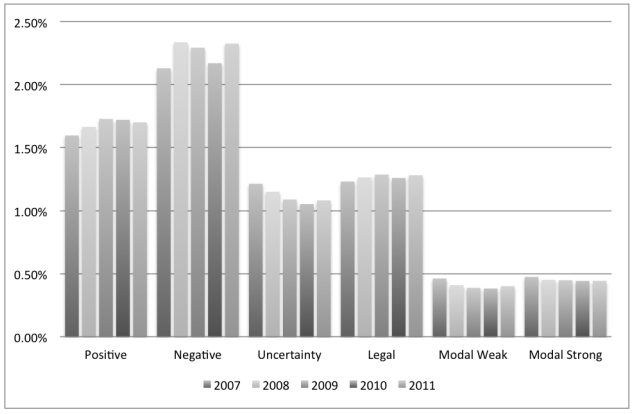

Figure 2: *Media sentiments per period. In this graph the mean of each tone is reported by year. Each tone is the result of the proportion of* tonal words *on the total number of words for each article.From the figure we may observe how the average level of positiveness, negativeness, uncertainty, et cetera vary across time.*

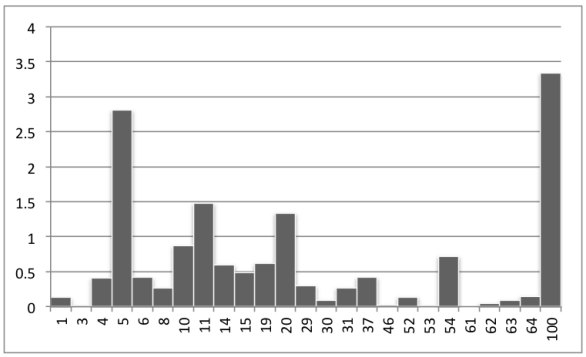

Figure 3: *Articles distribution for firms in conflict following the ownership share they own in publishing firm. On the* Y *axe number of articles are reported as percentage of the total number of article in the sample I considered.*

Table 1: *Sample composition. Number of articles published by each newspaper for all the firm (both in conflict and not in conflict firms). Showing only the firms that receive at least 100 articles on at least one newspaper.*

FIRM	Newspaper_name					Total
	Corriere della Sera	La Repubblica	La Stampa	Il Giornale	Il Resto del Carlino	
A2A SPA	743	593	288	696	633	2,953
ACEA SPA	181	278	49	175	71	754
AEFFE SPA	26	56	14	22	111	229
ANSALDO STS SPA	90	347	89	205	128	859
ARKIMEDICA SPA	226	2	6	117	26	377
ARNOLDO MONDADORI EDI	327	469	285	416	78	1,159
ASSICURAZIONI GENERAL	860	1,002	658	773	546	3,839
ASTALDI SPA	66	158	31	57	110	422
ATLANTIA SPA	424	377	152	302	909	2,164
AUTOGRILL SPA	192	110	36	182	62	582
BANCA CARIGE SPA	24	180	28	345	20	597
BANCA GENERALI SPA	16	274	14	40	61	405
BANCA MONTE DEI PASCHI DI SIENA	336	406	186	317	216	1,461
BANCA POPOLARE DI MILANO	474	309	233	543	249	1,808
BANCA POPOLARE EMILIA ROMAGNA	79	54	30	88	187	438
BANCO POPOLARE	692	326	249	340	191	1,798
BEGHELLI SPA	12	87	8	16	394	517
BENETTON GROUP SPA	482	579	276	505	936	2,778
BREMBO SPA - FRENI BR	129	109	72	94	101	505
CALTAGIRONE SPA	117	111	76	86	15	405
CIR SPA - COMPAGNIE INDUSTRIALI RIUNITE I	244	676	206	362	170	1,658
CNH	154	163	3	169	220	709
COFIDE - GRUPPO DE BE	127	238	115	172	116	768
CREDITO EMILIANO	44	114	25	46	207	436
DIASORIN SPA	48	166	37	26	40	317
EDISON SPA	320	281	187	252	373	1,413
ENEL GREEN POWER SPA	90	101	73	67	152	483
ENEL SPA	796	1,005	633	779	1,041	4,254

Continued on next page

Table 1: *Sample composition. Number of articles published by each newspaper for all the firm (both in conflict and not in conflict firms). Showing only the firms that receive at least 100 articles on at least one newspaper.*

FIRM	Newspaper_name					Total
	Corriere della Sera	La Repubblica	La Stampa	Il Giornale	Il Resto del Carlino	
ENI SPA	1,036	1,039	703	780	563	4,121
ERG SPA	37	151	59	94	128	469
EXOR SPA	389	195	256	238	91	1,169
FIAT INDUSTRIAL SPA	113	113	182	75	155	638
FIAT SPA	2,351	3,372	2,789	2,222	2,229	12,963
FIERA MILANO SPA	204	242	46	301	231	1,024
FINMECCANICA SPA	469	794	288	590	249	2,390
FONDIARIA - SAI SPA	534	988	542	564	579	3,207
GEMINA SPA - GENERALE	169	259	98	229	92	847
GEOX SPA	59	124	49	130	38	400
GREENVISION AMBIENTE	239	5	9	214	61	528
GRUPPO EDITORIALE L'ESPRESSO	70	358	86	149	23	686
GIULIO EINAUDI EDITORE	3	21	14	109	2	149
HERA SPA	34	230	29	49	2,044	2,386
IMMSI SPA	73	89	63	88	206	519
IMPREGILO SPA	335	893	289	330	360	2,207
INDESIT COMPANY SPA	56	105	25	45	120	351
INTEK SPA	3	6	2	221	2	234
INTERPUMP GROUP SPA	17	15	6	116	79	233
INTESA SANPAOLO SPA	1,180	1,340	901	1,090	893	5,404
IREN SPA	37	264	88	103	48	540
ITALCEMENTI SPA FABBR	308	396	244	283	294	1,525
ITALMOBILIARE SPA	329	403	255	297	298	1,582
LOTTOMATICA GROUP SPA	108	192	0	297	46	643
LUXOTTICA GROUP SPA	403	554	252	238	327	1,774
MARR SPA	8	28	6	20	128	190
MEDIASET SPA	1,062	1,824	1,282	852	756	5,776
MEDIOBANCA SPA	1,417	2,486	2,328	2,774	1,578	10,583
MEDIOLANUM SPA	175	308	249	638	137	1,507
MERIDIANA FLY SPA	54	236	66	170	84	610
MILANO ASSICURAZIONI	117	130	136	150	96	629
MITTEL SPA	450	208	244	56	8	966
MONRIF SPA	4	10	6	6	397	423
MEDUSA FILM	19	11	8	703	0	741
OLIDATA SPA	4	24	10	6	156	200
PARMALAT SPA	988	1,298	1,014	566	568	4,434
PIAGGIO & C. SPA	230	432	262	368	486	1,778
PININFARINA SPA	232	822	490	188	144	1,876
PIRELLI & C. SPA	541	1,924	1,650	1,272	1,258	6,645
POLIGRAFICI EDITORIAL	16	14	10	12	594	646
POLTRONA FRAU SPA	102	60	66	52	172	452
PREMAFIN FINANZIARIA	147	344	306	266	212	1,275
PRYSMIAN SPA	278	400	346	196	446	1,666
RCS MEDIAGROUP SPA	583	290	295	293	126	1,004

Continued on next page

Table 1: *Sample composition. Number of articles published by each newspaper for all the firm (both in conflict and not in conflict firms). Showing only the firms that receive at least 100 articles on at least one newspaper.*

FIRM	Newspaper_name					Total
	Corriere della Sera	La Repubblica	La Stampa	Il Giornale	Il Resto del Carlino	
SORGENIA SPA	73	138	34	65	72	382
TERNA SPA	240	441	243	173	342	1,439
TISCALI SPA	85	146	130	160	53	574
TOD'S SPA	453	370	171	68	130	1,192
TREVI - FINANZIARIA I	29	62	16	35	168	310
UBS FIDUCIARIA SPA	198	165	138	128	80	709
UNICREDIT SPA	1,133	1,489	1,322	1,095	911	5,950
UNIPOL GRUPPO FINANZIARIO SPA	302	671	431	335	604	2,343

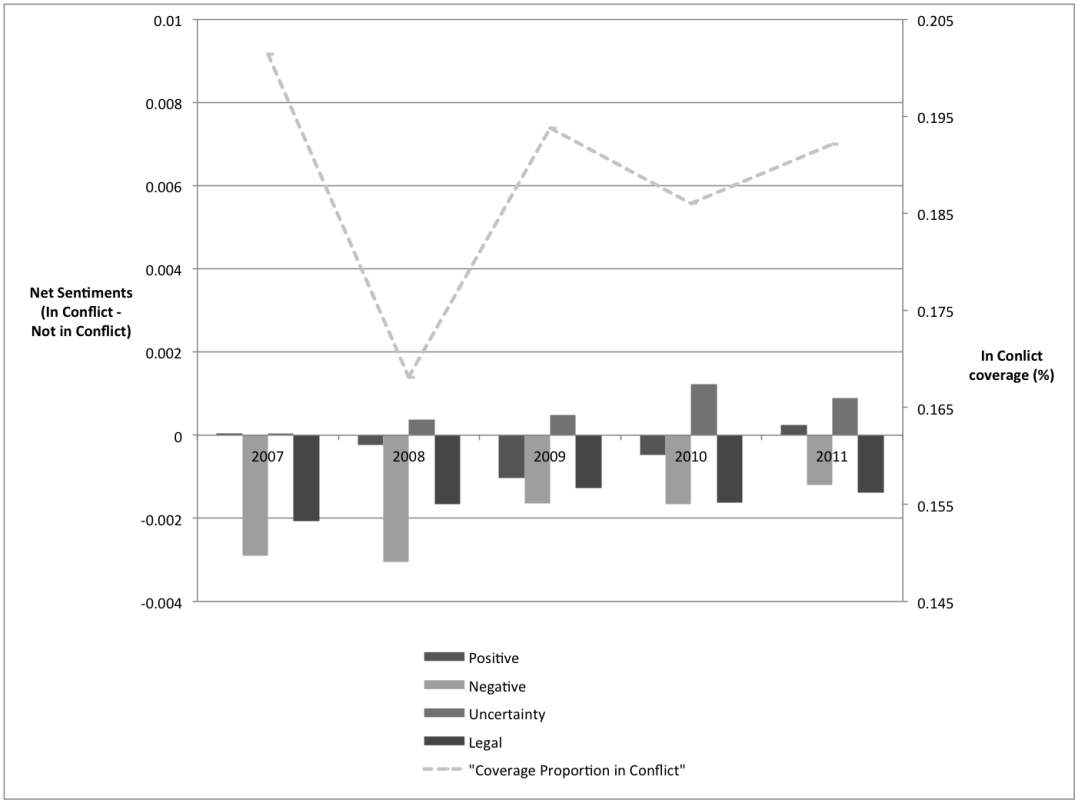

Figure 4: *Quantity and quality of coverage differences* in conflict *firms and* not in conflict *firms receive. The dotted line represents the percentage of articles published every year referred to in conflict firms. The histograms in the bottom part are representing the differences in the average sentiment between* in conflict *firms and* not in conflict *firms (every year). The first column in each histogram is the positive tone, the second the negative tone, the third the uncertainty tone, and the last one the legal tone. If a column has positive height, it means the average tone is greater for* in conflict *firms than* not in conflict *firms. If the height is negative, it is true the opposite. The most evident pattern here are those related to the negative and the legal tone: for both of them* in conflict *firms have on average significantly lower average levels.*

Table 2: *Coverage and media ownership. Descriptive statistics.*

	2007	2008	2009	2010	2011	*TOTAL*
			Corriere della Sera			
Articles in Conflict	2,653	1,519	1,698	1,970	2,050	*9,890*
Total Articles	6,110	3,998	4,165	5,117	5,729	*25,119*
Conflict Dimension	**43.42%**	**37.99%**	**40.77%**	**38.50%**	**35.78%**	***39.37%***
			La Repubblica			
Articles in Conflict	226	308	332	230	285	*1,381*
Total Articles	7,314	6,985	6,535	6,200	5,697	*32,731*
Conflict Dimension	**3.09%**	**4.41%**	**5.08%**	**3.71%**	**5.00%**	***4.22%***
			La Stampa			
Articles in Conflict	1,182	692	954	859	930	*4,617*
Total Articles	4,737	3,652	3,943	3,724	4,306	*20,362*
Conflict Dimension	**24.95%**	**18.95%**	**24.19%**	**23.07%**	**21.60%**	***22.67%***
			Il Giornale			
Articles in Conflict	818	673	720	670	473	*3,354*
Total Articles	5,692	4,389	5,014	5,140	3,893	*24,128*
Conflict Dimension	**14.37%**	**15.33%**	**14.36%**	**13.04%**	**12.15%**	***13.90%***
			Il Resto del Carlino			
Articles in Conflict	179	112	136	196	174	*797*
Total Articles	6,316	3,931	3,996	4,864	5,743	*24,850*
Conflict Dimension	**2.83%**	**2.85%**	**3.40%**	**4.03%**	**3.03%**	***3.21%***
			All Newspapers			
Articles in Conflict	5,058	3,304	3,840	3,925	3,912	*20,039*
Total Articles	30,169	22,955	23,653	25,045	25,368	*127,190*
Conflict Dimension	**16.77%**	**14.39%**	**16.23%**	**15.67%**	**15.42%**	***15.76%***

Table 3: *Summary statistics. Sample coverage: quantity and quality. For all the sample summary statistics are reported.Panel A reports the average, the standard deviation, minimum and maximum value, and media for both the quantity of coverage (Words) and quality of coverage (Positive, Negative, Uncertainty, Legal, Modal Weak, and Modal Strong tones). Quantity of coverage unit of measure is thousand of words. Quality of coverage unit of measure is percentage points representing the proportion of each category tonal words on the total number of words for each article. Panel B is instead reporting the proportion of articles published on each of the five considered newspaper, while Panel C is reporting the proportion of articles published on each of the five considered year.*

Variable	Obs	Mean	Std. Dev.	Min	Max	P50
PANEL A	*Quantity and quality of coverage*					
Words (1,000)	127190	.344	.247	.023	4.719	.265
Positive (%)	127190	.017	.013	-.009	.184	.014
Negative (%)	127190	.022	.015	-.015	.319	.02
Uncertainty (%)	127190	.011	.009	0	.243	.01
Legal (%)	127190	.013	.012	0	.193	.01
Modal_Weak (%)	127190	.004	.005	0	.204	.003
Modal_Strong (%)	127190	.005	.005	0	.106	.004
PANEL B	*Coverage by newspaper*					
	Newspaper	*Obs*	*Proportion*			
	CorriereDellaSera	127190	.188			
	LaRepubblica	127190	.261			
	LaStampa	127190	.169			
	IlGiornale	127190	.189			
	IlRestoDelCarlino	127190	.193			
PANEL C	*Coverage by year*					
	Year	*Obs*	*Proportion*			
	2007	127190	.237			
	2008	127190	.181			
	2009	127190	.182			
	2010	127190	.197			
	2011	127190	.203			

Table 4: *Correlation matrix table. All the indicators of quality and quantity of coverage are taken in account.*

	Words	Positive	Negative	Uncertainty	Legal	Modal_Weak	Modal_Strong
Words	1						
Positive	-0.1264	1					
Negative	-0.0367	0.0192	1				
Uncertainty	0.2008	-0.1072	0.0681	1			
Legal	-0.1166	-0.1205	0.2109	0.1128	1		
Modal_Weak	0.2121	-0.1308	0.0026	0.6847	0.0055	1	
Modal_Strong	0.2002	0.0741	0.062	0.147	-0.071	0.1412	1

Table 5: *Correlation matrix table. Correlations between all the indicators of quality and quantity of coverage and all the newspapers examined.*

	Words	Positive	Negative	Uncertainty	Legal
Corriere Della Sera	0.1013	-0.019	-0.03	0.0966	-0.0515
La Repubblica	0.0734	0.005	0.0736	-0.0273	-0.1162
La Stampa	0.0321	0.0734	0.0388	-0.0325	-0.0764
Il Giornale	-0.0542	-0.0321	-0.0866	0.0022	0.1271
Il Resto Del Carlino	-0.1588	-0.0245	-0.003	-0.0365	0.1266

Table 6: *Correlation matrix table. Correlations between all the indicators of quality and quantity of coverage and all the years in which the examined articles are published.*

	Words	Positive	Negative	Uncertainty	Legal
2007	0.0596	-0.0355	-0.0415	0.0573	-0.0144
2008	-0.0129	-0.005	0.0278	0.0152	0.0004
2009	-0.0424	0.0181	0.015	-0.0166	0.0091
2010	-0.0263	0.0163	-0.024	-0.0375	-0.0014
2011	0.016	0.0089	0.0266	-0.022	0.0074

Table 7: *Regression model explaing coverage as defined as total number of articles and total number of words.*

	(1) N_Articles	(2) N_Articles	(3) Total_Words	(4) Total_Words
Related parts firms dummy	559.4***		245526.9***	
	(96.22)		(102.25)	
Media Firm Share Owned		1032.1***		456013.9***
		(98.06)		(104.98)
Corriere della Sera	-160.9***	15.01**	16853.1***	93989.0***
	(-24.32)	(2.39)	(6.17)	(36.31)
La Repubblica	219.6***	224.2***	170815.0***	172817.0***
	(37.24)	(38.06)	(70.12)	(71.09)
La Stampa	12.88*	-7.830	74935.7***	65469.1***
	(1.92)	(-1.16)	(26.98)	(23.50)
Il Giornale	-106.7***	-52.88***	3410.5	27008.1***
	(-16.75)	(-8.35)	(1.30)	(10.34)
Constant	629.0***	619.3***	168253.9***	163900.1***
	(141.32)	(139.16)	(91.52)	(89.24)
Observations	127432	127432	127432	127432

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8: *Two-sample t test with equal variances: Negative Tone*

Group	Obs	Mean	Std. Err.	Std. Dev.	95 % Conf. Interval	
Not in Conflict	107358	0.0226444	0.0000478	0.0156767	0.0225507	0.0227382
In Conflict	20074	0.0204674	0.0000973	0.0137907	0.0202766	0.0206582
combined	127432	0.0223015	0.0000432	0.0154153	0.0222169	0.0223861
diff		0.0021771	0.0001184		0.001945	0.0024091
diff = mean(0) - mean(1)						t = 18.3902
Ho: diff = 0					Degrees of freedom = 127430	
Ha: diff < 0			Ha: diff != 0			Ha: diff > 0
Pr(T < t) = 1.0000			Pr(T > t) = 0.0000			Pr(T > t) = 0.0000

Table 9: *Regression table: OLS regressions with Conflict of interest as a dummy. Newspapers, Firms, and Quarters fixed effects considered.*

	(1) Words	(2) Positive	(3) Negative	(4) Uncertainty	(5) Legal	(6) Modal_Weak	(7) Modal_Strong
Related parts firms dummy	0.242	-0.000124	-0.00245***	-0.000260***	-0.000617***	-0.0000899*	0.0000103
	(0.10)	(-0.94)	(-15.61)	(-2.80)	(-5.15)	(-1.75)	(0.20)
Corriere della Sera	120.9***	0.000289**	-0.000549***	0.00237***	-0.00484***	0.00149***	0.000619***
	(49.76)	(2.28)	(-3.62)	(26.42)	(-41.77)	(29.94)	(12.45)
La Repubblica	109.0***	0.000505***	0.00113***	0.0000653	-0.00607***	0.000404***	0.000278***
	(51.67)	(4.59)	(8.60)	(0.84)	(-60.29)	(9.36)	(6.44)
La Stampa	90.43***	0.00236***	0.000385**	-0.000391***	-0.00575***	0.000197***	-0.000423***
	(37.27)	(18.64)	(2.54)	(-4.37)	(-49.71)	(3.98)	(-8.53)
Il Giornale	37.07***	-0.000126	-0.00292***	0.000621***	-0.0000198	0.000207***	-0.000369***
	(16.07)	(-1.04)	(-20.26)	(7.30)	(-0.18)	(4.40)	(-7.81)
Constant	321.1***	0.0154***	0.0252***	0.0143***	0.0162***	0.00526***	0.00517***
	(40.99)	(37.58)	(51.46)	(49.35)	(43.43)	(32.85)	(32.24)
Quarters Dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Firm Dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Observations	127190	127190	127190	127190	127190	127190	127190

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 10: *Regression table: OLS regressions with conflict of interest measured as the proportion of share owned. Newspapers, Firms, and Quarters fixed effects considered.*

	(1) Words	(2) Positive	(3) Negative	(4) Uncertainty	(5) Legal	(6) Modal_Weak	(7) Modal_Strong
Media Firm Share Owned	-6.392	0.0000359	-0.00463***	-0.000808***	-0.000747***	-0.000517***	0.0000393
	(-1.43)	(0.15)	(-16.59)	(-4.90)	(-3.50)	(-5.66)	(0.43)
Corriere della Sera	121.3***	0.000243**	-0.00124***	0.00231***	-0.00503***	0.00148***	0.000621***
	(53.70)	(2.06)	(-8.80)	(27.72)	(-46.73)	(32.00)	(13.44)
La Repubblica	109.1***	0.000502***	0.00116***	0.0000738	-0.00607***	0.000411***	0.000278***
	(51.70)	(4.56)	(8.83)	(0.95)	(-60.26)	(9.52)	(6.43)
La Stampa	91.37***	0.00234***	0.000764***	-0.000306***	-0.00572***	0.000261***	-0.000428***
	(36.67)	(18.00)	(4.92)	(-3.33)	(-48.10)	(5.11)	(-8.39)
Il Giornale	37.17***	-0.000134	-0.00301***	0.000615***	-0.0000486	0.000208***	-0.000369***
	(16.15)	(-1.11)	(-20.95)	(7.24)	(-0.44)	(4.43)	(-7.82)
Constant	320.5***	0.0153***	0.0223***	0.0139***	0.0156***	0.00511***	0.00518***
	(42.66)	(38.91)	(47.59)	(50.21)	(43.43)	(33.28)	(33.71)
Quarters Dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Firm Dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Observations	127190	127190	127190	127190	127190	127190	127190

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 11: *Regression table: OLS regressions with Conflict of interest as a dummy, including also the unlisted firm dummy variable.. Newspapers, Firms, and Quarters fixed effects considered.*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Words	Positive	Negative	Uncertainty	Legal	Modal_Weak	Modal_Strong
Related parts firms dummy	0.242	-0.000124	-0.00245***	-0.000260***	-0.000617***	-0.0000899*	0.0000103
	(0.10)	(-0.94)	(-15.61)	(-2.80)	(-5.15)	(-1.75)	(0.20)
Unlisted	-90.68***	-0.00339*	0.00161	-0.000602	0.000565	0.000733	0.000975
	(-2.71)	(-1.94)	(0.77)	(-0.49)	(0.35)	(1.07)	(1.42)
Corriere della Sera	120.9***	0.000289**	-0.000549***	0.00237***	-0.00484***	0.00149***	0.000619***
	(49.76)	(2.28)	(-3.62)	(26.42)	(-41.77)	(29.94)	(12.45)
La Repubblica	109.0***	0.000505***	0.00113***	0.0000653	-0.00607***	0.000404***	0.000278***
	(51.67)	(4.59)	(8.60)	(0.84)	(-60.29)	(9.36)	(6.44)
La Stampa	90.43***	0.00236***	0.000385**	-0.000391***	-0.00575***	0.000197***	-0.000423***
	(37.27)	(18.64)	(2.54)	(-4.37)	(-49.71)	(3.98)	(-8.53)
Il Giornale	37.07***	-0.000126	-0.00292***	0.000621***	-0.0000198	0.000207***	-0.000369***
	(16.07)	(-1.04)	(-20.26)	(7.30)	(-0.18)	(4.40)	(-7.81)
Constant	321.1***	0.0154***	0.0252***	0.0143***	0.0162***	0.00526***	0.00517***
	(40.99)	(37.58)	(51.46)	(49.35)	(43.43)	(32.85)	(32.24)
Quarters Dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Firm Dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Observations	127190	127190	127190	127190	127190	127190	127190

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 12: *Regression table: OLS regressions with Conflict of interest as a dummy, including also the unlisted firm dummy variable. Newspapers, Firms, and Quarters fixed effects considered.*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Words	Positive	Negative	Uncertainty	Legal	Modal_Weak	Modal_Strong
Media Firm Share Owned	-6.392	0.0000359	-0.00463***	-0.000808***	-0.000747***	-0.000517***	0.0000393
	(-1.43)	(0.15)	(-16.59)	(-4.90)	(-3.50)	(-5.66)	(0.43)
Unlisted	-87.35***	-0.00339*	0.00443**	-0.000136	0.00105	0.00102	0.000953
	(-2.60)	(-1.93)	(2.11)	(-0.11)	(0.66)	(1.48)	(1.39)
Corriere della Sera	121.3***	0.000243**	-0.00124***	0.00231***	-0.00503***	0.00148***	0.000621***
	(53.70)	(2.06)	(-8.80)	(27.72)	(-46.73)	(32.00)	(13.44)
La Repubblica	109.1***	0.000502***	0.00116***	0.0000738	-0.00607***	0.000411***	0.000278***
	(51.70)	(4.56)	(8.83)	(0.95)	(-60.26)	(9.52)	(6.43)
La Stampa	91.37***	0.00234***	0.000764***	-0.000306***	-0.00572***	0.000261***	-0.000428***
	(36.67)	(18.00)	(4.92)	(-3.33)	(-48.10)	(5.11)	(-8.39)
Il Giornale	37.17***	-0.000134	-0.00301***	0.000615***	-0.0000486	0.000208***	-0.000369***
	(16.15)	(-1.11)	(-20.95)	(7.24)	(-0.44)	(4.43)	(-7.82)
Constant	320.5***	0.0153***	0.0223***	0.0139***	0.0156***	0.00511***	0.00518***
	(42.66)	(38.91)	(47.59)	(50.21)	(43.43)	(33.28)	(33.71)
Quarters Dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Firm Dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Observations	127190	127190	127190	127190	127190	127190	127190

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 13: *Robusteness checks. TOBIT regressions with Conflict of interest as a dummy, including also the unlisted firm dummy variable. Newspapers, Firms, and Years fixed effects considered.*

	(1) Positive	(2) Negative	(3) Uncertainty	(4) Legal	(5) Modal_Weak	(6) Modal_Strong
Related parts firms dummy	-0.000134	-0.00270***	-0.000346***	-0.000652***	-0.000149*	-0.0000156
	(-0.95)	(-15.96)	(-3.07)	(-4.98)	(-1.75)	(-0.20)
Corriere della Sera	0.000468***	-0.000127	0.00295***	-0.00589***	0.00268***	0.00132***
	(3.42)	(-0.78)	(26.95)	(-46.72)	(32.15)	(17.42)
La Repubblica	0.000832***	0.00124***	-0.0000143	-0.00740***	0.000871***	0.000658***
	(6.99)	(8.69)	(-0.15)	(-67.50)	(11.86)	(9.92)
La Stampa	0.00281***	0.000761***	-0.000408***	-0.00674***	0.000563***	-0.000349***
	(20.56)	(4.66)	(-3.72)	(-53.52)	(6.67)	(-4.56)
Il Giornale	0.0000996	-0.00359***	0.000412***	-0.0000377	0.000256***	-0.000668***
	(0.77)	(-22.95)	(3.94)	(-0.32)	(3.18)	(-9.12)
Constant	0.0146***	0.0251***	0.0139***	0.0164***	0.00370***	0.00386***
	(33.28)	(47.63)	(39.52)	(40.27)	(14.04)	(15.81)
sigma						
Constant	0.0132***	0.0158***	0.0104***	0.0122***	0.00751***	0.00700***
	(474.13)	(472.59)	(430.21)	(471.32)	(345.85)	(367.39)
Years Dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Firm Dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Observations	127190	127190	127190	127190	127190	127190

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 14: *Robusteness checks. TOBIT regressions with conflict of interest measured as the proportion of share owned. Newspapers, Firms, and Years fixed effects considered.*

	(1)	(2)	(3)	(4)	(5)	(6)
	Positive	Negative	Uncertainty	Legal	Modal_Weak	Modal_Strong
Media Firm Share Owned	0.0000102	-0.00518***	-0.000994***	-0.00103***	-0.000769***	-0.0000105
	(0.04)	(-17.19)	(-4.94)	(-4.42)	(-5.07)	(-0.08)
Corriere della Sera	0.000419***	-0.000889***	0.00286***	-0.00609***	0.00266***	0.00132***
	(3.29)	(-5.84)	(28.19)	(-51.95)	(34.37)	(18.67)
La Repubblica	0.000829***	0.00127***	-0.00000452	-0.00739***	0.000881***	0.000658***
	(6.97)	(8.93)	(-0.05)	(-67.43)	(11.99)	(9.91)
La Stampa	0.00279***	0.00119***	-0.000306***	-0.00667***	0.000659***	-0.000350***
	(19.90)	(7.08)	(-2.71)	(-51.56)	(7.59)	(-4.44)
Il Giornale	0.0000911	-0.00369***	0.000403***	-0.0000647	0.000256***	-0.000669***
	(0.70)	(-23.66)	(3.87)	(-0.55)	(3.19)	(-9.15)
Constant	0.0145***	0.0219***	0.0134***	0.0157***	0.00346***	0.00384***
	(34.42)	(43.44)	(39.88)	(40.12)	(13.69)	(16.39)
sigma						
Constant	0.0132***	0.0158***	0.0104***	0.0122***	0.00751***	0.00700***
	(474.13)	(472.59)	(430.21)	(471.33)	(345.85)	(367.39)
Years Dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Firm Dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Observations	127190	127190	127190	127190	127190	127190

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 15: *Regression table regressions with interaction variable between conflict of interest dummy and newspaper dummies. Conflict of interest measured as the proportion of share owned. Newspapers, Firms, and Year fixed effects considered.*

	(1) Words	(2) Positive	(3) Negative	(4) Uncertainty	(5) Legal	(6) Modal_Weak	(7) Modal_Strong
Related parts firms dummy	44.35***	-0.000199	0.000458	-0.000411	0.000298	-0.000112	-0.0000586
	(2.76)	(-0.24)	(0.46)	(-0.69)	(0.39)	(-0.34)	(-0.18)
Corriere della Sera	125.4***	0.000477***	-0.000776***	0.00231***	-0.00512***	0.00144***	0.000731***
	(47.94)	(3.49)	(-4.75)	(23.95)	(-41.00)	(26.80)	(13.65)
La Repubblica	34.56*	0.00169*	-0.00320***	0.000726	-0.00637***	0.000272	0.000617*
	(1.91)	(1.79)	(-2.84)	(1.09)	(-7.39)	(0.73)	(1.67)
La Stampa	51.47***	0.00258***	-0.00345***	-0.000308	-0.00606***	-0.0000575	-0.000302
	(3.07)	(2.95)	(-3.29)	(-0.50)	(-7.58)	(-0.17)	(-0.88)
Il Giornale	31.34*	0.000340	-0.00577***	0.000282	-0.00421***	0.000355	0.000342
	(1.82)	(0.38)	(-5.36)	(0.44)	(-5.12)	(1.00)	(0.97)
Newspaper=1 × Related parts firms dummy=1	-54.66***	-0.000320	-0.00235**	0.000283	-0.000219	0.000160	-0.000231
	(-3.34)	(-0.38)	(-2.30)	(0.47)	(-0.28)	(0.48)	(-0.69)
Newspaper=2 × Related parts firms dummy=0	76.24***	-0.00119	0.00438***	-0.000687	0.000229	0.000136	-0.000345
	(4.20)	(-1.26)	(3.86)	(-1.02)	(0.26)	(0.37)	(-0.93)
Newspaper=3 × Related parts firms dummy=0	38.88**	-0.000209	0.00401***	-0.0000684	0.000147	0.000314	-0.000130
	(2.30)	(-0.24)	(3.79)	(-0.11)	(0.18)	(0.91)	(-0.38)
Newspaper=4 × Related parts firms dummy=0	2.910	-0.000453	0.00283***	0.000379	0.00443***	-0.000157	-0.000767**
	(0.17)	(-0.50)	(2.60)	(0.59)	(5.34)	(-0.44)	(-2.16)
Constant	311.7***	0.0153***	0.0260***	0.0141***	0.0155***	0.00549***	0.00513***
	(36.88)	(34.56)	(49.27)	(45.28)	(38.54)	(31.73)	(29.64)
Year Dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Firm Dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Observations	127190	127190	127190	127190	127190	127190	127190

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 16: *Regression table regressions with interaction variable between conflict of interest dummy and years dummies. Conflict of interest measured as the proportion of share owned. Newspapers, Firms, and Year fixed effects considered*

	(1) Words	(2) Positive	(3) Negative	(4) Uncertainty	(5) Legal	(6) Modal_Weak	(7) Modal_Strong
Related parts firms dummy	32.49***	0.000358	-0.00177***	0.000136	0.000277	-0.0000552	0.000230**
	(7.29)	(1.54)	(-6.37)	(0.83)	(1.30)	(-0.60)	(2.52)
Corriere della Sera	120.4***	0.000318**	-0.000584***	0.00236***	-0.00486***	0.00148***	0.000610***
	(49.58)	(2.51)	(-3.85)	(26.31)	(-41.91)	(29.84)	(12.26)
La Repubblica	108.4***	0.000519***	0.00110***	0.0000529	-0.00612***	0.000401***	0.000270***
	(51.42)	(4.71)	(8.34)	(0.68)	(-60.73)	(9.28)	(6.25)
La Stampa	90.04***	0.00239***	0.000352**	-0.000392***	-0.00579***	0.000200***	-0.000430***
	(37.13)	(18.82)	(2.32)	(-4.38)	(-49.98)	(4.03)	(-8.66)
Il Giornale	36.85***	-0.0000964	-0.00295***	0.000621***	-0.0000521	0.000207***	-0.000377***
	(15.99)	(-0.80)	(-20.48)	(7.29)	(-0.47)	(4.38)	(-7.98)
Year=2007 × Related parts firms dummy=1	-51.89***	-0.000687**	-0.00174***	-0.00112***	-0.00153***	-0.0000870	-0.000387***
	(-9.33)	(-2.37)	(-5.02)	(-5.46)	(-5.76)	(-0.76)	(-3.40)
Year=2008 × Related parts firms dummy=1	-56.40***	-0.000406	-0.00145***	-0.000484**	-0.00140***	-0.00000603	-0.000479***
	(-9.23)	(-1.27)	(-3.78)	(-2.14)	(-4.79)	(-0.05)	(-3.83)
Year=2009 × Related parts firms dummy=1	-46.58***	-0.000932***	0.0000238	-0.000339	-0.000619**	-0.000155	-0.000206*
	(-7.92)	(-3.03)	(0.06)	(-1.56)	(-2.20)	(-1.29)	(-1.71)
Year=2010 × Related parts firms dummy=1	-4.403	-0.000345	-0.0000574	0.000221	-0.000784***	0.000132	-0.00000661
	(-0.75)	(-1.13)	(-0.16)	(1.03)	(-2.81)	(1.10)	(-0.06)
Year Dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Firm Dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Constant	319.3***	0.0154***	0.0252***	0.0141***	0.0163***	0.00520***	0.00520***
	(42.61)	(39.39)	(53.87)	(51.02)	(45.47)	(33.90)	(33.89)
Years Dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Firm Dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Observations	127190	127190	127190	127190	127190	127190	127190

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 17: *Regression table regressions with interaction variable between conflict of interest dummy and unlisted dummy. Conflict of interest measured as the proportion of share owned. Newspapers, Firms, and Year fixed effects considered*

	(1) Words	(2) Positive	(3) Negative	(4) Uncertainty	(5) Legal	(6) Modal_Weak	(7) Modal_Strong
Media Firm Share Owned	-5.217	-0.0000475	-0.00464***	-0.000745***	-0.000497**	-0.000510***	0.0000232
	(-1.15)	(-0.20)	(-16.41)	(-4.46)	(-2.30)	(-5.51)	(0.25)
Unlisted	-76.07**	-0.00419**	0.00435**	0.000468	0.00346**	0.00109	0.000799
	(-2.22)	(-2.34)	(2.03)	(0.37)	(2.11)	(1.55)	(1.14)
Unlisted*Confl	-15.09	0.00107**	0.000109	-0.000808**	-0.00322***	-0.0000912	0.000206
	(-1.62)	(2.20)	(0.19)	(-2.35)	(-7.24)	(-0.48)	(1.08)
Corriere della Sera	121.3***	0.000239**	-0.00124***	0.00231***	-0.00502***	0.00148***	0.000621***
	(53.72)	(2.02)	(-8.80)	(27.75)	(-46.62)	(32.00)	(13.42)
La Repubblica	109.1***	0.000506***	0.00116***	0.0000706	-0.00608***	0.000411***	0.000279***
	(51.66)	(4.60)	(8.83)	(0.91)	(-60.38)	(9.51)	(6.45)
La Stampa	91.22***	0.00235***	0.000765***	-0.000314***	-0.00575***	0.000260***	-0.000426***
	(36.58)	(18.07)	(4.92)	(-3.41)	(-48.35)	(5.09)	(-8.34)
Il Giornale	37.52***	-0.000159	-0.00301***	0.000634***	0.0000261	0.000211***	-0.000373***
	(16.23)	(-1.31)	(-20.88)	(7.43)	(0.24)	(4.45)	(-7.89)
Constant	320.7***	0.0152***	0.0223***	0.0139***	0.0156***	0.00511***	0.00518***
	(42.68)	(38.88)	(47.58)	(50.24)	(43.52)	(33.28)	(33.70)
Querter Dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Firm Dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Observations	127190	127190	127190	127190	127190	127190	127190

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 18: *Regression table: propensity score matching based estimation. Propensity score is computed basing on the number of words of each article, the listed/unlisted status of the firm, and the newspaper publishing the article itself. I then apply the nearest neighbor matching method to compare the treated sample (articles regarding related parts) and the control one.*

	(1) Positive	(2) Negative	(3) Uncertainty	(4) Legal	(5) Modal_Weak	(6) Modal_Strong
Specification 1 - Propensity Score without taking in account for listing patterns						
Related parts firms dummy	0.000175	-0.00163***	-0.000770***	-0.00130***	-0.000187***	0.000130***
	(1.43)	(-13.11)	(-9.37)	(-10.65)	(-3.60)	(2.80)
Specification 2 - Propensity Score taking in account for listing patterns						
Related parts firms dummy	0.000221*	-0.00117***	-0.000694***	-0.00102***	-0.000208***	0.000140**
	(1.94)	(-8.19)	(-6.76)	(-8.51)	(-3.56)	(2.39)
Observations	127190	127190	127190	127190	127190	127190

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Media sentiment and the pricing of IPOs

Carlo Raimondo* Emanuele Bajo†

Abstract

We study a sample of 3.061 IPOs in the US stock markets in the period between 1995 and 2013 and the related coverage and tone from the New York Times. Based on 20.648 articles collected from one year before the filing to the IPO date we document that NYT coverage positively influences both underpricing and price revision. Likewise, applying a sentiment analysis based on financial word lists, we show how the tone of news also drives part of the first day returns and price revisions. Our results are consistent with a view in which more covered and positively reported IPOs are preferred by investors generating an increase on the demand side. These results hold after controlling for common underpricing and price revision economic rationales and also considering the informational effect of S-1 content disclosure.

I. Introduction

The pricing of IPOs as well as the price behavior during the first day of trade has been widely investigated phenomenon since Ibbotson 1975 (Ibbotson 1975). The information asymmetry between the issuer and the underwriter on the one side and the investors on the other side is the main rationale for the existence of the underpricing and for the practice of the bookbuilding. In a setting with no information asymmetry professional investors would not be able to profit at the expense of retail investors. Hence, a fairly priced IPO would produce no underpricing, no disparity among market participants and no market failure. Along this line, the wide use of bookbuilding, as a technique for underwriters to enhance the price discovery, confirms the pivotal role of the information in the IPO process. According to Hanley and Wilhelm, Cornelli and Goldreich, and Aggarwal et al. (Hanley and Wilhelm Jr 1995; Cornelli and Goldreich 2001; Aggarwal, Prabhala, and Puri 2002) underwriters are willing to underprice the IPO shares in order to reward institutional investors from revealing their evaluation of the offered shares. Clearly, in a informative efficient market, the bookbuilding would have no reason to exist.

As suggested, the IPO pricing literature is deeply rooted with the role of information asymmetry. However, quite surprisingly-to the best of our knowledge-past research has devoted little attention

*PhD Student, University of Bologna, Department of Management, Via Capo di lucca 34, 40126 Bologna, Italy. Corresponding author: carlo.raimondo@unibo.it.

†Associate Professor, University of Bologna, Department of Management, Via Capo di lucca 34, 40126 Bologna, Italy.

to the relationship between the media information production and the IPO valuation. More precisely, with the sole exception of Liu et al. (Liu, Sherman, and Zhang 2014a), who investigate the effect of media coverage on the level of underpricing, no other study analyzes the role of media as an avenue to mitigate the information problem. Instead, in this paper we show that both the coverage (information volume) and the tone of the articles (sentiment) are able to significantly affect the information dissemination and in turn the level of price revision and underpricing.

This study follows an increasing stream of literature (Tetlock 2007; Tetlock 2011; Bhattacharya et al. 2009; Dougal et al. 2012; Garcia 2013; Pollock and Rindova 2003; Pollock, Rindova, and Maggitti 2008) that looks at the relationship between media and finance. Likewise, we aim at investigating the role played by newspaper financial reporting in the big "IPO game". We restrict the analysis to the role of printed newspaper because, differently than other media, space is very limited and valuable. Rationing in space directly leads to rationing in news coverage making it more visible the information reported. Following the idea that too much information is no information, having a limited available space forces the newspaper to select and report the most informative events. From the reader perspective, having a definitive set of news allows her to process the information content. This effect is obviously greater, ceteris paribus, in increasing the newspaper specific reputation and the horizon of the interest of the newspaper itself, and in decreasing of total space availability on the newspaper.

We analyze the news coverage and tone from the New York Times (NYT), as in Garcia (2013) (Garcia 2013), for 3,061 U.S. IPOs in the period 1995-2013. For each new issue we consider any news published in the NYT in the time span ranging from one year before to the trade date. This procedure allows us to collect and examine 20,648 news articles. The most desirable feature of NYT is its audience discrimination, that is the self-selective mechanism which leads the readers to belong to a very broad social group. In this meaning we are more interested in measuring the information flow towards the general public or retail (mass) investors rather than focusing on the specialized sectorial information directed to professionals (institutional investors). The difference between retail and institutional/professional investors is a pivotal concept in this study. In fact we argue that information delivered by different means of communication to different kinds and segments of audience is able to produce different effects as measured by stock market valuations. Loughran McDonald (2013) (Loughran and McDonald 2013) show the importance and the effects of information disclosure for professional. They find that the level of negative tone used in the S-1 forms is associated with greater underpricing and price revision. We differ from Loughran McDonald (2013) as we consider the effect of newspaper reporting over the same IPO dimensions. This is not minor difference as we believe that the sentiment from the filing and the media is not

only independent (as diverse is the information source) but, most important, direct to different audience.

This difference in information effect is also related to the difference in the timing in which different investors may enter in the market, especially considering the complex and long dynamics of an IPO process. Better-informed (institutional) investors are more likely to invest during the pre-market period, also taking part in the book-building phase. We argue that informed investors, who have access to higher detail of information and possess and greater ability to process raw information, are less affected by the way financial newspaper present and convey the information on IPO shares. To the contrary, retail (uninformed) investors, who take part to the IPO process in a later stage, are more influenceable by the tone used from primary media source. Along this line we test whether a causality between media coverage/sentiment and IPO characteristics exists, taking in particular account the first-day returns dynamics. Our aim is to show that both a larger and a contently more positive coverage is able to significantly influence the IPO process, generating an increase in the not-qualified demand, leading so to a greater demand for the stock on the market and consequentially to a more severe underpricing.

Differently than the most recent improvements in the literature (like Liu et al. 2014 (Liu, Sherman, and Zhang 2014b)), we are able to determine the quality (and not only the volume) of the coverage each IPO receives. Media are indeed able to impact over their audience not only as a result of the mere coverage (*attention effect*) but also through the judgement embedded into the articles. Indeed, we show that the content (tone) effect is different and additional to the respect of the coverage effect. We measure the tone of the news detecting the degree of positiveness, negativeness and uncertainty *positive, negative and uncertain sentiment*. While we do not find any effects on the negative sentiment, we do find an effect of positive sentiment on both underpricing and price revision. The uncertainty is measured through different metrics, each of them capturing different angles: (a) a direct proxy of uncertain tone (*Uncertainty sentiment*), (b) a proxy for litigious/legal tone (*Litigious sentiment*) in order to understand to what extent each firm is associated which legal concerns and (c) *modal weak* and (d) *modal strong* tone, that measure the level of self-confidence and certainty in the way in which each text has been written, increasing for higher level of *modal strong sentiment* and lower level of *modal weak sentiment*. All these measures find their foundation and validity in Loughran and McDonald 2011. Even though the importance of uncertainty has been emphasized in the theoretical and empirical financial research, there is unanimity on which could be the direction of the effect of uncertainty on the IPO process. Though, we propose two competing hypotheses regarding the relationship between

media-driven uncertainty and IPO pricing.

This paper is organized as follow: in Section II a literature review will be proposed, merging together key contributions coming from both the two streams of related literature: IPO and the Media and finance. In Section III we provide detail on data and methodology, and we put forward our main hypotheses. Section IV presents the empirical results whilst in Section V we provide some robustness checks. Finally, Section VI concludes.

II. Related researches and hypotheses development

This paper grounds its theoretical foundations on two different streams of literature. On the one hand, under the idea that media act as a mean to reduce the information gap between issuing firm and retail investors, we aim at marginal contributing to the role of information asymmetry in shaping the pricing of IPOs. On the other hand, we want to contribute to the growing literature on the relationship between media and finance, shedding some light on the effect of newspaper coverage and tone over a uninvestigated financial dimension such as the *initial public offering*. We believe that this empirical setting we analyze in this paper provides a clean test to understand to what extent the media industry really influences the financial markets. In fact, before the listing the issuing firm is generally little known among the retail investors. To the contrary, in the process of going public any firm experiences a potential dramatic change in visibility, the degree at which that occurs is greatly determined by the amount and the quality of media coverage. As not all the IPO firms receive equal attention from the media, we are able to disentangle the effect of coverage from the other traditional investigated firm and deal's characteristics in determining the observed level of underpricing.

The role of media has become increasingly relevant. The cost of collecting, filtering and analyzing financial information acts as a barrier for uninformed investors, who can hence only rely on information that is publicly available and easy to access. Newspapers respond to both these characteristics as they (a) select - due to a scarcity of printed space - a limited amount of information and (b) present the information in a way that becomes accessible (comprehensible) to its readers. Over the last decade the political science and the financial literature has investigated the effects of media reporting, under the idea that media can both enhance the level of information among market participants but likewise produce cognitive bias. Beasly and Pratt (2006) (Besley and Prat 2006) propose a model for the relation between media and political power, underlying how dangerous could be the *noncoercive caption* of the media by the governments. Evidence about the problem of misreporting are documented also by Djankov et al (Djankov et al. 2003), and

by Mullainathan and Shleifer (Mullainathan and Shleifer 2005), who suggest the risk of having media with concentrated ownership, government-friendly attitude, and audience bearing biased news. The presence of systematically slanted news has been proven to be effective and potential dangerous in term of outcomes, as not-correctly informed individuals tend to take second best decisions. Della Vigna and Kaplan (DellaVigna and Kaplan 2007) show as republican friendly news reporting by the *Fox News* channel has critically changed the electoral outcomes in several electoral districts. Gentzkow et al. (Gentzkow, Glaeser, and Goldin 2006) highlight how the *bias* phenomenon could be limited in presence of specific market characteristics, like competition between media, audience capability to understand and eventually judge misreporting, and media specific reputational effect. Media may influence the audience also when they do not produce high-level products in terms of quality of information. This trait makes media potentially harmful but also potentially important as defense against illicit and damaging behaviours, being able to permanently jeopardize an organization's reputation. Dyck et al. (Dyck, Moss, and Zingales 2013) presents a model in which media successfully prevent legislators to act against public interest. Several papers (Miller 2006, Dyck et al. 2008, Dyck et al. 2010 (Miller 2006; Dyck, Volchkova, and Zingales 2008; Dyck, Morse, and Zingales 2010)) show how media may be able to act as *watchdog* against frauds, expropriations and other harmful behaviours. (Miller 2006; Dyck, Volchkova, and Zingales 2008; Dyck, Morse, and Zingales 2010).

Other scholar focus instead on the effectiveness on media on the asset pricing, from a wide set of viewpoints. Tetlock (Tetlock 2007) studies the relation between media outcomes, investors' sentiment and financial market, while Fang and Peress (Fang and Peress 2009) focus instead on media role in resolving informational frictions. Tetlock (Tetlock 2011) describes the difference between actual news effect and the effect of already available information on which media focus the public attention. Consistently with this paper, Liu et al. (Liu, Sherman, and Zhang 2014b) show as the *attention* effect after a company goes public is a mean reverted process when considering the long run. Dougal et al. (Dougal et al. 2012) instead find out how the tone of a *Wall Street Journal* popular column published all the days is related to stock market overall performance while Garcia (Garcia 2013) considers instead articles published on *The New York Times*. Media though are able to affect stock markets in many different ways. They are able to convey important information about irregular affairs (Dyck, Morse, and Zingales 2010; Dyck, Volchkova, and Zingales 2008), they may generate a common sense of trust or of absence of trust on market makers (Stulz 1999; Moy and Pfau 2000), they may also convey information that is totally already available generating in this way abnormal performances just for have mentioned something (Huberman and Regev 2001; Chan 2003; Barber and Odean 2008). Most importantly though media convey to financial markets

huge amount of real information, exercising their prominent role of collecting, elaborating, and dealing pieces of news. This holds also for *initial public offerings* as many other corporate events.

The other stream of literature this paper refers to is the IPO pricing. Explaining IPOs patterns and specifically the dynamics of IPOs underpricing has been long debated in the financial literature and several have been the offered rationales. However, the pivotal role of information and informative asymmetries represents a common thread in most of the used arguments. Starting from the evidence that there are investors better informed than others, Rock (Rock 1986) develops a model based on the lemon-model by Akerlof (Akerlof 1970). Rock argues that uniformed investors, being less able to assess the IPO quality, tend to overweight their portfolios with not-valuable IPOs, suffering in this way a sort of *winner curse*. In equilibrium uninformed investors earn zero initial returns while informed investors earn returns that can just cover the cost sustained to acquire information. Consistently with this view, Aggarwal, Prabhala,and Puri (Aggarwal, Prabhala, and Puri 2002) show that institutional investors earn greater returns than retail investors. Other scholars (Benveniste and Spindt 1989; Hanley 1993) underline how the disclosure of positive information before the settlement of the IPO offer price, positively influences the offer price[1], directly leading to a lower IPO underpricing. A recent paper by Loughran McDonald (Loughran and McDonald 2013) seems confirming this view, as they show that a more negative content in the IPO S-1 filings is associated with a more severe underpricing phenomenon. If information disclosure is able to produce more precise offer price and consequentially smaller IPO underpricing, why issuers do not reveal to the market a more accurate informative set, reducing in this way the information asymmetry? Sherman Titman (Sherman and Titman 2002) suggest the presence of a trade-off between issuer-specific benefits of a more accurate offer price and the cost of disclose information. Ljungqvist, Jenkinson,and Wilhelm (Ljungqvist, Jenkinson, and Wilhelm Jr 2003) and Loughran Ritter (Loughran and Ritter 2002; Loughran and Ritter 2004) both emphasize the role played by underwriters and investment-banks during the IPO process: the former highlighting the role of investment-banks in disclosing information aimed at mitigating underpricing, the latter in suggesting the existence of an agency problem between issuers and underwriters generated by the underwriters' payment mechanism[2]. Consistent with this view we argue that media coverage can play an important role in enhancing the information set available to investors. More specifically, we believe that IPOs receiving larger coverage (i.e., greater number of articles or wider spaces in the printed newspapers) become more visible to the investors and this effect in turns translate

[1]Positive information disclosure, if reliable, shifts the informational asymmetry in favour of IPO issuer, making possible for the issuer itself to increase the offer price.

[2]The underwriter fees are usually directly proportional to the total proceeds that are usually bigger in a expected-underpriced IPO.

into a higher demand for shares. Based on the demand upward shift, (a) underwriters are likely to upward revise the offer price and (b) the unmet demand during the offer period is likely to produce a splillover effect on the first day of trading, causing a larger positive returns. Accordingly, we postulate the following hypotheses:

A larger media coverage is positively associated with both price revision (H1) and underpricing (H2) .

The magnitude of underpricing or price revision is not only related to the amount of available information around the IPO process but also to the precision (quality) of the information disclosed. Initial public offerings associated with a greater level of uncertainty - especially about the IPO value itself - result in higher expected underpricing (Ritter 1984; Beatty and Ritter 1986). Consistent with this argument several scholars (Ross 1977; Ritter 1984; Hughes 1986; Allen and Faulhaber 1989) point out the importance for IPO issuers to signal their quality to the prospective investors as a means to reduce the expected underpricing. Along this line, hiring reputable underwriters (Booth and Smith II 1986) or having being VC-backed (Megginson and Weiss 1991; Lee and Wahal 2004) have been shown to be indirect way to signal the IPO quality. Hence, if the amount of information spread out by the media may serve as a means to make the issuing firm more visible for the uninformed investors, the way the same information is reported may likewise produce important and potentially diverse effects. Given a particular information reported in a newspaper, the tone used in the article may completely change the perception of the reader (investor) regarding the IPO quality/pricing and accordingly modify the investment behavior. Following this intuition we postulate a relationship between the sentiment used in the article and the two studied dimensions: underpricing and price revision. Specifically we argue that a more positive (negative) sentiment, everything else constant, make retail investors more (less) incline to demand shares either during the offering period or in the aftermarket. If most of the media information flows into the market during the offering period, we hypothesize a positive (negative) relationship between positive (negative) tone and price revision; if the media covers the IPO right before the first trade date, we postulate the same effect materializing over the level of underpricing. These effects are not necessarily mutually exclusive. If newspapers consistently send positive (negative) signals to the market during the entire IPO offering period the media effect (investor perception) may be continuously reinforced leading to upward (downward) price revision as well as positive (negative) first-day returns. Hence, we posit the following hypotheses:

A positive (negative) media sentiment is positively (negatively) associated with the level of price revision

(H3) and underpricing (H4)

The other investigated sentiment is the degree of uncertainty. The possible of effect of the uncertain tone is less clear-cut. On the one hand we may rely on the demand-for-share story. If media are successful in increasing the perceived level of uncertainty among retail investors, they should decrease their demand for shares. Consequently we could expect, in response to a higher uncertain media sentiment, lower underpricing. Consequently, we may suggest the following competing hypotheses:

A larger degree of media uncertainty is negatively associated with the level of underpricing (H5a)

On the other hand, the uncertainty *per se* does not necessarily reduce the appetite for shares if the expected return is higher. Underwriters may react to the higher perceived uncertainty by setting the offer price even further below its fundamental value, increasing the expected underpricing and compensating prospective investors with larger first-day returns. Accordingly, we should observe a positive relationship between media uncertainty and the level of underpricing, leading to the following hypothesis:

A larger degree of media uncertainty is positively associated with the level of underpricing (H5b)

The level of media uncertainty is also likely to affect price revision towards two possibly opposite directions. In response to the higher degree of perceived uncertainty, underwriters may voluntarily underprice the offer. This action is unlikely to take place during the filing period, when the filing range is determined, as most of the media coverage has not produced its effects. Therefore, we could expect that underwriters may react to the uncertain media sentiment by downward revising the offer price. If this might be the case, we could push forward the following hypothesis:

A larger degree of media uncertainty is negatively associated with the price revision (H6a)

However, a positive association between media uncertain sentiment and price revision is also possible. The increased level of uncertainty makes it the stock valuation a more arduous task for the price-setters. Underwriters may therefore revise the offer price, relative to the mid of the filing range, when they access to higher information set gathered through the bookbuilding process. Accordingly, we should posit a positive relationship between media uncertainty and price revision:

A larger degree of media uncertainty is positively associated with the price revision (H6b)

To sum up, we expect to find a positive relation between media coverage and both price revision and first-day returns (*visibility effect*), as well as a concordant effect of the media sentiment (positive/negative) over the same IPO phenomena. We consider retail investors to be influenced by media outcomes in a sense of being more likely to buy stocks they have heard about (*coverage effect*) and they have heard about in a positive way (*sentiment effect*). Instead, we posit a number of competing hypotheses regarding the level of uncertainty produced by the media coverage that might potentially lead to an increasing or decreasing effect over the price revision and the level of underpricing.

We also argue that media coverage produces an independent and different effect than what mandatory disclosure generates[3]. Media do not report news basing exclusively on the contents that firms voluntarily disclose during the IPO process.

III. Data and Methodology

I. Data sample

We collect a sample of 3,061 IPOs in the U.S. stock markets between January 1995 and December 2013 from the SDC/Platinum Global New Issues database. We exclude from sample real estate investment trusts (REIT), closed-end funds, unit IPOs and unit investment trusts, rights issues, spin-offs, equity carve-outs, financial firms (with SIC codes between 6000 and 6999), foreign firms, leveraged buy-out firms (LBO), shelf-registrations, withdrawn IPOs, IPOs with an offer price lower than five dollar, and duplicate observations. Accounting data from Compustat and stock price data from CRSP. We obtain firm age data from Jay Ritter's website as well as the analyst coverage data. News are collected from the Lexis-Nexis database. We search for articles related to each IPO on the *New York Times*, using the firm name as key word and allowing for common variations.[4] We furthermore exclude from the dataset all the articles with no real informational content, i.e. articles that are just a list of incoming or just happened IPOs, and we keep track about their presence[5]. News search range estands between one year before the filing date to the fist-day

[3]The informational content disclosed by the issuing firm through the S-1 files that have to be delivered to the SEC to enter the IPO process.

[4]We firstly search for the complete company name and then we allow for common variations and abbreviations like *Inc, Co, Corp, Intl, Lmt, Limited, Group*

[5]The *New York Times* also publishes on a weekly base a complete list of incoming or present IPOs, without providing further information other than the name of the issuing firms and the size of the IPOs (total proceeds and number of shares

of trading. This search generates a dataset of 20,648 articles covering 1,561 IPOs out of 3.061.
We measure the influence of media on the pricing of IPOs taking into consideration the news reported in the *New York Times* (as in Garcia 2013). The New York Times is first in circulation among daily U.S. newspaper, if we exclude specialized press such as *Wall Street Journal* and tabloids.[6]
Finally, S-1 forms are obtained form EDGAR. Using as search criterion the CIK (Central Index Key) we are able to match 2,082 S-1 forms out of our 3,061 IPOs that comprise our sample. As for the sample of news, we process these documents in order to infer the embedded sentiment, excluding not meaningful particles, such as ASCII- encoded segments (i.e. pdf's, jpg's), HTML, and XBRL (Loughran and McDonald 2013).

II. Textual analysis

Textual analysis for both articles and S-1 forms is carried out as in Loughran and McDonald 2011. This methodology, widely used and tested in the literature (Tetlock 2011; Dougal et al. 2012; Garcia 2013), is based on searching from a financial document a list of words that are likely to be associated with some *sentiment* attributes (*positiveness*, *negativeness*, *uncertainty and litigiousness*). The frequency of these particular words within each document allows infering the tone used by the writer. We first employ a "bags of words" approach[7], but keeping trace of the order in which words appear, in order to account for negations.[8]. We then parse documents into particles and we exclude not meaningful terms (like encoding parts that are not actual words). Analyzing the *New York Times* articles we also apply a *regular expressions* based script to correctly identify and separate for each piece of news some important data like title, author(s), date of publication, and body of the text.
From the sample of 20,648 news we then aggregate the information in order to obtain IPO-level measures of media coverage and sentiment. We estimate the IPO coverage by summing the total number of articles (*N. Articles*) and the total number of words used in the articles related to each

offered).

[6]Source: Audit Bureau of Circulation

[7]We consider each article as formed as a collection of terms that, at least in a first stage, are not ordered or related each other. Textual analysis is performed just considering the frequencies with which each word appears in the text. We then consider also the order when we allow for negations, since we are interest in understanding to which word is effectively referred the negations. In this way, given a list of *negation words*, we are able to count how many times a positive or negative word is negated, and we drop these cases from the total occurrences of positive and negative words.

[8]Similarly to what Loughran and McDonald did, we consider an extra list of negation words, like *not* or *nobody* and we count the frequency they appear before positive and negative word. We finally consider the positive and negative counts excluding the cases in which positive and negative words are negated in a range of 3 words.

IPO (*N. Words*). We also created a news dummy equal to one when the IPO is covered by the NYT. We compute the tone metrics as the average of the frequency of each sentiment attributes across the news reporting information of a specific IPO. For instance, if for IPO A we find three articles having 4, 5 and 6 percent of positive words (relative to the overall number of words) respectively, the positive sentiment (*positive sentiment*) will be equal to 5 percent. Consistently, we aggregate at the IPO-level the other three sentiment attributes, i.e. negativeness (*negative sentiment*), uncertainty (*uncertainty sentiment*), litigiousness (*litigious sentiment*) and level of confidence and certainty in the way in which news are written (*modal weak* and *modal strong*).

III. Data Summary Statistics

Our sample comprises 3,061 IPOs and 20,648 articles published on the *New York Times*. Table 1 shows that each IPO is covered on average by 6,21 articles in the whole period considered (from one year before the filing to the first trade date). However, the distribution is highly skewed as a large fraction of IPOs receive no coverage[9] and in median new issues are covered only once. To the contrary, some IPO receive large attention from the NYT and this is evident looking at the high standard deviation of the number of articles and at its maximum value. This finding is also aligned with our expectation of selective coverage. As expected a large part of information is reported after the filing date. The average number of news is in fact 4.44 between the filing to the IPO date, compared to 1.59 in the year before the filing. The difference appears even larger that these numbers can evidence if we take into account the different time-span they refer to. In fact, if the pre-filing period embraces 365 days, the post-filing period varies for each issue and it is equal to 107 (81) days in average (median). Dissecting the information before and after the filing date is also important for our investigation. Loughran and McDonald (2013) investigate the effect of the S-1 tone over the IPO pricing (namely, underpricing and price revision). Since S-1s become public at the filing date, we are to verify, comparing the sentiment pre- and post-filing, whether media are somehow influenced by the tone of mandatory information required by the SEC.

INSERT TABLE 1 HERE

[9] 1,500 IPOs we detect no articles reporting information about the new issue.

Table 1 (Panel B) also reports the summary statistics at the article level. *N_Words* represents the number of words contained in an article. Unlike Liu, Sherman, and Zhang 2014a, we proxy media coverage taking into account not only the number of articles but also the room the newspaper has reserved to the IPO (number of words) We argue that *N_Words* is a more robust coverage indicator than what has been proposed in Liu et al. paper (Liu, Sherman, and Zhang 2014a). The other variables reported in Table 1 (Panel B) proxy the quality of the *coverage*, i.e. the *sentiment*. They represent the percentage of positive, negative, uncertainty related, legal related (litigious), weakening (modal weak), and strengthening (modal strong) words on the total words of each articles. Both for *positive* and *negative* sentiment, the percentages reported here are at the net of the exclusions for *negated* sentiments (we exclude from the counts the occasions in which a positive or a negative word is actually negated and its real meaning is different than expected.). Each article is on average (in median) composed by 685 (570) words, among which 0.006 % (0.005 %) are positive, 0.010 % (0.008 %) are negative, 0.006 % (0.005 %) are uncertainty related, 0.002 % (0.000 %) are litigious related, 0.003 % (0.003 %) are modal weak, and 0.004 % (0.003 %) are modal strong.

INSERT FIGURE 1 HERE

Figure 1 displays (bold line) the time-path of news reporting the (IPO-average) cumulative frequency of articles. In order to compare IPOs having different lengths of the IPO process, we standardize by the number of days occurred between filing to the first trade date. Accordingly, the time interval between -1 to 0 correspond the the year before the filing date (with no cross-variation among IPOs); the time-interval between 0 to 1 represents instead the (relative) period between filing to the first trade date. Consistent with what earlier reported, the coverage largely increases after the filing date[10] but, interestingly, the slope of the cumulative distribution of articles does not change during the post-filing period. This shape suggests that the media coverage does not exhibit peaks around the filing or immediately prior the first trade date but it shows a quite steady trend throughout the entire post-filing period.

Coverage provided before and after the filing date does not change only in terms of quantity of coverage (from now on *coverage*), but also in terms of quality of coverage (from now on *sentiment*). As shown in Table 2 articles published before and after the filing date are significantly different.

[10] 4,867 out of 20,648 articles are published before the filing date.

INSERT TABLE 2 HERE

Comparing the two panels of the table (the *Pre Filing Period* and the *Post Filing Period*) we note how both the *coverage* (number of words) and the *sentiment* indicators are more pronounced in the first subset. This evidence suggests that, despite the fewer number of article before the filing date, on average each article contains a greater number of word and a more pronounced sentiment traits. We believe the motivation is twofold. First, before the filing, only a certain type of prospective IPO firm is likely to be media covered, that is a company with a potential for attracting the interest of the newspaper readers. Generally speaking, these firms appear to be larger and more glamour. After the filing date, even smaller and less known firms begin to be media followed, but presumably the limited reader's interest produces both a decrease of the average length of each article and a more neutral writer's tone. Second, as soon as the IPO process begins, it is likely that newspapers report potential price-sensitive information more prudentially, smoothing and diluting the influence of the journalists' beliefs.

INSERT TABLE 3 HERE

Table 3 presents the descriptive statistics at the IPO level. The first panel reports the statistics for the dependents (*Underpricing* and *Price Revision*) and control variables, whilst panel B for *coverage* and *sentiment*, both at the article and at the S-1 level. In order to have the second set of metrics at the IPO level, coverage measures account for the cumulative distribution of news per IPO, while sentiment metrics reflect the average across the articles associated to each offer. *First-day returns* is the percentage different between first day of trade closing price and the offer price; *Price revision* is the ratio between the final offer price and the mid of the bookbuiling price range (minus one). The former denotes an average (median) underpricing of 27 (11) percent, while the latter shows that offer price has been in average (and in median) set 1 percent above the mid-price. The third variable (*Proceeds*) proxies for the IPO size and shows that the average firm collects nearly 100 million dollar through the offer. *VC dummy* shows that approximately 46 percent of the IPO in our sample have received a financing from venture capitalists. The *Reputation* measure is computed as in Carter and Manaster (Carter and Manaster 1990) and accounts for the market share of the leading underwriter among all IPOs in the four year before the IPO date or in the

largest period available. The remaining variables indicate that the average firm's age (*Age at trade*) when it becomes public is 15 years (8 years in median), and rougly 8 percent of our IPO firms is analyst-followed (*Analyst Dummy*). *Panel B* lists the variables obtained from textual analysis, regarding both newspapers articles and S-1 forms. The first section the Table reports the *coverage* variables, i.e. variables that explain if the IPO is followed by the New York Times (*News Dummy*), how many articles an IPO has received (*Number of Articles*), and how many words have been used by the newspaper to describe an IPO (*Number of Words*). In the second section of Panel B the sentiment variables are reported. The fewer number of observations results from the partial NYT coverage of our sample of IPOs. The last section reports the sentiment embedded in the firm-generated S-1 documents. These variables represent the degree of positiveness (as well as the degree of negativeness and uncertainty) of firms when disclosing SEC mandatory information and they are important to detect the separate effect of media coverage and firm specific disclosure (Loughran and McDonald 2013).

INSERT TABLE 4 HERE

From Table 3 we know that approximately half of our sample has received the attention of NYT. This evidence is important for two reasons. First, we still little know about what type of firm more than others is more likely to be media-covered. Intuitively, we might think that larger and more glamour firms should lead to a greater newspaper coverage but no prior investigation has clearly showed which firms' characteristics explain the media coverage. In this paper we aim at shedding some light on this regard. The IPO setting is particularly suitable to address this research question as firms, becoming publicly traded, experience a visibility shock. Second, as the sub-sample of covered IPOs is potentially highly diverse from the media-neglected firms, any inference drawn on the sample of IPOs with associated news might be suffering from a selection problem and hence being biased. This suspect is strongly confirmed by looking at Table 4, reporting descriptive statistics for the two sub-samples of covered and media-neglected IPOs. Covered IPOs have on average greater first-day returns (more than double) and larger price revisions. Although most of this gap may be driven by idiosyncratic IPO characteristics, the difference in the IPO pricing is likely to be also affected by media influence. Clearly these findings need to be corroborated with a multivariate analysis, as the two sub-samples differ in a number of other characteristics that might potentially drive the investigated phenomena: media-followed IPOs are indeed able to

collect larger proceeds, are more likely VC-backed, taken public by more reputable underwriter and followed by analysts in the aftermarket..

IV. Results

In order to understand the role of the media (*coverage* and *sentiment*) on some highly investigated IPO phenomena we undertake a three-step approach. First, we wish comprehending whether media firms tend to follow more in depth specific type of firms as well as showing whether the media sentiment is associated with some firm's characteristics. As said, the answer to this question is at the same time a novel contribution and the base to assess the existence of selection bias in inferring from a sample of covered IPOs. Second, under the hypothesis that media can enhance the information flow to uninformed investors, we investigate the marginal explanatory power of media coverage and sentiment over the level of observed underpricing and price revision. Third, we provide some robustness to our findings controlling for the informative effect of S-1 forms, for potential correlation between the media measures and the other independent variables and for the potential sample selection bias.

I. The Determinants of Media Coverage

Intuitively it is reasonable to assume that larger and more known firms receive greater attention from media when they decide to go public. However, to the best of our knowledge, we still lack of a clear evidence if this intuition is anyhow grounded and which other firms' characteristics are associated with media coverage. Furthermore, if it is somehow possible postulating casual relations between firms' characteristics and number of articles (Table 5), to foresee the same relationships to the news sentiment is certainly harder (Table 6).

INSERT TABLE 5 HERE

Table 5 clearly shows that news coverage is significantly affected by several *ex ante* IPO-firms traits, regardless the proxy used to account for the coverage itself. In model (1) we investigate coverage determinants using a *logit* regression where we use, as dependent variable, a dummy

variable taking the value of one when an IPO is covered (i.e., we collect at least one article reporting information on the offer). Consistent with our expectations, results show that the firm size (*Log-Proceeds*) positively affect the probability of being covered by the *The New York Times*, as well as being taken public by an highly reputed underwriter and being VC-backed. Quite surprisingly firm age does not seem influencing the propensity of *The New York Times* to cover the IPO firm. Model (2) and model (3) are *ordinary least-squares* linear models, where the number of articles and the overall number of words are used as dependent variables. In both case results confirm previous findings from model (1). Larger IPO firms result being more actively covered (both in terms of articles and in terms of total number of words) as well as IPO underwritten by reputable investment banks or VC-backed receive more space on the *New York Times* . Model (3) also evidences some statistical negative significance of firm's age on coverage, suggesting that, everything else constant, older issuers are less likely to attract the interest from the *The New York Times*.

INSERT TABLE 6 HERE

The next step aims to investigating the determinants of media sentiment. Table 6 shows *ordinary least-squares* linear regression models using as dependent variables the six sentiment indicators of *New York Times* articles' tone. First, we may observe how the *sentiment* is not affected by the log of proceeds. This is to some extent a unexpected but important insight as it might have been suspected that firm size may be associated with article's tone. For instance, smaller firms, due to their higher information asymmetry, could have been associated with a larger degree of news uncertainty. Furthermore, bigger firms are potentially larger advertisement spender and potentially associated with a more positive treatment from the media. Instead, none of these effects is actually in place.[11] *Underwriter reputation*, *tech dummy* and *venture capitalist backed dummy* seem having some explanatory power over of the sentiment structure, even if none of them is able to to influence the *positive tone*. Among the sentiment attributes, the one that seems to be more easily predicted based on the firm's characteristics is the level of uncertainty. Younger, high tech and VC-backed firms, for which higher is the expected information asymmetry to the market, exhibit a

[11] The lack of sizeable correlation between proceeds and sentiment allows us to use IPO size as an instrument when we control for possible selection bias.

more pronounced level of news uncertainty. The tech dummy is also importantly associated with the negative tone used by the *New York Times* . VC-backed dummy is negatively associated with the news litigiousness, to suggest that firm that are followed by a professional venture capitalist are generally less involved in litigations and other legal procedures, given the reputational and professional support provided by the venute capital firm.

II. Media Coverage and First-Day Returns

One of the core contribution of this paper is to be able to correctly understand the linkage between the informative set delivered by the media (proxied by the *New York Times*) and the level of IPO *underpricing* and *price revision*. We argue that the *New York Times* is able to convey a large set of information to the market and, specifically, to retail and not-professional investors that are not otherwise able or willing to autonomously and independently reach important and valuable information. The value of the information provided by the media comes from the costly selection activity among the virtually endless amount of available information and from potentially uncovering insider information. It is not uncommon that journalists, particularly from reputable media providers, have special and reserved informative sources that would not disclose their information if not in the anonymous way guaranteed by journalists themselves (what Dyck and Zingales call the *qui pro quo condition* (Dyck and Zingales 2003)). Information contents of *New York Times* articles is hence richer than the public available information, selected by high reputed newspapers organizations and commented by specialist and highly-trained financial journalists. Accordingly, if media can successfully enrich the information set of retail investors, it is plausible to establish a casual relationship between the quantity (*coverage*) and the quality (*sentiment*) of news and some IPO information-driven phenomena, such as the *underpricing* and *price revision*.
In terms of coverage, consistently with Liu, Sherman, and Zhang 2014a, we postulate a *visibility effect*. More covered IPOs generate more interest on retail investors, more demand for IPO shares and in turn higher likelihood of an upward revision of the offer price and larger first-day returns. Unlike Liu, Sherman, and Zhang 2014a, who only provide a rough measure of coverage (based on the sole count of the numeber of articles), we are able to test this hypothesis employing a larger and finer spectrum of metrics. We are able in fact not only to measure the numeber of published article, but also the size of the articles (as measured as total number of words) and time distribution of the articles across the different phases of the IPO process. However, we believe that the most important and novel contribution arises from the analysis of the news *sentiment* over these IPO phenomena. Retail investors are surely influenced by the media hype but the tone of

the news potentially produces a even larger effect in shaping the reader's beliefs. Accordingly, we postulate that first-day returns and the price revision should be larger when associated with disclosed positive information (*positive sentiment*) and decreasing following negative information (*negative sentiment*), while the effects following information about litigiousness (*litigious sentiment*), uncertainty-related information indicators (*uncertainty and model weak sentiment* and the backwards of *modal strong sentiment*) may potentially lead to both these directions.

INSERT TABLE 7 HERE

Table 7 shows *ordinary least-squares* linear regression models using first-day returns (i.e. the *underpricing*) as dependent variable. As a first and undoubted insight, we notice a coverage effect, present, significant, and persistent in all the model proposed. The positive coefficient is consistent with our hypothesis (H2) and economically sizeable as one standard deviation on the number of words increases the underpricing by approximately 9 percentage points. Switching our attention to the sentiment indicators, model (1) evidences a strong and significant positive relation between *positive sentiment* and the *underpricing*. Also the economic magnitude of these effects is important as one standard deviation in the *positive sentiment* increases the underpricing by 0.0473, while the same change in the *uncertainty* increases the underpricing by 0.033885.This result confirms the conjecture that a benevolent treatment from media increases the investors' interest towards the offer and in turn the demand for shares and the firs-day price. Consistent with our hypotheses, also the degree of uncertainty embedded in the news seems to play a role in determining the level of underpricing as both *uncertainty sentiment* and *modal weak sentiment* show a positive a statistical (at 10 percent level) significance. The positive relationship seems supporting hypothesis H5b over the competing hypothesis H5a, suggesting that, when media report information with a more uncertain tone, underwriters more severely underprice the offer in response to the larger returns required by the investors to be compensated for the higher uncertainty.[12]
These results hold after controlling for year and industry effects, as well as a number of variables commonly used to explain the underpricing: log of proceeds, offer price revision, tech dummy, venture capital backed dummy, reputation of the leading underwriter, presence of analyst following

[12]It could be argued that these findings may suffer from endogeneity problems or reverse causality. For instance, the underpricing could produce a more favorable media treatment and not the opposite. However, the time-structure of the data offers shelter from this potential concern. The news we collect refer to a period before the trade date when the underpricing becomes known.

the IPO, log of age of the IPO firm, average returns in the month of IPO, and percentage of IPOs above the midpoint in the considered month. The sign and the statistical significance of the control variables appear in line with previous studies.
Table 8 reports instead a number of regression models where we use price revision as dependent variable and as explanatory variables the same set of media coverage, sentiment and controls we have previously employed.

INSERT TABLE 8 HERE

Starting with the coverage effect, we report a virtually absent role of the number of words over the level of price revision.[13] The measure of media coverage shows a slightly statistical significance (at 10 percent level) in just two out of six models, precisely in those in which sentiment indicators turn to be insignificant.This result is explainable in the light of our proposed framework: we argue that media is mainly able to influence uninformed/retail agents, more than professional/institutional investors who build their beliefs using different information sources. In particular, institutional investors benefit from a direct contact with the issuing firm, even if mediated by the filter of the underwriters. Moreover, given their superior skills, professional investors only marginally rely on the information processed by the media provider but they rather base their judgments over their interpretation of the raw available information. Having this consideration in mind, it is to little extent surprising that phenomena greatly driven by retail demand , such as the first-day returns, are more affected by media-effects than other dimensions, like price revision, more influenced by the action of institutional investors.
Switching to the analysis of the tone attributes, Table 8 presents similar findings to what showed for the determination of the underpricing. Price revision appears to be positively related to (a) the *positive sentiment* and (b) the level of *uncertainty*, supporting hypotheses H3 and H6b, respectively. The level of statistical significance is for the *positive sentiment* and for the two proxies of uncertainty, *uncertainty* and *modal weak*, at the 5 percent level. Also the economic magnitude of these effects is somehow similar as one standard deviation in the *positive sentiment* increases the price revision by 0.0067, while the same change in the *uncertainty* produces a 0.0064 upward revision.[14] The *positive sentiment* seems producing an effect throughout the entire IPO process, being able to affect both price revision and first-day returns, likely due to a demand-driven outcome. Similarly, the

[13]The other proxy of media coverage, the news dummy and the number of articles, produce similar effects.
[14]The same effect for *modal weak* is 0.0068.

media uncertain sentiment causes higher observed underpricing (we argue due to an increase of investors' required returns) and larger price revision (probably produced by the higher difficulties in fairly setting the offer price). As for previous regressions, these results hold after controlling for IPO size, firm's riskiness, underwriter reputation, IPO market conditions, industry and year effects.

V. Robustness Checks

Previous findings have shown that media *coverage* and *sentiment* are able to produce effects over the observed level of *underpricing* as well as the *price revision*. In this section we aim at providing a number of robustness checks in order to show that the presented insights are not driven by unconsidered issues. The first control we offer concerns the informative role of S-1 forms, in order to verify that the *New York Times* is delivering a brand new informative set, and it is not just translating and publicly reporting technical information already disclosed by the IPO firm itself. The second concern we have relates to the correlation between the sentiment indicators and the firm and IPO's characteristics. As shown in this paper, media sentiment is vaguely associated with some specific type of firms. In order to control for it, we repeat the analysis using as sentiment indicators the residuals from the models where we regress the sentiment measures over the pre-market IPO firm characteristics. Finally, applying an Heckman selection model (Heckman 1979), we address the concern of possible selection bias due to the fact that just for half of the sample is possible to extract information about related-news tone.

I. Controlling for S-1 informative contents

The first robustness check aims at dispelling the suspect that articles from the *New York Times* do not provide incremental valuable information to the market, as what reported in the media may be simply greatly anticipated by the legal documents that issuing firms are required to fill in (S-1 forms). We present robust and diffuse evidence on our data this concern is unjustified. The first evidence comes from analysing the sign of media sentiment coefficients relative to the corresponding results computed using the S-1 files. Loughran and McDonald (2003) find that negative S-1 sentiment increases the underpricing, while we find the same phenomeon is caused by a positive media sentiment. This difference suggests that media and S-1 sentiments are not correlated measures as we might think. In other words, it seems dispelled the doubt that the way the underwriter communicates to the SEC trhough the filings is able to influence the tone that journalists use when they report IPO information.

INSERT TABLE 9 HERE

Furthermore, for sake of completeness, we collect the same data as in Loughran and McDonald (2013) to verify that media sentiment produces an effect over and above what S-1 sentiment can offer. Table 9 confirms our predictions as *New York Times* articles positive sentiment maintains its positive and significant effect also after controlling for S-1 sentiment. The fact that a *positive sentiment* (on the daily newspaper) has a same sign effect than a negative one (on the S-1 form) may be explained within the framework of market segmentation. Investors who have access to the S-1 forms differ from those who read IPO information on the *New York Times* and, most important, different is the timing at which the two set of information are disclosed. Unlike the *positive sentiment*, the proxies of *uncertainty* lose their statistical significance at least at the conventional levels. *Uncertainty* and *modal weak* keep their positive coefficients but they are not longer significant, probably due to the lower number of observations.[15]

INSERT TABLE 10 HERE

Table 10 reports the same controls for price revision and we notice a similar patter to what previously discussed. The effect of positive *New York Times* articles sentiment on price revision is still present and robust. However, as for the underpricing, the media uncertainty loses statistical power, even though coefficients are in line with what earlier presented.

II. Controlling using tone residuals

The first part of our analyses show that the media coverage and partially the media sentiment is *ex-ante* explained by some firm and market characteristics. Even though, in explaining the level of underpricing and price revision, we do control for such characteristics, a potential concern may nevertheless arise due to the correlation among our explanatory variables. Therefore, we compute

[15]These regressions show a poorer sample size due to the drop out caused by the merge of our sample with the number of companies for which we find information on S1.

tone residuals after regressing each sentiment indicator against pre-IPO and market characteristics. We then regress both first-day returns and price revision using these tone residuals as a proxy of unpredictable content of the news, also controlling for news coverage. As we show in Table 11 and in Table 12 positive sentiment remains a good and significant predictor for both first-day returns and price revision.

INSERT TABLE 11 HERE

INSERT TABLE 12 HERE

As mentioned before the positive and significant coefficient associated with the residuals of positive *sentiment* shows us how the positive *sentiment* embedded in the *New York Times* articles effectively brings some extra informational content that cannot be explained by ex-ante IPO firm characteristics and it is furthermore effective in shaping the dynamics of the IPO process. However, unlike positive *sentiment*, the level of uncertainty appears to be less important in the explanation of these phenomena. To some extent this result is not completely unexpected as, among the sentiment indicators, *uncertainty* proxies were the most correlated with firms characteristics.

III. Controlling for endogeneity issue

We have previously suggested that our results might be potentially driven by a sample selection bias. Since sentiment indicators can be only computed if media *coverage* is positive[16], we want to control if the reported association between sentiment and IPO phenomena is not caused by a non-random sample selection. In order to do carry out such a check, we apply an Heckman two-stage model (Heckman 1979) where in the first step we run a probit regression using the News dummy as dependent variable, the same set of control variables but taking the *LogProceed* as instrument. In the second step we use the predicted variables from the first stage regression to estimate the corrected models for *underpricing* and *price revision*. Consistent with the other two robustness checks, Table 13 shows how the relation between positive sentiment and *underpricing* and *price revision* patterns firmly holds, whilst we do not longer evidence a statistically robust association with the level of *uncertainty*.

[16]In other words, only if we find any article reporting information on that specific IPO.

INSERT TABLE 13 HERE

INSERT TABLE 14 HERE

In Table 14 we observe the same phenomenon for price revision. Positive sentiment is the only one still significant and useful in explaining the dependent variable, while uncertainty and negative indicators are not producing any significant appreciable effects.

VI. Conclusions

In this paper we show how influential and effective could be the media on the financial markets. In order to test this relation we choose the IPO process empirical setting. In this framework the information plays a pivotal role in shaping and conditioning the whole process. In the IPO process a large amount of private information is gradually released to the public, together with a well define set of ownership and administrative rights on the firms (i.e. stock shares). In this process there are some individuals and groups that are able to reach and to process the information before others. These last ones gradually start to disclose pieces of private information to the public, transforming private information into public information. We argue in this last part of the process the pivotal role to involve the large part of not-professional investors is played by the media industry. In this study we aim at investigating the effect of media information over the IPO process, analysing the coverage and the tone of the articles published in the *New York Times*. We provide evidence that *New York Times* articles reporting information on an incoming IPO are able to influence the IPO ina double fashion. First, the simple fact the the *New York Times* dedicates spaces to a prospective publicly traded firm, generates by itself a *visibility effect* and in turns an increase in the IPO *first-day returns* and *price revision*. Second, we show how articles with more positive sentiment influence retail investors, the demand for IPO shares and a further increase in the first-day returns patterns. After controlling for a number of compounding factors, including S-1 informational contents, pre-market IPO firm characteristics, and possible selection bias issues, we robustly show that media coverage and positive sentiment are important drivers in determining the level of IPO underpricing and price revision.

References

Aggarwal, Reena, Nagpurnanand R Prabhala, and Manju Puri (2002). "Institutional allocation in initial public offerings: Empirical evidence". In: *The Journal of Finance* 57.3, pp. 1421–1442.

Akerlof, George A (1970). "The market for" lemons": Quality uncertainty and the market mechanism". In: *The quarterly journal of economics*, pp. 488–500.

Allen, Franklin and Gerald R Faulhaber (1989). "Signalling by underpricing in the IPO market". In: *Journal of financial Economics* 23.2, pp. 303–323.

Barber, Brad M and Terrance Odean (2008). "All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors". In: *Review of Financial Studies* 21.2, pp. 785–818.

Beatty, Randolph P and Jay R Ritter (1986). "Investment banking, reputation, and the underpricing of initial public offerings". In: *Journal of financial economics* 15.1, pp. 213–232.

Benveniste, Lawrence M and Paul A Spindt (1989). "How investment bankers determine the offer price and allocation of new issues". In: *Journal of Financial Economics* 24.2, pp. 343–361.

Besley, Timothy and Andrea Prat (2006). "Handcuffs for the grabbing hand? Media capture and government accountability". In: *The American Economic Review*, pp. 720–736.

Bhattacharya, Utpal et al. (2009). "The role of the media in the internet IPO bubble". In: *Journal of Financial and Quantitative Analysis* 44.03, pp. 657–682.

Booth, James R and Richard L Smith II (1986). "Capital raising, underwriting and the certification hypothesis". In: *Journal of Financial Economics* 15.1, pp. 261–281.

Carter, Richard and Steven Manaster (1990). "Initial public offerings and underwriter reputation". In: *The Journal of Finance* 45.4, pp. 1045–1067.

Chan, Wesley S (2003). "Stock price reaction to news and no-news: drift and reversal after headlines". In: *Journal of Financial Economics* 70.2, pp. 223–260.

Cornelli, Francesca and David Goldreich (2001). "Bookbuilding and strategic allocation". In: *The Journal of Finance* 56.6, pp. 2337–2369.

DellaVigna, Stefano and Ethan Daniel Kaplan (2007). "The Fox News Effect: Media Bias and Voting". In: *The Quarterly Journal of Economics* 122.3, pp. 1187–1234.

Djankov, Simeon et al. (2003). "Who owns the media?" In: *Journal of Law and Economics* 46.

Dougal, Casey et al. (2012). "Journalists and the stock market". In: *Review of Financial Studies* 25.3, pp. 639–679.

Dyck, Alexander, Adair Morse, and Luigi Zingales (2010). "Who blows the whistle on corporate fraud?" In: *The Journal of Finance* 65.6, pp. 2213–2253.

Dyck, Alexander, David Moss, and Luigi Zingales (2013). "Media versus Special Interests". In: *Journal of Law and Economics* 56.3, pp. 521–553.

Dyck, Alexander, Natalya Volchkova, and Luigi Zingales (2008). "The corporate governance role of the media: Evidence from Russia". In: *The Journal of Finance* 63.3, pp. 1093–1135.

Dyck, Alexander and Luigi Zingales (2003). "The Media and Asset Prices". In:

Fang, Lily and Joel Peress (2009). "Media Coverage and the Cross-section of Stock Returns". In: *The Journal of Finance* 64.5, pp. 2023–2052.

Garcia, Diego (2013). "Sentiment during recessions". In: *The Journal of Finance* 68.3, pp. 1267–1300.

Gentzkow, Matthew, Edward L Glaeser, and Claudia Goldin (2006). "The rise of the fourth estate. How newspapers became informative and why it mattered". In: *Corruption and Reform: Lessons from America's Economic History*. University of Chicago Press, pp. 187–230.

Hanley, Kathleen Weiss (1993). "The underpricing of initial public offerings and the partial adjustment phenomenon". In: *Journal of financial economics* 34.2, pp. 231–250.

Hanley, Kathleen Weiss and William J Wilhelm Jr (1995). "Evidence on the strategic allocation of initial public offerings". In: *Journal of financial economics* 37.2, pp. 239–257.

Heckman, James J (1979). "Sample selection bias as a specification error". In: *Econometrica: Journal of the econometric society*, pp. 153–161.

Huberman, Gur and Tomer Regev (2001). "Contagious speculation and a cure for cancer: A nonevent that made stock prices soar". In: *The Journal of Finance* 56.1, pp. 387–396.

Hughes, Patricia J (1986). "Signalling by direct disclosure under asymmetric information". In: *Journal of Accounting and Economics* 8.2, pp. 119–142.

Ibbotson, Roger G (1975). "Price performance of common stock new issues". In: *Journal of financial economics* 2.3, pp. 235–272.

Lee, Peggy M and Sunil Wahal (2004). "Grandstanding, certification and the underpricing of venture capital backed IPOs". In: *Journal of Financial Economics* 73.2, pp. 375–407.

Liu, Laura Xiaolei, Ann E Sherman, and Yong Zhang (2014a). "An attention model of IPO underpricing, with evidence on media coverage". In: *unpublished*.

— (2014b). "The Long-Run Role of the Media: Evidence from Initial Public Offerings". In: *Management Science*.

Ljungqvist, Alexander P, Tim Jenkinson, and William J Wilhelm Jr (2003). "Global integration in primary equity markets: The role of US banks and US investors". In: *Review of Financial studies* 16.1, pp. 63–99.

Loughran, Tim and Bill McDonald (2011). "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks". In: *The Journal of Finance* 66.1, pp. 35–65.

Loughran, Tim and Bill McDonald (2013). "IPO first-day returns, offer price revisions, volatility, and form S-1 language". In: *Journal of Financial Economics* 109.2, pp. 307–326.

Loughran, Tim and Jay R Ritter (2002). "Why don't issuers get upset about leaving money on the table in IPOs?" In: *Review of Financial Studies* 15.2, pp. 413–444.

— (2004). "Why has IPO underpricing changed over time?" In: *Financial management* 33.3.

Megginson, William L and Kathleen A Weiss (1991). "Venture capitalist certification in initial public offerings". In: *The Journal of Finance* 46.3, pp. 879–903.

Miller, Gregory S (2006). "The press as a watchdog for accounting fraud". In: *Journal of Accounting Research* 44.5, pp. 1001–1033.

Moy, Patricia and Michael Pfau (2000). *With malice toward all?: The media and public confidence in democratic institutions*. Greenwood Publishing Group.

Mullainathan, Sendhil and Andrei Shleifer (2005). "The market for news". In: *American Economic Review*, pp. 1031–1053.

Pollock, Timothy G and Violina P Rindova (2003). "Media legitimation effects in the market for initial public offerings". In: *Academy of Management Journal* 46.5, pp. 631–642.

Pollock, Timothy G, Violina P Rindova, and Patrick G Maggitti (2008). "Market watch: Information and availability cascades among the media and investors in the US IPO market". In: *Academy of Management Journal* 51.2, pp. 335–358.

Ritter, Jay R (1984). "The" hot issue" market of 1980". In: *Journal of Business*, pp. 215–240.

Rock, Kevin (1986). "Why new issues are underpriced". In: *Journal of financial economics* 15.1, pp. 187–212.

Ross, Stephen A (1977). "The determination of financial structure: the incentive-signalling approach". In: *The Bell Journal of Economics*, pp. 23–40.

Sherman, Ann E and Sheridan Titman (2002). "Building the IPO order book: underpricing and participation limits with costly information". In: *Journal of Financial Economics* 65.1, pp. 3–29.

Stulz, René M (1999). "Globalization, corporate finance, and the cost of capital". In: *Journal of applied corporate finance* 12.3, pp. 8–25.

Tetlock, Paul C (2007). "Giving content to investor sentiment: The role of media in the stock market". In: *The Journal of Finance* 62.3, pp. 1139–1168.

— (2011). "All the news that's fit to reprint: Do investors react to stale information?" In: *Review of Financial Studies* 24.5, pp. 1481–1512.

VII. Tables and Figures

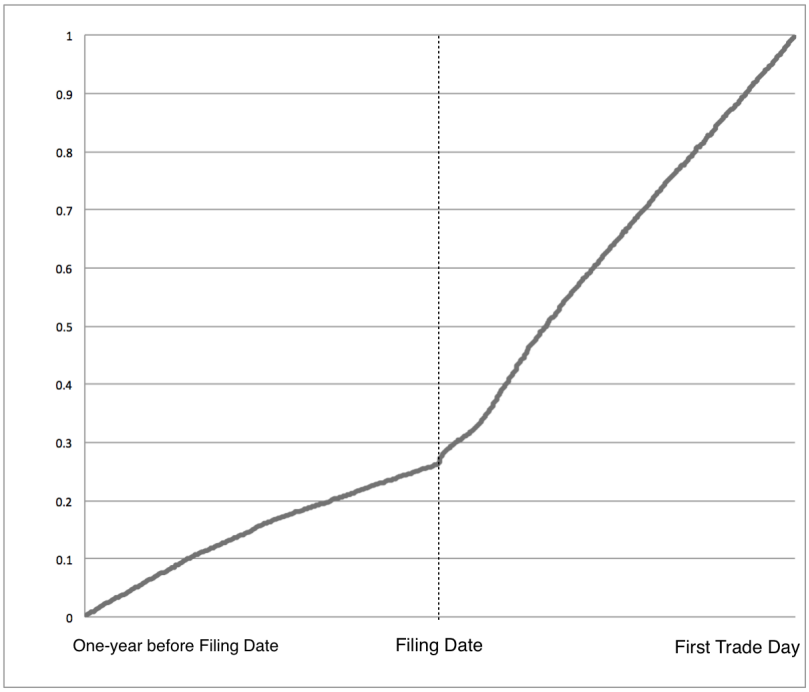

Figure 1: *IPOs coverage per period. In this graph the distribution of the coverage is reported. We standardized the period before filing date and between filing date and first day of trade date. In the picture a cusp point is evident in correspondence with the filing date which becomes a sort of watershed in the media attention patterns.The effect of officially enter the IPO process by filing the required documentation at the filing date is evident and it changes dramaticlly the slope of the line.*

Table 1: *Summary statistics. The first panel (Panel A) reports the number of articles. In this panel the number of article published in each period is reported. Whole sample refers to the total number of articles published in the whole period, pre-filing refers to number of articles published from one year before the filing date to the filing date itself, and post period refers to the period after the filing date until the first day of trade. In the second panel (Panel B) some features of the articles are reported. In this panel the data about articles analyzed are reported for the whole sample. All the articles are published on the printed version of the New York Times and they involve specifically the firm going to become public. For each period N_Words is the total number of words used in each article and the other variables are the sentiment indicators we are using. They represent respectively: the percentage of positive words, the percentage of negative words, the percentage of uncertainty-related word, and the percentage of legal words, the percentage of modal weak words, the percentage of modal strong words. Both the percentages of positive and negative word are reported at the net of negations (i.e. the cases in which positive/negative word are actually negated assuming so a different meaning).*

Variable	Obs	Mean	Std. Dev.	Min	Max	Median
Panel A	*Number of articles per period*					
whole sample	3061	6.21	39.06	0	1094	1
pre-filing	3061	1.59	14.11	0	453	0
post-filing	3061	4.44	25.41	0	650	0
Panel B	*Articles characteristics*					
N_Words	20648	685	662.2	0	16432	570.0
positive sentiment (%)	20648	0.006	0.005	0	0.087	0.005
negative sentiment (%)	20648	0.010	0.010	0	0.174	0.008
uncertain sentiment (%)	20648	0.006	0.005	0	0.085	0.005
litigious sentiment (%)	20648	0.002	0.005	0	0.130	0.000
modal weak sentiment (%)	20648	0.003	0.004	0	0.071	0.003
modal strong sentiment (%)	20648	0.004	0.005	0	0.125	0.003

Table 2: *Summary Statistic and periods comparison. In this table* coverage *and* sentiment *variables are reported at the article level, distinguishing for the period. The* Pre Filing Period *is the period between one-year before the* filing date *and the* filing date*. The* Post Filing *period is instead the period after the* filing date *and before the* first day of trade*. All the articles are published on the printed version of the New York Times and they involve specifically the firm going to become public. For each period N_Words is the total number of words used in each article and the other variables are the sentiment indicators we are using. They represent respectively: the percentage of positive words, the percentage of negative words, the percentage of uncertainty-related word, the percentage of legal words, the percentage of modal weak words, and the percentage of modal strong words.Both the percentages of positive and negative word are reported at the net of negations (i.e. the cases in which positive/negative word are actually negated assuming so a different meaning).*

Sentiment patterns across pre and post filing periods						
	Obs	Mean	St Dev	Min	Max	Median
Pre Filing Period						
N_Words	4915	769.6	606.7	18	6025	689
positive sentiment (%)	4915	0.006	0.005	0	0.067	0.006
negative sentiment (%)	4915	0.010	0.010	0	0.111	0.008
uncertain sentiment (%)	4915	0.006	0.005	0	0.085	0.005
litigious sentiment (%)	4915	0.002	0.005	0	0.069	0.000
modal weak sentiment (%)	4915	0.004	0.004	0	0.056	0.003
modal strong sentiment (%)	4915	0.005	0.005	0	0.083	0.003
Post Filing Period						
N_Words	13711	724.4	662.3	7	10479	601
positive sentiment (%)	13711	0.006	0.005	0	0.087	0.005
negative sentiment (%)	13711	0.010	0.010	0	0.174	0.008
uncertain sentiment (%)	13711	0.006	0.005	0	0.077	0.005
litigious sentiment (%)	13711	0.002	0.005	0	0.130	0.000
modal weak sentiment (%)	13711	0.003	0.004	0	0.063	0.003
modal strong sentiment (%)	13711	0.004	0.006	0	0.125	0.003

Table 3: *Summary statistics for IPO sample, 1995 - 2013.The sample includes 3061 U.S. IPOs with an offer price at least $5 per share, excluding financial firms, REITs, and closed-end funds. The sample is obtained from Thompson Financial Securities Data.In Panel A all variable all variables related to IPO are reported. In Panel B we report instead all variable related to media coverage and mandatory disclosure by the firm (S-1 forms).*

Variable	Obs	Mean	Std. Dev.	Min	Max	P50
Panel A: IPO characteristics						
First-day Returns	2,864	0.27	0.55	-0.88	6.97	0.11
Proceeds	3,023	99.70	123.27	3.45	933.80	59.48
Price Revision	3,061	0.01	0.13	-0.37	0.33	0.00
VC Dummy	3,061	0.46	0.49	0.00	1.00	0.00
Reputation	3,061	0.05	0.05	0.00	0.23	0.03
Institutional Investors (number of)	3,061	27.99	26.10	0.00	421.00	23.00
Market Valuation	2,123	3.24	3.29	0.86	108.48	2.73
Market Valuation (*Ind-Adjusted*)	2,123	0.47	3.21	-2.56	105.98	0.00
Age at Trade	2,921	15.02	21.01	0.00	165.00	8.00
Analyst Dummy	2,522	0.23	0.42	0.00	1.00	0.00
Avg First Day Return	3059	26.70	26.10	-19.9	163.20	17.70
Rate IPOs Above Midpoint	3059	46.83	21.02	0.00	100.00	48.00
Panel B: Coverage and sentiment						
News Dummy	3,061	0.51	0.50	0	1	1
Number of Articles	3,061	6.18	39.06	0	1,094	1
Number of Words	3,061	4,561.93	30,342.68	0	791,770	82
Positive Words (% *points*)	1,576	0.48	0.39	0	3.79	0.47
Negative Words (% *points*)	1,576	0.88	0.68	0	5.88	0.80
Uncertainty Words (% *points*)	1,576	0.46	0.36	0	2.70	0.44
Litigious Words (% *points*)	1,576	0.15	0.25	0	3.05	0.05
Modal Weak Words (% *points*)	1,576	0.26	0.24	0	2.06	0.24
Modal Strong Words (% *points*)	1,576	0.39	0.38	0	3.01	0.33
S-1 Positive Words (% *points*)	2307	0.3	0.2	0	1.3	0.3
S-1 Negative Words (% *points*)	2307	0.8	0.4	0	2.0	0.7
S-1 Modal Weak Words (% *points*)	2307	0.4	0.2	0	1.0	0.4
S-1 Modal Strong Words (% *points*)	2307	0.3	0.1	0	0.9	0.2
S-1 Litigious Words (% *points*)	2307	1.2	0.8	0	3.3	1.0
S1 Uncertainty Words (% *points*)	2307	0.6	0.3	0	1.5	0.6

Table 4: *Main variable summary statistics if covered by New York Times. We report all the most important variables related to the IPO process putting in evidence the diffrences they experience if covered or not by the NYT. Panel A reports summary statistics for IPO issues covered by NYT, Panel B reports the same for IPO issues that are not covered. Covered IPOs are slightly bigger, they experience larger first-day returns as well as larger price revisions. They are also associated with more reputed underwriters, with a larger proportion of venture capitalists and analysts, and a larger number of institutional investors. Furthermore covered IPOs receive better market valuation and they are, on average, related to older firms.*

Panel A: IPO covered by The New York Times						
	Obs	*Mean*	*Min*	*Max*	*Median*	*St. Dev*
First-Day Returns	1466	0.38	-0.87	6.98	0.16	0.68
Proceeds	1537	118.37	5.16	933.8	71.88	140.68
Price Revision	1575	0.03	-0.37	0.33	0.05	0.13
VC Dummy	1575	0.49	0	1	0	0.50
Tech Dummy	1575	0.49	0	1	0	0.50
Reputation	1575	0.07	0	0.34	0.06	0.06
Institutional Investors (n)	1575	33.34	0	421	28	30.43
Market Valuation (IndAdj)	1085	0.70	-2.42	105.98	0.04	4.13
Age	1516	15.86	0	158	7	22.55
Analyst Dummy	1365	0.29	0	1	0	0.45
Panel B: IPO not-covered by The New York Times						
	Obs	*Mean*	*Min*	*Max*	*Median*	*St. Dev*
First-Day Returns	1397	0.17	-0.88	3.57	0-08	0.33
Proceeds	1485	80.39	3.45	870	47.5	98.63
Price Revision	1485	-0.01	-0.37	0.33	0	0.13
VC Dummy	1485	0.45	0	1	0	0.49
Tech Dummy	1485	0.52	0	1	1	0.50
Reputation	1485	0.05	0.00	0.34	0.02	0.06
Institutional Investors (n)	1485	22.33	0	117	18	18.96
Market Valuation (IndAdj)	1037	0.23	-2.56	21.70	-0.04	1.77
Age	1404	14.12	0	165	8	19.18
Analyst Dummy	1157	0.17	0	1	0	0.38

Table 5: *Regression table. Explaining the coverage received by Initial Public Offering in terms of ex ante characteristics of the firm going public. In this set of regressions the dependent variables are all measuring the coverage the IPOs receive, as measured as a NYT related* coverage dummy *(Column 1), as total number of articles published on the NYT about each IPO (Column 2), and as total space the NYT devotes to each IPO, computed as total number of words (column 3). The table shows how the NYT coverage of an IPO is positively associated with the IPO dimension (*LogProceeds*), the underwriters reputation, and the presence a venture capitalist involved in the process. The age of the issuing firm seems to be significant only to the extent of total space the NYT devotes with a marginal negative effect on it. In all the regressions industry and year fixed effects are considered.*

	(1) Coverage Dummy	(2) N. articles	(3) N. Words
LogProceeds	0.929***	3.862***	2779.1***
	(13.10)	(6.45)	(5.25)
Reputation	3.301***	32.13***	28476.1***
	(3.79)	(3.97)	(3.98)
VC Dummy	0.242**	2.394***	2376.7***
	(2.47)	(2.62)	(2.94)
Tech Dummy	-0.138	-1.802	-1397.0
	(-1.04)	(-1.46)	(-1.28)
LogAge	-0.0766	-0.606	-898.8**
	(-1.49)	(-1.28)	(-2.15)
Constant	-3.250***	-14.02	-9233.0
	(-8.95)	(-0.67)	(-0.50)
Year Dummies	*Yes*	*Yes*	*Yes*
Industry Dummies	*Yes*	*Yes*	*Yes*
Observations	2867	2879	2879

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: *Regression table: sentiment determinants. Explaining the sentiment received by Initial Public Offering in terms of ex ante characteristics of the firm going public.In this set of regressions the dependent variables are all measuring the quality of the coverage the IPOs receive, as measured as degree of positiveness, negativeness, litigious, uncertainty and certainty in the way in which the articles are written (modal weak and modal strong.). Each sentiment (corresponding to each column) has its own pattern when associated with the explanatory variables. To be noticed how none of the sentiment is significantly associated with the dimension of the IPO (here represented by the LogProceeds).*

	(1)	(2)	(3)	(4)	(5)	(6)
Sentiment:	positive	negative	litigious	uncertainty	modal weak	modal strong
LogProceeds	0.00001	-0.00023	0.00014	-0.00012	-0.00006	-0.00013
	(0.03)	(-0.83)	(1.28)	(-0.83)	(-0.64)	(-0.82)
Reputation	0.0032	0.0080**	0.0006	0.0003	0.0007	-0.0001
	(1.52)	(2.20)	(0.41)	(0.13)	(0.54)	(-0.03)
VC Dummy	0.0003	-0.0003	-0.0004**	0.0004*	0.0003**	0.0006***
	(1.11)	(-0.66)	(-2.23)	(1.95)	(2.35)	(2.64)
Tech Dummy	0.0004	0.0015***	-0.0001	0.0006**	0.0004**	-0.0005
	(1.19)	(2.73)	(-0.38)	(2.18)	(2.34)	(-1.62)
LogAge	-0.0001	-0.0002	-0.0001	-0.0003***	-0.0001	-0.0001
	(-1.01)	(-0.81)	(-0.81)	(-2.62)	(-1.32)	(-0.50)
Constant	0.0039	0.0102	0.0006	0.0075**	0.0045**	0.0045
	(1.03)	(1.53)	(0.25)	(2.17)	(1.96)	(1.20)
Year Dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Industry Dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Observations	1474	1474	1474	1474	1474	1474

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: *Regression table: First day returns. This table shows the results for a set of ordinary least squares regressions explaining the underpricing phenomenon. The most important result is that evident from the first column: the positive sentiment influences positively e significantly the first-day returns, also after controlling for the usual set of underpricing explanatory variable, industry and year fixed effect, and also the coverage effect as measures as number of words.*

	(1)	(2)	(3)	(4)	(5)	(6)
	First-day returns	First-day returns	First-day returns	First-day returns	First-day returns	First-day returns
N. words	0.000003***	0.000003***	0.000003***	0.000003***	0.000003***	0.000003***
	(3.33)	(3.49)	(3.37)	(3.52)	(3.33)	(3.54)
positive sentiment	12.13***					
	(2.79)					
negative sentiment		2.936				
		(1.13)				
uncertainty sentiment			9.425*			
			(1.92)			
litigious sentiment				-1.720		
				(-0.25)		
modalweak sentiment					14.36*	
					(1.92)	
modalstrong sentiment						-7.092
						(-1.60)
Log of Proceeds	0.0170	0.0139	0.0163	0.0138	0.0159	0.0116
	(0.60)	(0.49)	(0.58)	(0.49)	(0.56)	(0.41)
Offer price revision	1.176***	1.189***	1.176***	1.189***	1.174***	1.185***
	(7.99)	(8.06)	(7.98)	(8.06)	(7.96)	(8.04)
Tech Dummy	0.166***	0.166***	0.165***	0.172***	0.165***	0.168***
	(3.24)	(3.22)	(3.21)	(3.35)	(3.21)	(3.27)
VC Dummy	0.0640	0.0635	0.0610	0.0623	0.0605	0.0672*
	(1.62)	(1.60)	(1.54)	(1.57)	(1.53)	(1.70)
Reputation	1.031***	1.034***	1.065***	1.064***	1.055***	1.060***
	(2.78)	(2.77)	(2.87)	(2.86)	(2.84)	(2.85)
Analyst Dummy	0.113***	0.116***	0.114***	0.116***	0.116***	0.115***
	(2.82)	(2.88)	(2.84)	(2.87)	(2.88)	(2.85)
Age	-0.00178*	-0.00188**	-0.00178*	-0.00188**	-0.00182*	-0.00183**
	(-1.92)	(-2.01)	(-1.90)	(-2.01)	(-1.95)	(-1.97)
Avg_Return	0.0119***	0.0118***	0.0117***	0.0118***	0.0118***	0.0118***
	(9.85)	(9.69)	(9.67)	(9.70)	(9.74)	(9.74)
%IPO above mid	-0.00436***	-0.00415***	-0.00407***	-0.00413***	-0.00414***	-0.00417***
	(-3.64)	(-3.46)	(-3.39)	(-3.44)	(-3.45)	(-3.48)
Constant	0.0682	0.0971	0.0440	0.123	0.0528	0.158
	(0.11)	(0.16)	(0.07)	(0.21)	(0.09)	(0.26)
Year Dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Industry Dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Observations	1271	1271	1271	1271	1271	1271

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8: *Regression table: Price Revision. This table shows the results for a set of ordinary least squares regressions explaining the price revision phenomenon. The most important result is that evident from the first column: the positive sentiment influences positively e significantly the price revision, also after controlling for the usual set of underpricing explanatory variable, industry and year fixed effect, and also the coverage effect as measures as number of words.*

	(1)	(2)	(3)	(4)	(5)	(6)
	price revision	price revision	price revision	price revision	price revision	price revision
N. words	0.0000002	0.0000002	0.0000002	0.0000002*	0.0000002	0.0000002*
	(1.54)	(1.64)	(1.52)	(1.65)	(1.50)	(1.66)
positive sentiment	1.725**					
	(2.17)					
negative sentiment		0.223				
		(0.49)				
uncertainty sentiment			1.773**			
			(1.99)			
litigious sentiment				0.0970		
				(0.08)		
modal weak sentiment					2.844**	
					(2.09)	
modal strong sentiment						-0.315
						(-0.38)
Log of Proceeds	0.0395***	0.0396***	0.0398***	0.0395***	0.0398***	0.0395***
	(8.31)	(8.30)	(8.36)	(8.29)	(8.36)	(8.28)
Tech Dummy	0.0130*	0.0131*	0.0123	0.0134*	0.0121	0.0133*
	(1.65)	(1.66)	(1.56)	(1.70)	(1.53)	(1.69)
VC Dummy	0.0186***	0.0192***	0.0184***	0.0192***	0.0182**	0.0193***
	(2.61)	(2.69)	(2.58)	(2.68)	(2.54)	(2.70)
Reputation	0.173***	0.174***	0.177***	0.176***	0.175***	0.175***
	(2.76)	(2.77)	(2.82)	(2.80)	(2.80)	(2.80)
Age	-0.000516***	-0.000527***	-0.000510***	-0.000527***	-0.000515***	-0.000526***
	(-3.22)	(-3.28)	(-3.17)	(-3.28)	(-3.21)	(-3.28)
AvgReturn	0.000282	0.000265	0.000263	0.000268	0.000273	0.000270
	(1.26)	(1.18)	(1.17)	(1.19)	(1.22)	(1.20)
%IPO above mid	0.00170***	0.00174***	0.00174***	0.00174***	0.00173***	0.00174***
	(8.16)	(8.34)	(8.36)	(8.33)	(8.31)	(8.33)
Constant	-0.273***	-0.271***	-0.277***	-0.269***	-0.274***	-0.268***
	(-9.45)	(-9.28)	(-9.51)	(-9.31)	(-9.46)	(-9.22)
Year Dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Industry Dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Observations	1474	1474	1474	1474	1474	1474

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9: *Regression table: First day returns controlling for the informational content of S-1 files. This table shows the results for a set of ordinary least squares regressions explaining the underpricing phenomenon adding also the informational effect of S-1 files. The most important results hold: the positive sentiment influences positively e significantly the first-day returns, also after controlling for the usual set of underpricing explanatory variable, industry and year fixed effect, and also the coverage effect as measures as number of words. Furthermore the S-1 negative sentiment is negatively associated with the first-day returns, consistently with our explanation of first-day returns generated by an increase on the demand side.*

	(1)	(2)	(3)	(4)	(5)	(6)
	First-day returns	First-day returns	First-day returns	First-day returns	First-day returns	First-day returns
N. words	0.000001	0.000002	0.000001	0.000002	0.000001	0.000002
	(0.98)	(1.12)	(1.06)	(1.14)	(1.01)	(1.13)
positive sentiment	16.43**					
	(2.43)					
negative sentiment		4.620				
		(1.08)				
uncertainty sentiment			9.166			
			(1.20)			
litigious sentiment				0.651		
				(0.07)		
modal weak sentiment					15.78	
					(1.35)	
modal strong sentiment						-8.553
						(-1.25)
S1 negative sentiment	19.55**	19.95**	19.23**	20.16**	19.58**	20.10**
	(2.36)	(2.40)	(2.30)	(2.42)	(2.35)	(2.42)
Constant	-0.200	-0.190	-0.203	-0.141	-0.199	-0.102
	(-0.28)	(-0.26)	(-0.28)	(-0.20)	(-0.28)	(-0.14)
Usual Checks	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Year Dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Industry Dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Observations	697	697	697	697	697	697

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 10: *Regression table: Price Revision controlling for the informational content of S-1 files. This table shows the results for a set of ordinary least squares regressions explaining the price revision phenomenon adding also the informational effect of S-1 files. The most important results hold: the positive sentiment influences positively e significantly the first-day returns, also after controlling for the usual set of underpricing explanatory variable, industry and year fixed effect, and also the coverage effect as measures as number of words. Furthermore the S-1 negative sentiment is negatively associated with the price revision, consistently with our expectations.*

	(1) price revision	(2) price revision	(3) price revision	(4) price revision	(5) price revision	(6) price revision
N. words	0.0000002	0.0000002	0.0000002	0.0000002	0.0000002	0.0000002
	(1.14)	(1.24)	(1.18)	(1.23)	(1.13)	(1.24)
positive sentiment	2.163*					
	(1.94)					
negative sentiment		-0.681				
		(-0.98)				
uncertainty sentiment			1.107			
			(0.89)			
litigious sentiment				-2.197		
				(-1.42)		
modal weak sentiment					2.867	
					(1.51)	
modal strong sentiment						-0.315
						(-0.28)
S1 negative sentiment	0.799	0.907	0.786	0.820	0.816	0.871
	(0.59)	(0.66)	(0.57)	(0.60)	(0.60)	(0.64)
Constant	-0.251***	-0.238***	-0.253***	-0.245***	-0.252***	-0.247***
	(-4.72)	(-4.40)	(-4.72)	(-4.60)	(-4.73)	(-4.63)
Usual Checks	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Year Dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Industry Dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Observations	797	797	797	797	797	797

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 11: *Regression table: First day returns using the residuals. In this table the results of ordinary least squares regressions are reported. Differently than before instead of the sentiment variables here are reported the residuals of them after having regressed the sentiment variables on the ex-ante IPO firm characteristics. Also after this control the results in explaining the underpricing are confirmed.*

	(1)	(2)	(3)	(4)	(5)	(6)
	First-day returns	First-day returns	First-day returns	First-day returns	First-day returns	First-day returns
N. Words	0.00001***	0.00001***	0.00001***	0.00001***	0.00001***	0.00001***
	(5.81)	(5.91)	(5.85)	(5.94)	(5.82)	(5.94)
xpositive	9.873*					
	(1.84)					
x_negative		2.119				
		(0.67)				
x_uncertain			6.373			
			(1.06)			
x_litigious				-4.585		
				(-0.55)		
x_modal weak					9.859	
					(1.07)	
x_modal strong						-8.222
						(-1.51)
Constant	0.370***	0.370***	0.370***	0.370***	0.370***	0.371***
	(17.56)	(17.54)	(17.55)	(17.56)	(17.56)	(17.59)
Year Dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Industry Dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Observations	1271	1271	1271	1271	1271	1271

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 12: *Regression table: Price Revision using the residuals. In this table the results of ordinary least squares regressions are reported. Differently than before instead of the sentiment variables here are reported the residuals of them after having regressed the sentiment variables on the ex-ante IPO firm characteristics. Also after this control the results in explaining the price revision are confirmed. hold.*

	(1) Price Revision	(2) Price Revision	(3) Price Revision	(4) Price Revision	(5) Price Revision	(6) Price Revision
N. Words	0.000001***	0.000001***	0.000001***	0.000001***	0.000001***	0.000001***
	(4.05)	(4.13)	(4.04)	(4.14)	(4.02)	(4.14)
x_positive	1.577*					
	(1.73)					
x_negative		0.132				
		(0.25)				
x_uncertain			1.641			
			(1.61)			
x_litigious				-0.0843		
				(-0.06)		
x_modal weak					2.452	
					(1.57)	
x_modal strong						-0.276
						(-0.29)
Constant	0.0282***	0.0282***	0.0283***	0.0282***	0.0283***	0.0282***
	(7.96)	(7.93)	(7.96)	(7.92)	(7.96)	(7.92)
Year Dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Industry Dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Observations	1473	1473	1473	1473	1473	1473

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 13: *Regression table . Heckman selection model for First Day Returns – two-step estimates (regression model with sample selection.). With this model we want to test for possible sample selection bias due to the fact that only for the IPOs that receive a coverage we are able to determine the quality of the coverage (i.e. the sentiment). For this reason we apply a Heckman two.stage model where in the first step we run a probit using as dependent variable the News dummy and the LogProceeds as an instrument, while in the second step we use the predicted values to correctly estimate models for the underpricing phenomenon.*

	(1)	(2)	(3)	(4)	(5)	(6)
	Underpricing	Underpricing	Underpricing	Underpricing	Underpricing	Underpricing
positive sentiment	13.16***					
	(3.08)					
negative sentiment		0.990				
		(0.40)				
litigious sentiment			-5.930			
			(-0.95)			
uncertainty sentiment				2.427		
				(0.51)		
modal weak sentiment					0.918	
					(0.13)	
modal strong sentiment						3.377
						(0.77)
Offer price revision	1.585***	1.608***	1.597***	1.608***	1.609***	1.611***
	(8.07)	(8.16)	(8.08)	(8.17)	(8.16)	(8.18)
Tech Dummy	0.117***	0.123***	0.124***	0.122***	0.124***	0.124***
	(3.06)	(3.20)	(3.24)	(3.18)	(3.21)	(3.24)
VC Dummy	0.137***	0.143***	0.141***	0.142***	0.143***	0.141***
	(3.61)	(3.76)	(3.68)	(3.72)	(3.73)	(3.70)
Reputation	0.536	0.587	0.610	0.592	0.597	0.602
	(1.40)	(1.53)	(1.59)	(1.54)	(1.55)	(1.57)
LogAge	-0.0923***	-0.0935***	-0.0936***	-0.0933***	-0.0936***	-0.0930***
	(-5.07)	(-5.11)	(-5.12)	(-5.10)	(-5.12)	(-5.08)
Constant	0.390***	0.443***	0.467***	0.440***	0.448***	0.434***
	(2.62)	(2.97)	(3.12)	(2.94)	(3.00)	(2.90)
	News	News	News	News	News	News
LogProceeds	0.254***	0.254***	0.254***	0.254***	0.254***	0.254***
	(7.53)	(7.53)	(7.53)	(7.53)	(7.53)	(7.53)
Price revision	1.334***	1.334***	1.334***	1.334***	1.334***	1.334***
	(7.03)	(7.03)	(7.03)	(7.03)	(7.03)	(7.03)
Tech Dummy	-0.0953	-0.0953	-0.0953	-0.0953	-0.0953	-0.0953
	(-1.75)	(-1.75)	(-1.75)	(-1.75)	(-1.75)	(-1.75)
VC Dummy	0.141***	0.141***	0.141***	0.141***	0.141***	0.141***
	(2.59)	(2.59)	(2.59)	(2.59)	(2.59)	(2.59)
Reputation	1.138**	1.138**	1.138**	1.138**	1.138**	1.138**
	(2.36)	(2.36)	(2.36)	(2.36)	(2.36)	(2.36)
LogAge	-0.0302	-0.0302	-0.0302	-0.0302	-0.0302	-0.0302
	(-1.13)	(-1.13)	(-1.13)	(-1.13)	(-1.13)	(-1.13)
Constant	-1.097***	-1.097***	-1.097***	-1.097***	-1.097***	-1.097***
	(-7.58)	(-7.58)	(-7.58)	(-7.58)	(-7.58)	(-7.58)
mills						
lambda	-0.119	-0.122	-0.131	-0.121	-0.121	-0.119
	(-0.79)	(-0.81)	(-0.87)	(-0.80)	(-0.80)	(-0.79)
Observations	2810	2810	2810	2810	2810	2810

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 14: *Regression table . Heckman selection model for Price Revision – two-step estimates (regression model with sample selection. With this model we want to test for possible sample selection bias due to the fact that only for the IPOs that receive a coverage we are able to determine the quality of the coverage (i.e. the sentiment). For this reason we apply a Heckman two.stage model where in the first step we run a probit using as dependent variable the News dummy and the LogProceeds as an instrument, while in the second step we use the predicted values to correctly estimate models for the price revision pattern.*

	(1)	(2)	(3)	(4)	(5)	(6)
	Price Revision	Price Revision	Price Revision	Price Revision	Price Revision	Price Revision
positive sentiment	2.101**					
	(2.50)					
negative sentiment		-0.296				
		(-0.61)				
litigious sentiment			-1.171			
			(-0.91)			
uncertainty sentiment				-0.0222		
				(-0.02)		
modal weak sentiment					-0.366	
					(-0.26)	
modal strong sentiment						-0.131
						(-0.15)
Tech Dummy	0.0311***	0.0326***	0.0323***	0.0323***	0.0325***	0.0323***
	(3.16)	(3.27)	(3.21)	(3.24)	(3.25)	(3.24)
VC Dummy	0.00702	0.00817	0.00772	0.00820	0.00836	0.00827
	(0.70)	(0.80)	(0.75)	(0.81)	(0.82)	(0.81)
Reputation	-0.0651	-0.0540	-0.0553	-0.0572	-0.0564	-0.0573
	(-0.65)	(-0.53)	(-0.54)	(-0.56)	(-0.56)	(-0.57)
LogAge	-0.0119**	-0.0121**	-0.0121**	-0.0121**	-0.0121**	-0.0121**
	(-2.47)	(-2.49)	(-2.47)	(-2.49)	(-2.49)	(-2.49)
Constant	0.183***	0.196***	0.197***	0.194***	0.195***	0.194***
	(5.35)	(5.70)	(5.68)	(5.62)	(5.66)	(5.64)
	News	News	News	News	News	News
LogProceeds	0.291***	0.291***	0.291***	0.291***	0.291***	0.291***
	(8.91)	(8.91)	(8.91)	(8.91)	(8.91)	(8.91)
Dummy Tech	-0.0505	-0.0505	-0.0505	-0.0505	-0.0505	-0.0505
	(-0.95)	(-0.95)	(-0.95)	(-0.95)	(-0.95)	(-0.95)
VC Dummy	0.145***	0.145***	0.145***	0.145***	0.145***	0.145***
	(2.71)	(2.71)	(2.71)	(2.71)	(2.71)	(2.71)
Reputation	1.123**	1.123**	1.123**	1.123**	1.123**	1.123**
	(2.37)	(2.37)	(2.37)	(2.37)	(2.37)	(2.37)
LogAge	-0.0408	-0.0408	-0.0408	-0.0408	-0.0408	-0.0408
	(-1.56)	(-1.56)	(-1.56)	(-1.56)	(-1.56)	(-1.56)
Constant	-1.208***	-1.208***	-1.208***	-1.208***	-1.208***	-1.208***
	(-8.55)	(-8.55)	(-8.55)	(-8.55)	(-8.55)	(-8.55)
mills						
lambda	-0.201***	-0.203***	-0.205***	-0.203***	-0.203***	-0.203***
	(-5.90)	(-5.90)	(-5.90)	(-5.90)	(-5.90)	(-5.91)
Observations	2877	2877	2877	2877	2877	2877

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Media and Finance: A Review

Carlo Raimondo*

Abstract

Understanding which is the role played by media in modern economies and, specifically, in financial markets has become a central point in the economic and finance academia. Even though an increasingly number of researches are proposed every year, we still lack a systematic review of what has been done and what has still to be searched. This paper aims to mind this gap, systematically and completely review related research papers.

I. Introduction

> One should hardly have to tell academicians that information is a valuable resource: knowledge *is* power.
>
> *George J. Stigler*
> The Economic of Information,
> *The Journal of Political Economy*, 1961

In the last decades the relationship between the media industry and the financial one has become increasingly important and studied. To understand which role media outlets play to the respect of markets and investors has been and it still is very challenging aim among economic and financial scholars. The old-fashion textbook idea of perfectly rational markets in which media are not substantially able to provide any real new information, due to the arbitrage-free condition[1], is day by day challenged and denied. Empirical studies have proved how media are effectively able to influence financial markets in several ways. Research community classified the nature of this relationship in different ways, according to different views on the role of media and on the quality of media products.

*PhD Student, University of Bologna, Department of Management, Via Capo di Lucca 34, 40126 Bologna, Italy. Corresponding author: carlo.raimondo@unibo.it.

[1]Following the arbitrage-free theory all the information is immediately used and discounted on the financial markets in the exact same moment in which it becomes available to anyone. Media require time to detect, process, package, and deliver the same information and therefore they should not be able to provide any market-sensitive information. Even though an efficient media is able to arrive to delivering very quickly, it should be by construction always late.

In this paper I am not going to argue the effectiveness of media on financial markets or the importance of media industry in the Internet era. Main aim of this review is to put together the numerous and sometimes discordant contributions to this field of research with the ambition of detect and underline common paths in order to actually understand the findings of this important stream of research. Moreover I will reserve a special section to summarize the main contributions to the stream of research which explains the media industry functioning, with a special focus on the media decision in terms of quality of produced information. Finally .a brief part of this paper will be devoted to the description and discussion of the main methodologies that have been used and created to process media outcomes data, namely data that assume the form of a simple text, like a newspaper's article or the transcription of a TV broadcast.

II. Why Media Matter

I. What is *information* and the role of media

The importance of information has been at length discussed in both empirical and theoretical literature. Starting from Von Hayek (1937; 1945) a large body of literature has been developed in economics about the role of information, with a special focus on economic theory. George Stigler and the *University of Chicago school* systematically addressed the topic, considering the role of information not *per se* but as an improvement to the *theory of General Equilibrium*[2]. In this framework, information and information spread and availability have to be considered as a crucial issue *inside* the model, but without discussing the main assumptions, logic, and results of the model itself. The contribution by Stigler (1961) points out which is the price of being not-informed and which is the role played by information dealers' to actually delivering and sharing reliable information[3]. Different is the viewpoint of Joseph Stiglitz, who devoted a large part of his prolific research activity to the role of information in the economy (Grossman and J. E. Stiglitz 1976; Grossman and J. E. Stiglitz 1980; J. E. Stiglitz and Weiss 1981; B. Greenwald, J. E. Stiglitz, and Weiss 1984; B. C. Greenwald and J. E. Stiglitz 1986; J. E. Stiglitz 1987; J. E. Stiglitz 2000; J. E. Stiglitz 2002). Stiglitz's main idea is focused on the necessity to devote a special attention to the information in modern economies: information has not to be considered a part of previous paradigm but

[2]Basing on the Arrow-Debreu model (Arrow and Debreu 1954) and its improvements (e.g. McKenzie 1959)

[3]"*Reputation* is a word which denotes the persistency of quality, and reputation commands a price because it economizes on search." (Stigler 1961). Here the author is referring to the markets of goods but it is easy to understand the generalizability of this theory. Accessing information, about a good's price or about a listed firm that issues financial instruments, is high costly and the best possible outcome involves the presence of a form of *information intermediary* which consumers or investors may actually trust. In this way the cost of accessing information is lowered.

it has to become the central focus of what Stiglitz named the *information economy*. Information and information asymmetries are at the centre of the research activities of two other outstanding economists, who contribute dramatically to change academic view on information topic. Akerlof (1970) introduced the idea of informational asymmetries and their role in competitive markets; Spence (1973) firstly addressed the idea of costly signalling mechanisms in presence of difficulties and costs in measuring unknown qualities (e.g. in the job market).

II. The media informative role

Understanding the importance and the value of information is critic in understanding the importance of information communication, creation, and distribution. In modern economies professional figures have been created to deal with this issue, creating the *media industry*. As soon as individuals and organizations understood the importance and the worth of accessing a good information setting, media firms were founded. Media owners and entrepreneurs found this form of economic activity profitable, investors and individuals found important to access their product. Looking at the media sector as an industry *inside* the economy and not *on a side* of it is crucial to understand the media industry functioning. If media industry is nothing more and nothing different than a set of firms maximizing profits and not a sort of *public service* devoted to distribute a public good, it is easy to understand the importance of accurately analysing its functioning mechanisms. Jensen in his 1976 paper (1979) first identified the main problem that could arise in the media industry between information quality and markets forces. In the media sector the idea of optimal equilibrium following a fair market competition is, as the matter of the facts, almost impossible to be applied. Jensen argues the phenomena and topics usually described by newspapers' articles are, in most cases, too difficult and to complex to be honestly and completely reported to the audience. In this situation a newspaper has to be understood by its readers and this leads to a sort of demand for slant. Jensen uses this finding to demonstrate that a special regulation for this kind of industry is necessary. This is a first key-point of this review: media industry is an important and developed industry formed by firms among which the vast majority are in business for profit, not for ethical reasons or for providing a public good. Furthermore even the quality of the outcome, i.e. *the information*, cannot be evaluated unequivocally. As I will show in the proceedings of this review, not all the audience wants the highest quality information, a relevant part of it wants instead to have their beliefs confirmed[4], independently to what actually happened.

[4]This phenomenon has been widely investigated in the economic and financial literature, among the others by Mullainathan and Shleifer 2005. In their paper the authors show how if the readers in a newspaper's audience have strong shared beliefs, for the newspaper is very unlikely and difficult to go against these beliefs, even though there could be

The very special relationship between competition among firms and outcomes' quality, specific of the media sector, will be one of the key-points of Section IV. Hereinafter this research paper will try to disentangle the different ideas scholars have produced about the effects of media on financial markets and some possible feature at the base of media industry functioning.

III. The Media Industry

Media are information dealers, able to select facts, happenings, judgements, firms' results, and transform it in pieces of news to be delivered to their audience. Articles and all kinds of piece of news come from facts and opinions that media judge enough important and interesting to be delivered to their audience. In this process media judging is very vast and crucial: it is what transform in a news something that could also not become a news. To these extents media power is huge and indisputable. Which are the boundaries to this free will?

Classical view of media industry is based on the reputation effect that relies on multiple observation of single media outlet's outcomes. Consumers (i.e. audience, or readers) are willing to pay to be informed by a media firm if they judge it in a positive way. Judging should be based on past observation of the media itself,providing in this way the right incentive to the media to reach and maintain an high quality standard in order to both keep its audience and, possibly, conquest new consumers in the competition against other media firms. According to this framework the presence of free market competition, several media firms, and rationality among consumers, produce a dynamic equilibrium in which low reputed media are pushed out from the market by the higher reputed ones. Furthermore the information produced by this kind of situation is levelled on the highest quality[5], that basically means effective information, able to describe and even predict political and economic facts and changes at the best. No room has been left for irrationality, or for a different role of media different than the delivering of the best informative set available in each single moment.

Is media sector actually working in this way? And, more importantly, it is delivering *actually new* news? The novelty of media produced pieces of news assumes a more important value if referred to economic news and, particularly, to financial markets related contents. If media are delivering exclusively pieces of news already known by market participants and already discounted in assets' prices, no reactions on financial markets should follow the disclosure of new

strong evidence and supportive sources supporting it.

[5]In equilibrium all not reputed media firms, producing low-quality information, should be overcome by highly reputed media firms producing high-quality information. As a result, we have on the market only high quality information produced by the most competitive media firms.

contents. At the contrary, if media are actually able to deal new information or, at least, valuable signals to markets' participants, financial markets should react systematically to news' disclosure by reliable media outlets. Several scholars did study this *phenomenon* in order to test the so called *markets efficiency hypothesis* in several contexts and applying different methodologies. Furthermore, when markets efficiency hypothesis has been rejected, a very wide set of explanations have been proposed. In the next section I will go to describe the most important research works in more details, always trying to consider at the same time their contribution to the field of research.

I. Effects on Financial Markets

Curiously, the first contribution about role of media on financial markets is able to put together media inefficiency in delivering information and media effectiveness on financial markets. Huberman and Regev in their 2001 paper analyzed the effect of an article published on the Sunday's edition of the *New York Times* about a small firm fighting to find a cure for cancer. No real news, no new facts or analysis were involved in the article. All information was already available from five months earlier, but stock markets did react hugely to the news[6], arriving to haul all the bio-tech sector on the stock exchange. Their findings were motivated by a view in which the induced public attention is able to generate a permanent higher stock price, even though no new information is actually delivered. A somehow coherent result is contained in the paper by Dyck and Zingales (2003): the authors find that earnings announcements that are massively and optimistically reported by the press are more likely to generate rises in stock prices, and with a greater magnitude. They also find the effect to be greater, fewer are the media following the news and higher is the reputation of the media itself. Media firms are here supposed to follow sources that are internal to the *reported* firms: in order to obtain internal (i.e. private) information, journalists are willing to reward their sources with a better (i.e. *positively slanted*) coverage[7].
Johnson et al. (2005) studied the effect of being covered for a firm's executives in the ratings published periodically by *Business Week* magazine. Both positive and negative judgements for a firm's executives produce positive abnormal returns for its stock price. The effect is not produced by the quality of the information but by the simple fact to be covered and to *jump* to the investors' attention (*coverage effect*). Tetlock (2007) tried to systematically and objectively quantify the content of textual information[8]. In its paper published on *The Journal of Finance*, he finds that media

[6]The *EntreMed* stock price grew up from 12 USD closing price on Friday to 85 USD opening price on Monday.

[7]Dyck and Zingales named this relationship the "qui pro quo" condition that is effectively a source of conflict of interest between media outlet and firm the media's articles are about.

[8]Addressing this issue has been for a long time an obstacole to the development of this kind of research. In Section IV I will provide a brief overview of the methodologies used, and I will discuss the most important advantages and

pessimism is a good predictor of financial markets' downwards associated with an increase in the trading volume. Tetlock provided in this way a strong evidence of media effectiveness on the financial markets, even though the mechanism of propagation between media and investors is still uncertain. Similarly but in a different setting DellaVigna and Kaplan (2007) proved the effectiveness of media referring to the politics[9]. Political elections may be easily view as a sort of *market* in which individuals decide not about their wealth investment but about their communities governments. Furthermore electors have to decide basing on the information they have together with the beliefs they hold. DellaVigna and Kaplan showed how electors are heavily and significantly influenced by news' quality and news preferences toward a specific political party. In their paper the authors focus on the effects of Republican-friendly *Fox-News TV* spread in the U.S., finding an increase in votes for Republican candidates after the arrive of *Fox-News TV* in each single county. Media are then effective, at least in generating a shared consensus on specific beliefs. Following the reasoning of the authors, we may argue that investors do not completely change their mind toward the opposite party, but they were effectively motivated to sustain their party after observing their ideas to be sustained by media. Translating this reasoning in financial terms, we do not observe to a change in already active investors, but we see how inactive investors may go to the markets when they observe their idea to be sustained by the media, generating in this way an increase in volume (in politics, more people to vote) and generating more demand on the chosen stocks (the supported parties receive more votes). Engelberg (2008) studies tried to disentangle the effect of *hard* (i.e. numbers, results that can be easily quantified) to *soft* (comment, opinions, and all what is not possible to be quantified) information in the earnings announcements empirical setting. He found the *qualitative* information to be effective on stock markets and, on the top of that, to be more effective than quantified information.

Joe et al. (2009), relying on the findings of the previous paper of Allen et al. (2006), developed a model in which media play an important role towards financial markets by deciding to cover or not firms' board ineffectiveness. When media devote space and attention to boards' ineffectiveness issues, financial markets react differently according to their nature: individual investors react penalizing the firm; at the contrary institutional investors and professional seem to be able to

disadvantages of each of them.

[9]Why reporting here something not directly linked to financial markets? The main reason is that in term of information disclosure patterns and media role financial markets and *political markets* are not so dissimilar. Precisely referring to the occurrence of elections a large numbers of individuals (the electors) have to decide about something important with a limited informative set. Therefore they will see and understand their choice's results only in a later stage. These patterns are very close to those involving an individual investor who goes to the financial markets deciding about something important basing on a limited informative set.

anticipate the news' delivering by media. This contribution somehow links this literature to that one about corporate governance role of media I am going to present in a special section of this paper. Both the paper by Bushee et al. (2010) and Bushman et al. (2010) look at the media as essentially *information dealers*. In their view media are actually able to deliver new information to the market, using their professional skill to find the not already public information. Bushman et al. (2010) find out a dissemination of private information due to the relationship between landers and borrower; Bushee et al. (2010) instead provide evidence of the media role in solving information asymmettries in the case of earnings announcement. Tetlock (2010) has provided further investigation about the possibility and the ability of media in resolving informational asymmetries among investors in financial markets. In his paper the author analyses to which extents media are effectively able to deliver to markets and investors new information. Basing on a set of price-sensitive events, Tetlock showed how part of financial markets participants are actually predicting media contents, reaching though sensitive news before media publish it. These investors then exploit this information trading consequentially. The other part of investors is instead responsive to media-delivered news, reacting to it in terms of trading activity only after media disclosure. In this view media play the important role of *setting the field*, aligning informed and not-informed investors' informative sets. Media action results in strengthening the market efficiency, resolving the provisional information misalignment among investors. The main challenge in analyzing media role and effects on financial markets consists in correctly understanding what is due to the events reported by the media and what instead is due to the media reporting itself. In order to address this issue, Engelberg and Parson (2011) exploited the differences in the media content diffusion in different geographical areas inside the United States. Due to the role of local-diffused newspapers, different sets of investors receive different informational sets. Using data from local trading, the authors show how local newspapers' coverage on earnings announcements strongly predicts local trading, even considering for a wide set of possible compounding factors. Furthermore the authors provide evidence about the linkage between local trading activity and the local media news disclosure timing. Birz and Lott (2011) analyzed how financial markets react to real sector economic news. In their paper they provided empirical evidence about the role of media in information diffusion. They showed how economic news covered by and reported by media outlets are more likely to be able to influence financial markets massively. Consistently whit the previously mentioned research papers, it may be said that media are not only able to be effective in focusing the public attention on some specific topics, but they have a wider role in financial markets. Media in fact may also spread not-new pieces of information, pretending they are actually news and conditioning in this way

their audiences. To understand if news' consumers (i.e. investors) rely on media disclosure, Tetlock (2011) analyzed what happens when media publish *stale* information. Evidence from his paper shows how individual investors over-react to stale information generating in this way stronger momentum for stocks with higher participation of individual investors' trading. In the media industry professional journalists play a pivotal role, contributing to the generation of news and to the information processing process. Furthermore some of them are able to become well known and followed public figures, characterizing themselves with specific characteristics. Understanding the role of the single indivudual journalist in the big puzzle of financial media seems to be important, since they can influence importantly the information disclosure process. Dougal et al (2012) take in account for such a phenomenon analyzing the stock markets reactions to media delivered news including a journalist fixed-effect. In their framework the authors have been able to show how financial news impact on stock markets is effective and significant. Similarly Ahern and Sosyura (2015) study the single journalist effect in terms of reputation and expertise they bring to the articles they authored. According to them, accuracy of news is well predicted by journalist's experience, specialized education, and industry expertise. Furthermore they show how media are often able to light the spot on important and effective facts, generating important consequences on the financial markets. Solomon and Soltes (2012) underline the risk of manipulated media, in a view in which firms' managers try to influence and manipulate the press in their favor. Empirical evidence though weakly supports their framework. Aman (2013) provides additional evidences about media role in financial markets, especially during bubbles and crashes, focusing on the Japanese stock market. Papers by Boudoukh (2013), Ferguson et al. (2014), Kraussl and Mirgorodskaya (2014), and Solomon, Soltes, and Sosyura (2014) contribute to strength the empirical evidences of a causal and significant effect of media on financial markets. Nevertheless detecting media content causal effects on financial markets is a critical point in the financial literature: Peress (2014) used newspapers strikes to understand when and how media influence the markets, providing a convincing proof of media effectiveness in a cross-country sample. For completeness papers showing not relevant effect of media on financial markets have to been mentioned here. Actually just few papers show an not effective role of media on financial market. The most important is that by Bhattacharya et al. (2009). The paper finds no important role of media during the IPO bubble in the nineties. Investors' irrationality was not generated or at least, not fully explained by media behaviours in processing information. Also in this case media effectivenessss, even if negated, is actually confirmed. Media are noresponsiblele for market irrationality during the high-tech IPO bubble but they just contribute to spread investors' sentiment.

II. The Corporate Governance Role of Media

Media has been often viewed as a very special industry. Among the many, the media ability to condition other industries business is the prominent one. Furthermore media play a very important role towards all the markets and the institutions whose functioning is guaranteed by a reputation based equilibrium. If reputation is important for firms, media are able to influence it and firms have to properly consider the reputational risk if their bad bahaviour could be covered by media. This is what is usually called the *Corporate Governance Role of Media*. First research paper underling this issue refer to the public sector and to the administration of States and Local Governments. Islam (2002; 2003; 2006) analyzed how countries with more developed media sectors are associated with better governments and smaller levels of corruption. Besley and Prat (2006) find similar conclusion developing a formal economic model for the corruption framework in which media role is questioned when the media sector is deeply conditioned by the governments itself. Egerov et al. (2009) proved how media are relevant in fighting public corruption and bribing in the setting of developing economies states ruled by dictatorships: the authors show how dictatorships located in resoruces poor countries are less likely to try to control the press and more likely to allow free press in order to provide the right incentives to their bureaucrats and therefore to improve the quality of government.

Media role has a pivotal role also when we refer to the private sector and to the business administration. The papers by Becht et al. (2003), Johnson et al. (2005), and Miller (2006) shift the focus on the media role on firms and firms' stakeholders. Both of the papers show how efficient media activity is able to preserve and guarantee stakeholders' wealth against possible frauds or manipulations operated by firms executives. They are the first to properly emphasize the role of media in the business sector governance. Dyck et al. (2008) studied the role of media in the Russia corporate sector, finding the foreign media to be more effective than the domestic ones due to some reputational issues and to the fact of notefficient and scarcely reliable doemstic media industry. The importance of media industry characteristics and efficency to be able to play the corporate governance role is crucial. Only reputed and followed media may obtain the effect of influence efficiently and sistmaticaly private sector corporate governace, acting like *watchdog* (Miller 2006) and not like a *lapdog* (Bednar 2012) in dislosing and preventing corporate frauds (Dyck, Morse, and Zingales 2010). Several scholars studied the role of media in corporate governance referring to peculiar phenomena, like executive copensation, board efficency and executive decisions. Media play also a role in the executive compensation puzzle, leveraging on the reputational cost of incongruous compensations. Core et al. (2008) linked media coverage of firms' activities to a

correct and efficient executive compensation. Media effects on executive compensation is also put in evidence by the work by Kunhen and Niessen (2012) where the role of public opinion is properly considered. Related to this topic Joe et. al. (2009) undelined how media can be important also in covering board inefficencies. Other scholar studied the role of media in case of dual class shares unification and shareholders protection (Lauterbach and Pajuste 2012), or in cases of capital allocation decisions (B. Liu and McConnell 2013), or also in case of insider trading disclosure and prevention (Dai, Parwada, and Zhang 2015).

IV. Textual Analysis in Finance

A big boost in the literature about role of information and media in the financial sector was generated by the creation and the subsequent availability of methodologies to process textual data. Media outcomes as well as the vast majority of what we call information is usually in form of written text. Analyzing this kind of content has been challenging for a long period: it was difficult and time-consuming to process a sufficient big size of data *by hand* and, at the same time, objectivity and replicability of this kind of researches were very disputable[10]. One of the first attempt to apply automatic (i.e. not performed by *humans*) analysis of textual is that by Paul Tetlock (2007). He built some measures of texts' *sentiment* essentially based on *Hardvard Dictionaries*. Tetlock also introduced the *word list* methodology, where a set of words indicating positiveness, negativeness, or other connotations is defined and then used *counting* in each article or generic text the sum of the occurrences of all the words of each list. Linguistically pretty simple and even *dirty*, this is still the most used and effective approach in financial textual analysis. Tetlock et al. (2008) tried to improve the textual analysis method finding an effective way to measure the *negative* tone of an article or a text, and they applied their mothodology to earnings announcements finding significant results. The real watershed in this methodological debate is represented by the paper published by Loughran and McDonald (2011), where they investigate about financial language specific characteristics and they developed several word lists able to detect different textual tones, like the positive tone, the negative tone, the uncertainty tone, and others. This methodology has been later widely used by several scholars (for instance Dougal et al. 2012 and Ahern and Sosyura 2015). Most recent improvements are those by Kearney and Liu (2014) that tries to put together all the improvements in the field, and by the two last paper by Loughran and McDonald (2014; 2014) in which they stressed and enlarged the validity of their methodology published before (Loughran and McDonald 2011).

[10]Main issue in this case is related to the human judgement of a content.

References

Ahern, Kenneth R and Denis Sosyura (2014). "Rumor Has It: Sensationalism in Financial Media". In: *Available at SSRN 2264468*.

— (2015). "Rumor Has It: Sensationalism in Financial Media". In: *Review of Financial Studies*.

Akerlof, George A (1970). "The market for" lemons": Quality uncertainty and the market mechanism". In: *The quarterly journal of economics*, pp. 488–500.

Allen, Franklin, Stephen Morris, and Hyun Song Shin (2006). "Beauty contests and iterated expectations in asset markets". In: *Review of Financial Studies* 19.3, pp. 719–752.

Aman, Hiroyuki (2013). "An analysis of the impact of media coverage on stock price crashes and jumps: Evidence from Japan". In: *Pacific-Basin Finance Journal* 24, pp. 22–38.

Arrow, Kenneth J and Gerard Debreu (1954). "Existence of an equilibrium for a competitive economy". In: *Econometrica: Journal of the Econometric Society*, pp. 265–290.

Baron, David P (2005). "Competing for the public through the news media". In: *Journal of Economics & Management Strategy* 14.2, pp. 339–376.

— (2006). "Persistent media bias". In: *Journal of Public Economics* 90.1, pp. 1–36.

Becht, Marco, Patrick Bolton, and Ailsa Röell (2003). "Corporate governance and control". In: *Handbook of the Economics of Finance* 1, pp. 1–109.

Bednar, Michael K (2012). "Watchdog or lapdog? A behavioral view of the media as a corporate governance mechanism". In: *Academy of Management Journal* 55.1, pp. 131–150.

Besley, Timothy and Andrea Prat (2006). "Handcuffs for the grabbing hand? Media capture and government accountability". In: *The American Economic Review*, pp. 720–736.

Bhattacharya, Utpal et al. (2009). "The role of the media in the internet IPO bubble". In: *Journal of Financial and Quantitative Analysis* 44.03, pp. 657–682.

Binderkrantz, Anne Skorkjær (2012). "Interest groups in the media: Bias and diversity over time". In: *European Journal of Political Research* 51.1, pp. 117–139.

Birz, Gene and John R Lott Jr (2011). "The effect of macroeconomic news on stock returns: New evidence from newspaper coverage". In: *Journal of Banking & Finance* 35.11, pp. 2791–2800.

Boudoukh, Jacob et al. (2013). "Which news moves stock prices? A textual analysis". In:

Brockman, Paul, Ivonne Liebenberg, and Maria Schutte (2010). "Comovement, information production, and the business cycle". In: *Journal of Financial Economics* 97.1, pp. 107–129.

Bushee, Brian J et al. (2010). "The role of the business press as an information intermediary". In: *Journal of Accounting Research* 48.1, pp. 1–19.

Bushman, Robert M, Abbie J Smith, and Regina Wittemberg-Mortman (2010). "Price discovery and dissemination of private information by loan syndicate participants". In: *Journal of Accounting Research* 48.5, pp. 921–972.

Chiang, Chun-Fang and Brian Knight (2011). "Media bias and influence: Evidence from newspaper endorsements". In: *The Review of Economic Studies*.

Core, John E, Wayne Guay, and David F Larcker (2008). "The power of the pen and executive compensation". In: *Journal of Financial Economics* 88.1, pp. 1–25.

Dai, Lili, Jerry T Parwada, and Bohui Zhang (2015). "The Governance Effect of the Media's News Dissemination Role: Evidence from Insider Trading". In: *Journal of Accounting Research*.

DellaVigna, Stefano and Ethan Kaplan (2007). "The Fox News Effect: Media Bias and Voting". In: *The Quarterly Journal of Economics* 122.3, pp. 1187–1234.

DellaVigna, Stefano and Alec Kennedy (2011). "Does media concentration lead to biased coverage? Evidence from movie reviews". In: *Unpublished*.

DellaVigna, Stefano and Joshua M Pollet (2009). "Investor inattention and Friday earnings announcements". In: *The Journal of Finance* 64.2, pp. 709–749.

Djankov, Simeon et al. (2001). "Who owns the media?" In: *National Bureau of Economic Research Working Paper*.

Dougal, Casey et al. (2012). "Journalists and the stock market". In: *Review of Financial Studies*.

Durante, Ruben and Brian Knight (2012). "Partisan control, media bias, and viewer responses: Evidence from Berlusconi's Italy". In: *Journal of the European Economic Association* 10.3, pp. 451–481.

Dyck, Alexander, Adair Morse, and Luigi Zingales (2010). "Who blows the whistle on corporate fraud?" In: *The Journal of Finance* 65.6, pp. 2213–2253.

Dyck, Alexander, David Moss, and Luigi Zingales (2008). "Media versus special interests". In:

Dyck, Alexander, Natalya Volchkova, and Luigi Zingales (2008). "The corporate governance role of the media: Evidence from Russia". In: *The Journal of Finance* 63.3, pp. 1093–1135.

Dyck, Alexander and Luigi Zingales (2003). "The Media and Asset Prices". In:

Egorov, Georgy, Sergei Guriev, and Konstantin Sonin (2009). "Why resource-poor dictators allow freer media: a theory and evidence from panel data". In: *American Political Science Review* 103.04, pp. 645–668.

Engelberg, Joseph (2008). "Costly information processing: Evidence from earnings announcements". In:

Engelberg, Joseph E and Christopher A Parsons (2011). "The causal impact of media in financial markets". In: *The Journal of Finance* 66.1, pp. 67–97.

Entman, Robert M (2007). "Framing bias: Media in the distribution of power". In: *Journal of communication* 57.1, pp. 163–173.

Ferguson, Nicky J et al. (2014). "Media Content and Stock Returns: The Predictive Power of Press". In: *Midwest Finance Association 2013 Annual Meeting Paper*.

Fergusson, Leopoldo (2014). "Media markets, special interests, and voters". In: *Journal of Public Economics* 109, pp. 13–26.

Gal-Or, Esther, Tansev Geylani, and Tuba Pinar Yildirim (2012). "The impact of advertising on media bias". In: *Journal of Marketing Research* 49.1, pp. 92–99.

Gentzkow, Matthew and Jesse M Shapiro (2006). "Media Bias and Reputation". In: *Journal of Political Economy* 114.2, pp. 280–316.

— (2008). "Competition and Truth in the Market for News". In: *The Journal of Economic Perspectives*, pp. 133–154.

— (2010). "What drives media slant? Evidence from US daily newspapers". In: *Econometrica* 78.1, pp. 35–71.

George, Lisa (2007). "What's fit to print: The effect of ownership concentration on product variety in daily newspaper markets". In: *Information Economics and Policy* 19.3, pp. 285–303.

George, Lisa M and Joel Waldfogel (2006). "The" New York Times" and the Market for Local Newspapers". In: *The American economic review*, pp. 435–447.

George, Lisa and Joel Waldfogel (2003). "Who affects whom in daily newspaper markets?" In: *Journal of Political Economy* 111.4, pp. 765–784.

Gerber, Alan S, Dean Karlan, and Daniel Bergan (2009). "Does the Media Matter? A Field Experiment Measuring the Effect of Newspapers on Voting Behavior and Political Opinions". In: *American Economic Journal: Applied Economics* 1.2, pp. 35–52.

Greenwald, Bruce C and Joseph E Stiglitz (1986). "Externalities in economies with imperfect information and incomplete markets". In: *The quarterly journal of economics*, pp. 229–264.

Greenwald, Bruce, Joseph E Stiglitz, and Andrew Weiss (1984). "Informational imperfections in the capital market and macroeconomic fluctuations". In: *The American Economic Review*, pp. 194–199.

Groseclose, Tim and Jeffrey Milyo (2005). "A measure of media bias". In: *The Quarterly Journal of Economics*, pp. 1191–1237.

Grossman, Sanford J and Joseph E Stiglitz (1976). "Information and competitive price systems". In: *The American Economic Review*, pp. 246–253.

— (1980). "On the impossibility of informationally efficient markets". In: *The American economic review*, pp. 393–408.

Gurun, Umit G and Alexander W Butler (2012). "Don't believe the hype: Local media slant, local advertising, and firm value". In: *The Journal of Finance* 67.2, pp. 561–598.

Hamilton, James T and Richard Zeckhauser (2004). "Media Coverage of CEOs: Who? What? Where? When? Why?" In: *Unpublished working paper, Stanford Institute of International Studies*.

Houston, Joel F, Chen Lin, and Yue Ma (2011). "Media ownership, concentration and corruption in bank lending". In: *Journal of Financial Economics* 100.2, pp. 326–350.

Huberman, Gur and Tomer Regev (2001). "Contagious speculation and a cure for cancer: A nonevent that made stock prices soar". In: *The Journal of Finance* 56.1, pp. 387–396.

Islam, Roumeen (2002). *The right to tell: the role of mass media in economic development*. World Bank Publications.

— (2003). *Do more transparent governments govern better?* World Bank, Washington, DC.

— (2006). "Does more transparency go along with better governance?" In: *Economics & Politics* 18.2, pp. 121–167.

Jensen, Michael C (1979). "Toward a Theory of the Press". In: *Economics Social Institutions*. Springer, pp. 267–287.

Joe, Jennifer R, Henock Louis, and Dahlia Robinson (2009). "Managers' and investors' responses to media exposure of board ineffectiveness". In: *Journal of Financial and Quantitative Analysis* 44.03, pp. 579–605.

Johnson, Jonathan L et al. (2005). "The influence of the financial press on stockholder wealth: The case of corporate governance". In: *Strategic Management Journal* 26.5, pp. 461–471.

Kearney, Colm and Sha Liu (2014). "Textual sentiment in finance: A survey of methods and models". In: *International Review of Financial Analysis*.

Kräussl, Roman and Elizaveta Mirgorodskaya (2014). "News Media Sentiment and Investor Behavior". In: *Available at SSRN 2246295*.

Kuhnen, Camelia M and Alexandra Niessen (2012). "Public opinion and executive compensation". In: *Management Science* 58.7, pp. 1249–1272.

Lauterbach, Beni and Anete Pajuste (2012). "The media role in corporate governance improvement: Lessons from dual class share unifications". In: *Unpublished manuscript. Bar Ilan University*.

Liu, Baixiao and John J McConnell (2013). "The role of the media in corporate governance: Do the media influence managers' capital allocation decisions?" In: *Journal of Financial Economics* 110.1, pp. 1–17.

Lott Jr, John R and Kevin A Hassett (2014). "Is newspaper coverage of economic events politically biased?" In: *Public Choice* 160.1-2, pp. 65–108.

Loughran, Tim and Bill McDonald (2011). "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks". In: *The Journal of Finance* 66.1, pp. 35–65.

— (2013). "IPO first-day returns, offer price revisions, volatility, and form S-1 language". In: *Journal of Financial Economics* 109.2, pp. 307–326.

— (2014a). "Measuring readability in financial disclosures". In: *The Journal of Finance*.

— (2014b). "The use of word lists in textual analysis". In: *Available at SSRN 2467519*.

McKenzie, Lionel W (1959). "On the existence of general equilibrium for a competitive market". In: *Econometrica: journal of the Econometric Society*, pp. 54–71.

Miller, Gregory S (2006). "The press as a watchdog for accounting fraud". In: *Journal of Accounting Research* 44.5, pp. 1001–1033.

Mullainathan, Sendhil and Andrei Shleifer (2002). "Media bias". In:

— (2005). "The market for news". In: *American Economic Review*, pp. 1031–1053.

Peress, Joel (2014). "The media and the diffusion of information in financial markets: Evidence from newspaper strikes". In: *The Journal of Finance*.

Petrova, Maria (2011). "Newspapers and parties: How advertising revenues created an independent press". In: *American Political Science Review* 105.04, pp. 790–808.

— (2012). "Mass media and special interest groups". In: *Journal of Economic Behavior & Organization* 84.1, pp. 17–38.

Reuter, Jonathan and Eric Zitzewitz (2006). "Do ads influence editors? Advertising and bias in the financial media". In: *The Quarterly Journal of Economics*, pp. 197–227.

Rothschild, Micheal and Joseph Stiglitz (1976). "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information". In: *The Quarterly Journal of Economics*, pp. 629–649.

Sobbrio, Francesco (2011). "Indirect Lobbying and Media Bias". In: *Quarterly Journal of Political Science* 6, pp. 235–274.

Solomon, David H (2012). "Selective publicity and stock prices". In: *The Journal of Finance* 67.2, pp. 599–638.

Solomon, David H and Eugene F Soltes (2011). "The determinants of coverage in the business press". In: *SSRN eLibrary*.

Solomon, David H, Eugene Soltes, and Denis Sosyura (2014). "Winners in the spotlight: Media coverage of fund holdings as a driver of flows". In: *Journal of Financial Economics* 113.1, pp. 53–72.

Solomon, David and Eugene Soltes (2012). "Managerial Control of Business Press Coverage". In: *Available at SSRN 1918138*.

Spence, Michael (1973). "Job market signaling". In: *The quarterly journal of Economics*, pp. 355–374.

Stigler, George J (1961). "The economics of information". In: *The journal of political economy*, pp. 213–225.

Stiglitz, Joseph E (1987). "The causes and consequences of the dependence of quality on price". In: *Journal of economic literature*, pp. 1–48.

— (2000). "The contributions of the economics of information to twentieth century economics". In: *Quarterly Journal of economics*, pp. 1441–1478.

— (2002). "Information and the Change in the Paradigm in Economics". In: *American Economic Review*, pp. 460–501.

Stiglitz, Joseph E and Andrew Weiss (1981). "Credit rationing in markets with imperfect information". In: *The American economic review*, pp. 393–410.

Tahoun, Ahmed (2014). "The role of stock ownership by US members of Congress on the market for political favors". In: *Journal of Financial Economics* 111.1, pp. 86–110.

Tetlock, Paul C (2007). "Giving content to investor sentiment: The role of media in the stock market". In: *The Journal of Finance* 62.3, pp. 1139–1168.

— (2010). "Does public financial news resolve asymmetric information?" In: *Review of Financial Studies* 23.9, pp. 3520–3557.

— (2011). "All the news that's fit to reprint: Do investors react to stale information?" In: *Review of Financial Studies* 24.5, pp. 1481–1512.

Tetlock, Paul C, Maytal Saar-Tsechansky, and Sofus Macskassy (2008). "More than words: Quantifying language to measure firms' fundamentals". In: *The Journal of Finance* 63.3, pp. 1437–1467.

Veldkamp, Laura L (2006). "Media frenzies in markets for financial information". In: *The American economic review*, pp. 577–601.

Von Hayek, Friedrich August (1937). "Economics and knowledge". In: *Economica*, pp. 33–54.

— (1945). "The use of knowledge in society". In: *The American economic review*, pp. 519–530.

Walker, Clive B (2014). "Housing booms and media coverage". In: *Applied Economics* 46.32, pp. 3954–3967.

Westphal, James D and David L Deephouse (2011). "Avoiding bad press: Interpersonal influence in relations between CEOs and journalists and the consequences for press reporting about firms and their leadership". In: *Organization Science* 22.4, pp. 1061–1086.

Xiang, Yi and Miklos Sarvary (2007). "News consumption and media bias". In: *Marketing Science* 26.5, pp. 611–628.