AARHUS
UNIVERSITY
BUSINESS AND SOCIAL SCIENCES
DEPARTMENT OF ECONOMICS AND BUSINESS

CREATES
Center for Research in Econometric Analysis of Time Series

And Now, The Rest of the News: Volatility and Firm Specific News Arrival

Robert F. Engle, Martin Klint Hansen and Asger Lunde

CREATES Research Paper 2012-56

Department of Economics and Business
Aarhus University
Fuglesangs Allé 4
DK-8210 Aarhus V
Denmark

Email: oekonomi@au.dk
Tel: +45 8716 5515

And Now, The Rest of the News: Volatility and Firm Specific News Arrival *

Robert F. Engle
Stern School of Business, New York University

Martin Klint Hansen and Asger Lunde†
CREATES & Aarhus University, Denmark

December 4, 2012

Abstract

Starting with the advent of the event study methodology, the puzzle of how public information relates to changes in asset prices has unraveled gradually. Using a sample of 28 large US companies, we investigate how more than 3 million firm specific news items are related to firm specific stock return volatility. We specify a return generating process in conformance with the mixture of distributions hypothesis, where stock return volatility has a public and a private information processing component. Following public information arrival, prices incorporate public information contemporaneously while private processing of public information generates private information that is incorporated sequentially. We refer to this model as the information processing hypothesis of return volatility and test it using time series regression. Our results are evidence that public information arrival is related to increases in volatility and volatility clustering. Even so, clustering in public information does not fully explain volatility clustering. Instead, the presence of significant lagged public information effects suggest private information, generated following the arrival of public information, plays an important role. Including indicators of public information arrival explains an incremental 5 to 20 percent of variation in the changes of firm specific return volatility. Contrary to prior financial information research, our investigation favors the view that return volatility is related to public information arrival.

JEL Classification: G14
Keywords: Firm Specific News, Realized Volatility, Public Information Arrival

*Acknowledgements. We are grateful to Torben Andersen, Tim Bollerslev, Christian Brownlees, Nikolaus Hautsch, Yu Jun, Andrew Patton, Peter Phillips, Kevin Sheppard, and Carsten Tanggaard for helpful discussions. We are also thankful for helpful comments received at the HUKU 2011 conference in Copenhagen, and at seminars at CREATES, Copenhagen Business School and Duke University Singapore Management University. The second author acknowledges financial support from the Bikuben, Oticon and Tuborg foundations. We acknowledge financial support by the Center for Research in Econometric Analysis of Time Series, CREATES, funded by the Danish National Research Foundation. All errors are ours.

†Corresponding author: alunde@econ.au.dk

1. Introduction

Empirical research, confronting the paradigm that changes in stock prices are related to the arrival of new economic public information, has reached mixed conclusions. While both financial theory and empirical research suggest unanticipated public information affect stock prices, research using all observable macroeconomic and industry specific information has found high levels of inexplicable price volatility (Roll 1988; Cutler, Poterba, and Summers 1989).

Initially, research, spurred by the advent of the event study methodology, explored how corporate information events relate to changes in stock prices. Among many things, these efforts led to the first studies using aggregate news counts to examine the relationship between market activity and the flow of public information (Ederington and Lee 1993; Berry and Howe 1994; Mitchell and Mulherin 1994). Motivated by the recent availability of extensive electronic news databases, current branches of research investigate the relationships between news arrival and conditional volatility (Kalev, Liu, Pham, and Jarnecic 2004), investor sentiment and market activity (Tetlock 2007), news coverage and return predictability (Fang and Peress 2009), as well as news readership and realized volatility (Lumsdaine 2009).

Financial information research faces two key challenges (Boudoukh, Richardson, Shen, and Whitelaw 2007). The first challenge is to identify and observe all relevant fundamental information for a specific financial asset. The second challenge consists of correctly quantifying and measuring the fundamental. This paper builds on existing financial information research by investigating the relationship between economic information arrival and changes in stock return volatility. Our primary contribution is to identify a collectively exhaustive measure of firm specific newsflow by collecting all firm specific news in the cross-section of approximately thirty thousand different news sources accessible through the Dow Jones Factiva database. Our dataset is probably one of the most comprehensive news datasets employed in financial research.

We depart from a simple three distribution mixture model for the return generating process, which suggests stock return volatility has two rational information processing components as well as a noise component. Public information is incorporated contemporaneously while private information is generated from the processing of public information and therefore incorporated sequentially. We refer to this specification as the information processing hypothesis of return volatility and test it using time series regressions. For a sample of 28 large US stocks we construct indicators of economic information arrival and investigate their relationship with measures of firm specific realized variance by use of time-series regression. Our approach allows us to test whether public information arrival is related to increases in volatility and subsequent volatility clustering.

Although firm specific news clustering exists, our results show that the clustering of news items is not enough to capture the extent of volatility clustering. In addition, we find that lagged news indicators are significant alongside contemporaneous news indicators. We interpret this effect as evidence that after the arrival of public information, additional private information is generated and incorporated into asset prices. Three robustness checks confirm that our measures of information arrival capture news relevant for a particular stock. In general, time series regressions of firm specific volatility on firm specific information arrival provide evidence that, ex-post, large changes in return volatility and successive volatility clustering are related to information arrival.

Section 2 outlines an economic framework for thinking about microeconomic sources of firm specific volatility. Next, in Section 3, we suggest a time series regression approach for testing the relationship between the arrival of economic information and firm specific return volatility. Section 4 presents our results and contrasts our findings with previous studies. We support our investigation with a series of robustness checks in Section 5, conclude in Section 6 and the appendix provides further detail.

2. A Framework for Thinking About the Microeconomic Sources of Equity Volatility

This section introduces a return specification describing sources of equity volatility. We conclude with a description of the model's intuition and its relationship with established financial information theory.

The natural starting point for our investigation is a conceptual model of unexpected changes in the stock price of the firm. Campbell and Shiller (1988) and Campbell (1991) provide a conceptual model of the drivers of unexpected returns from financial assets. Let r_t be the log return and d_t the log dividend received by owning the asset from time $t-1$ to t, then the following expression specifies unexpected returns:

$$r_t - E_{t-1}(r_t) = (1-\rho)\sum_{j=0}^{\infty} \rho^j (E_t - E_{t-1})(d_{t+1+j}) - \sum_{j=0}^{\infty} \rho^j (E_t - E_{t-1})(r_{t+1+j}) \qquad (1)$$

where E is the expectations operator. r is the continuously compounded return. d is the log dividend per share. The difference between the realized return and the expected return at time $t-1$ is the surprise in returns, the unexpected return. The difference between expectations at time t and the expectations at time $t-1$ is the revision in expectations, the change in expectations $E_t - E_{t-1}$. Based on these definitions, expression (2) below establishes that due to a simple accounting identity surprises in returns must correspond to a mix of changes in expectations of future dividends and changes in expected future returns. In short,

$$r_t - E_{t-1}(r_t) = \nu_t^d - \nu_t^r \qquad (2)$$

allows us to characterize unexpected returns as revisions in expectations related to future cash flows, ν_t^d, and changes in expected returns, ν_t^r. In this paper we are particularly interested in understanding the relationship between firm specific news item arrival and firm specific revisions in expectations about future dividends. As such it is not necessary to investigate the direct relationship between returns and news item arrival, in fact it is sufficient to investigate the relationship between news item arrival and the variance in stock returns related to firm specific revisions in expected future dividends. In this context, it is relevant to considered what implications the log-linear model has for the variance of returns. Given equation 2 the variance of returns is

$$Var(r_t) = Var(\nu_t^d) + Var(\nu_t^r) - 2Cov(\nu_t^d, \nu_t^r) \tag{3}$$

where $Var(r_t)$ represents the variance of stock returns at time t. This relationship simply says that the variance of returns has three components. The first component is related to variance of revisions to expectations about future dividends. The second component is related to the variance of revisions to expected returns. The third component is a covariance term between the first two components. The challenge for the econometrician interested in investigating the relationship between a possible indicator of revisions in expected dividends, in this case news items, and stock return variance is to isolate and measure stock return variance related to revisions in expected dividends. As will be described in the following paragraphs, in this investigation we use a simple market model to decompose total stock return variance into a component which is firm specific and a component which is common to the market. Throughout our investigation we assume that this approach is capable of isolating the variance associated with revisions in expected future dividends. In other words, we equate idiosyncratic variance from a market model with the variance of stock returns associated with firm specific revisions in expected future dividends.

Before proceeding it is important to note that this investigation uses the term information to mean facts, knowledge or intelligence. This is an important definition since it is important to distinguish it from the concept of information in an econometric sense. Although the term information set will not be used throughout the paper, it is an implicit assumption that everything is conditioned on the econometricians information set. That said it is also important to note that different types of information will be discussed. Public information will refer to facts, knowledge or intelligence that published into the public domain by news sources such as news wires, newspapers, press release wires and others. Private information will refer to information generated by investors from the processing of information in the public domain. Information processing in turn refers to the collection of facts, knowledge or intelligence and its subsequent examination, investigation, study or analysis in the context of setting new expectations for future dividends of the company in question. This study does not equate private information with inside information. Inside information may however be a source of contextual information for the processing of new public information. However this investigation does not treat inside information as a separate object of analysis.

We motivate our model by departing from equation (2) and proceed to add further detail. In this paper we consider equities, financial instruments linked directly to the economic performance of a company. Generally, when we think of information about a company's economic performance we use a broad typology with three main categories: General and macroeconomic, industry and company specific. Changes in expectations about future dividends and returns occur due to the arrival of new information. The expected return includes a constant and a contemporaneous relation with market wide returns, $r_{m,t}$, such that unexpected returns only correspond to a firm specific component. Let i denote a given company, $r_{i,t}$ the return on its common stock and $\varepsilon_{i,t}$ the unexpected firm specific return component, then

$$r_t - E_{t-1}(r_t) = r_t - \mu - \beta r_{m,t} = \varepsilon_t \tag{4}$$

where β, equal to $\frac{Cov(r_{i,t}, r_{m,t})}{Var(r_{m,t})}$, is a coefficient which measures the response of $r_{i,t}$ to the return on the chosen market portfolio, $r_{m,t}$ and μ is a constant. This corresponds to a more detailed specification of the return generating process.

Following French and Roll (1986) we include a distinction between private and public information, so that returns have a public and a private information component. Applying this conceptual breakdown and specifying the dynamics of each component, $\varepsilon_{private,t}$ and $\varepsilon_{public,t}$, we consider the following simple mixture model in terms of the firm specific return variance, for simplicity we drop the notation identifying a given company

$$\begin{aligned} \varepsilon_t &= \varepsilon_{private,t} + \varepsilon_{public,t} \\ \varepsilon_{private,t} &= \sigma_t \epsilon_{1,t} \qquad \sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \theta \sigma_{t-1}^2 \\ \varepsilon_{public,t} &= \sqrt{\left(\sum_{k=1}^{K} \delta_k n_{t,k} \right)} \epsilon_{2,t} \\ \epsilon_{1,t}, \epsilon_{2,t} &\sim iid(0,1) \end{aligned} \tag{5}$$

where $n_{t,k}$, for the time being, is defined broadly as an indicator of public information arrival. In the case where we have only one type of public information and a time invariant effect on expected returns and expected cashflows, we have

$$\varepsilon_t = \sigma_t \epsilon_{1,t} + \sqrt{\delta n_t} \epsilon_{2,t} \tag{6}$$

a specification that captures the private information processing component, σ_t, as a GARCH (1,1) process and the relationship with the arrival of public information as being contemporaneous. Any mispricing and other noise components are left entangled and unidentified in the two error terms $\epsilon_{1,t}$ and $\epsilon_{2,t}$. The conditional variance of the idiosyncratic return component will then be

$$Var(\varepsilon_t | \Omega_{t-1}, n_t, x_t) = \sigma_t^2 + \delta n_t \tag{7}$$

where x_t is a vector of exogenous information such as general market conditions as captured by the return on the market, $r_{m,t}$. Since ε_t is defined as per equation 4 its variance will also be conditioned on x_t although it does not appear explicitly on the right hand side of equation 7. Our model does not imply a causal relationship between volatility and news items. Rather, we suggest that news items and volatility share a common latent factor, new economic information.

We refer to this specification as the information processing hypothesis of return volatility. The intuition follows in two steps. First, market participants evaluate the signal inherent in new economic information and, due to intense competition between numerous informed traders, incorporate an imprecise and unbiased estimate into market prices almost immediately. Next, investors pursuing active information-based strategies conduct further economic assessments and revise their expectations and portfolios accordingly. In what can be characterized as an adaptive price discovery mechanism triggered by the arrival of public information, this behavior results in stock return volatility decreasing over time. The speed of mean reversion of the GARCH component then corresponds to the speed at which investors process and incorporate private information into prices. Lagged effects of economic information arrival will be related to information processing activities by investors.

The information processing hypothesis is in line with three mechanisms modeled by financial information theory. First, when multiple informed traders observe the exact same signal, prices may reflect new information almost instantaneously (Holden and Subrahmanyam 1992). Second, skilled processing of public information may identify profitable trading opportunities (Admati and Pfleiderer 1988). Third, a large information advantage of multiple informed traders, combined with the subsequent arrival of informative public information, may lead to decreasing return volatility on an interday basis (Foster and Viswanathan 1990).

Holden and Subrahmanyam (1992) develop a multi-period auction model with multiple privately informed traders. In this model every trader is assumed to observe the same information, interpret it similarly and trade on it in the same way. Their results suggest that new information will be incorporated almost instantaneously and that the speed at which prices incorporate information is proportional to the number of informed traders.

Admati and Pfleiderer (1988) examine how informed traders and liquidity traders interact. The information structure of their model allows the interpretation that skilled information processors will be able to identify profitable trading opportunities due to faster and more efficient processing of public information. They show that endogenous information acquisition intensifies the concentration of trading and that increased competition among informed traders increases price informativeness.

Foster and Viswanathan (1990) present a model where a single informed trader enjoys a large information advantage over non-informed traders at the start of the week and where noisy public information is made available at the end of each day. In a model where public information provides no trading signal, across the week, price change variance will be constant. On the other hand, when public information is informative, over time, return volatility will decline. The quality of public information available to non-information based traders determines the decay rate of the information advantage.

Theory is supported by anecdotal evidence from disclosure by one of the world's largest and most transparent sovereign wealth funds[1]. Disclosure regarding their information based active investment strategies suggests that processing of public information is the primary source of private information in equity markets. The disclosure further suggests that information based investors seek to process public information better than the average market participant. Investor behavior related to active investment management is important given that over the years from 2000 to 2007 less than 20% of US public equity was managed passively, as presented in French (2008).

Theory and anecdotal evidence support the specification of the information processing

[1]Appendix A provides excerpts of Norges Bank Investment Management's disclosure.

hypothesis of return volatility. To summarize, it suggests that information observed by a broad set of market participants is likely to be incorporated almost instantaneously. In addition, the arrival of public information creates a trading opportunity for investors with exceptional information processing capabilities. The information advantage is largest in the period after public information arrival and decreases with time as multiple information based investors compete with each other. The speed of variance decline may be further increased by the arrival of public information providing analysis of previously disclosed economic information. Conclusively, our specification captures the intuition that, after the arrival of public information, the change in stock return volatility will reflect changes in the rate at which private information is revealed in prices.

In a broader context, established financial research suggests that there are several possible sources of volatility. The prevailing hypothesis involves the arrival of new information. The mixture of distributions hypothesis suggests that volume and volatility are driven by a common factor which can be interpreted as the rate of information arrival (Clark 1973). If the underlying information flow is clustered this will lead to clustering in volatility. This hypothesis is general and can be thought of as subsuming several more detailed hypotheses that differ primarily in their suggested explanation for the source of volatility clustering. Possible sources of volatility clustering are: Sequential readership of public information (Copeland 1976; Copeland and Friedman 1987), clustering in private information generated by information processing in the form of signal extraction from noisy public information (Admati and Pfleiderer 1988), resolution of information asymmetry between informed and uninformed traders due to the arrival of revealing public information (Foster and Viswanathan 1990; Tetlock 2010), diminishing difference in opinion among traders about the impact of public information (Kandel and Pearson 1995), clustering in newspaper media attention (Huberman and Regev 2001; Barber and Odean 2008; Engelberg and Parsons 2011) and clustering in public information (Engle 2004). An alternate perspective interprets volatility as economic uncertainty related to the fundamental value of the underlying assets of the firm (Merton 1974; Black 1976; Christie 1982). Some theories suggesting volatility as a measure of economic uncertainty reverse the causal relation and propose that news is the result of the media producing news items to resolve uncertainty or information asymmetry (Veldkamp 2006). This hypothesis provides no predictions on the source of volatility clustering and does not predict why most news items are the result of scheduled announcements (Reuters 2006; Winkler 2011). A possible source of volatility not directly related to information is noise. Noise as a source of volatility arises from trading by non-information based traders (Black 1986). Noise therefore includes any irrational reason for trading as well as trades motivated by liquidity needs. We can consider the noise hypothesis as covering all irrational and non-information based sources of volatility.

In this light, our three distribution mixture model captures the salient hypotheses of the sources of return volatility. The specification allows us to nest all of the above hypotheses in one model. If volatility is primarily driven by public information in such a way that volatility clustering directly reflects the clustering in the public news process, then lagged effects of public news would be insignificant. On the other hand, if volatility is driven mainly by private information unrelated to public information, then contemporaneous public information should be unrelated to increases in volatility. Likewise if volatility reflects economic uncertainty in the form of information asymmetry then the relationship between contemporaneous public information and changes in volatility should be negative, reflecting the resolution of uncertainty. If adding news to the return specification reduces the persistence of past volatility, then it may be reasonable to assume that the volatility clustering phenomena is related to information arrival. This paper presents a simple empirical investigation where firm specific volatility is regressed on firm specific news. While our investigation will not enable us to say anything about the specific behavior of agents subject to public information arrival, results from time-series regressions of our two measures of interest, news and volatility, provide evidence of whether the arrival of information is related to changes in volatility.

3. Time Series Estimation of Public and Private Information Processing Components of Return Volatility

This section presents time series regression as a convenient way of testing the information processing hypothesis of return volatility. First, we discuss how time series regression allows us to estimate and test the significance of equity volatility's public and private information processing components. Next, five subsections proceed to treat the following aspects in more detail: How we measure economic information arrival, our choice of information indicators, the sample's characteristics, our procedure for model selection, and some preliminary insight on the relationship between volatility and information arrival.

We estimate the public information component in order to test whether the arrival of economic information is related to contemporaneous changes in firm specific variance. We interpret a significant proportion of covariation between realized variance and contemporaneous indicators as implying that prices are moving in conjunction with the arrival of relevant information about the firm's economic prospects.

In addition, we aim to test if private processing of public information occurs following the arrival of economic information. We argue that any lagged effects of economic information arrival will be related to information processing activities of information based investors. For example, institutional investors and investment managers pursuing active investment management strategies will process newly arrived information, evaluate its impact on company value, infer the initial impact estimate already incorporated in prices, and finally devise a trading strategy to exploit the derived signal.

Furthermore, recent research into media, news and financial markets has brought forth an alternate hypothesis for the origin of price movements and market activity. It has been suggested that a proportion of stock returns and market activity may be driven by a media effect reflecting the attention that a company achieves in the media (Huberman and Regev 2001; Barber and Odean 2008; Engelberg and Parsons 2011). In order to control for this hypothesis we include a set of indicators measuring increases in media attention. Our aim is to disentangle media attention effects from economic content effects.

We test the information processing hypotheses by approximating equation (7) with the following time series model

$$\Delta ln(FV_t) = \omega + \sum_{l=1}^{L} \rho_k ln(FV_{t-k}) + \sum_{k=1}^{K} \gamma_k n_{k,t} + \sum_{k=1}^{M} \sum_{j=1}^{J} \tilde{\gamma}_{k,j} n_{k,t-j} + \epsilon_t, \tag{8}$$

where we add indicators of public information arrival, $n_{k,t}$, that we believe may either have economic relevance or measure media attention. This empirical approach is motivated both by the aim of our investigation and the complexity arising from the number of indicators which we intend to consider in our estimation efforts. The model in equation (8) has three components of particular interest: an autoregressive part, $\sum_{l=1}^{L} \rho_l ln(FV_{t-l})$; contemporaneous news, $\sum_{k=1}^{K} \gamma_k n_{k,t}$; and lagged effects of public news, $\sum_{k=1}^{K} \sum_{j=1}^{J} \tilde{\gamma}_{k,j} n_{k,t-j}$. Our analysis of this model, center on applying linear restrictions to a subset of parameters in order to test whether a particular component is significant. For our test of linear restrictions we employ Wald statistics.

Time series regression is a convenient way of investigating how firm specific realized variance is related to the arrival of different types of economic information. Our central hypothesis is that public information is incorporated contemporaneously while private information resulting from the processing of new public information is incorporated sequentially through an adaptive mechanism where market prices gradually reveal private information. Time-series regressions provide direct evidence of whether such a dynamic relationship between realized variance and economic information arrival exists. In particular, the R^2 estimates and Wald statistics indicate whether proxies for information arrival capture common variation with firm specific realized variance.

In our analysis the variable of interest is idiosyncratic variance which captures firm specific price movements within the trading day. We call this measure Firm Specific Variance (FV) and suggest that it corresponds to a realized version of (7). We estimate it as

$$FV_{i,t} = RV_t^i - \beta_t^2 RV_t^{SPY}, \qquad \text{using} \quad \beta_t = \frac{RCov_t^{i,SPY}}{RV_t^{SPY}}, \tag{9}$$

where the individual parts are computed using intraday data so that RV_t^i, RV_t^{SPY} and $RCov_t^{i,SPY}$ represent the covariance matrix between the individual asset and the S&P 500 index as represented by the SPY exchange traded fund, ensuring FV_t is always positive. We interpret β_t as a realized beta and RV_t^{SPY} as the realized variance of the market index. Realized measures are computed by aggregating squared five minute returns within each trading day as done in Andersen, Bollerslev, Diebold, and Ebens (2001). In addition, the realized covariance matrix between the individual asset and the market index is estimated using *Refresh Time* sampling as discussed in Barndorff-Nielsen, Hansen, Lunde, and Shephard (2011). All realized measures are based on transaction prices according to the cleaning rules presented in Barndorff-Nielsen, Hansen, Lunde, and Shephard (2009).

While there are several possibilities for econometric implementation of the specification suggested in (7), time-series regression provides the flexibility necessary for our research question. The chosen approach is inspired by the ability to rewrite the specification from

(7) in terms of an autoregressive distributed lag model. This notion was first introduced in Bollerslev (1986) and discussed further in terms of the integrated GARCH model of Engle and Bollerslev (1986). Equation (8) has a similar form as an augmented Dickey-Fuller regression. If the log-transformed firm specific realized variance is non-stationary and corresponds to a random walk, then ρ_1 would be approximately zero. On the other hand, if the firm specific realized variance instead is a stationary process, then equation (8) is a valid respecification of a model in levels and the ρ_1 coefficient would be significantly negative.

3.1. Measuring Economic Information Arrival

Previous empirical work by Ederington and Lee (1993), Mitchell and Mulherin (1994), Berry and Howe (1994) and DeGennaro and Shrieves (1997) suggests that a relevant measure of public information arrival is a simple count of the number of news items. We argue that economic information arrival is best proxied by positive surprises to newsflow, a measurement approach that takes into account how economic information is transformed into news by the media industry. Our indicator corresponds to

$$n_{i,t} \equiv max(\Delta c_{i,t}, 0) \tag{10}$$

where $max(\cdot)$ is the maximum function, and $c_{i,t}$ counts from time $t-1$ to t the number of news items for a given subject or media attention category[2]. Positive changes in newsflow will reflect an increase in information arrival. A larger increase in news items will reflect a higher level of content materiality. If media industry participants reliably evaluate the materiality of new information and consistently initiate editing and distribution of news items of material economic content, then it is reasonable to assume that a positive change in the number of news items will proxy for the revisions in expectations about future dividends[3]. Our choice of indicator is also dictated by the data. Contemporaneous and lagged levels resulted in virtually the same estimated coefficients but of opposite signs.

The structure of the media industry implies that most news items are the result of information release. This is consistent with the description of the news disclosure process provided by Thompson, Olsen, and Dietrich (1987). They describe how firms typically initiate firm specific news stories through press releases and direct contact with journalists. New economic information is created by participants in the information environment

[2] All news items published after close of the exchange are transferred to the news count observation for the next trading day.

[3] Preliminary analysis of the relationship between news counts and changes in firm specific volatility also supports the use of positive changes in news counts. Figure 2 illustrates the autocorrelation functions and cross-correlation functions between firm specific news and changes in firm specific volatility. It illustrates that a reasonable specification is first differenced news counts. Separating positive and negative changes in news counts reveals that: positive changes in news counts are contemporaneously related to changes in volatility, while negative changes have a correlation of a similar magnitude for the lead relationship, $Corr(ln(FV_t), min(\Delta c_{i,t+1}, 0)))$, meaning that negative changes in news counts add little extra information. Positive changes in news counts is therefore a simple proxy for the unexpected arrival of information when news counts are aggregated across the cross section of news sources.

of a firm. These participants include: The corporation, competitors, suppliers, customers, strategic partners, government agencies, financial institutions, credit rating agencies, industry associations and other original data providers. New economic information therefore takes the form of press releases, public announcements, transcripts, fillings with the U.S. Securities and Exchange Commission (SEC) and other disclosures, economic data, reports, indicators and estimates. This is supported by the following statements from two of the largest real-time news wire services. In its journalistic handbook, Winkler (2011), Bloomberg's news agency states:

> Much of what Bloomberg writes is based on scheduled events. Companies release their earnings reports every quarter or half a year. Government departments schedule economic releases well in advance. Press conferences are announced hours or days before they happen. Elections and economic summits are scheduled long ahead of time. [p. 59]
>
> The most newsworthy releases are used as the basis for a short story, edited and transmitted as quickly as possible. Then we strive to add value through background, context, perspectives and voices. [p. 68]

Similarly, Reuters, another leading news wire service, in its journalistic handbook, Reuters (2006), states:

> A large proportion of daily news comes from events known about in advance, such as government news conferences, visits by foreign leaders, companies announcing annual results and court cases. [p. 6]

Economic information is transformed into news items by the media. The disclosure of material economic information triggers information processing and distribution activities among media industry participants. When information is released, news agencies may start by summarizing its content in a short version and instantly redistribute it to end-users. The news agency then gathers information from various sources, eliciting comments from industry experts and adding other contextual information, this results in a second distribution of relatively longer news items within a couple of hours of the first one. Successive editing and distribution based on the original information release may continue depending on its level of materiality. Often, following large corporate events, equity research and credit rating analysts publish a report containing their immediate analysis and comments to the event. Subsequently, news agencies and newswires distribute news items discussing or summarizing the contents of these reports.

Simultaneously, newspaper journalists gather news for the next issue of their publication. Some news items included in the next daily publication will reflect information that has been processed and distributed through newswires the day prior to publication. Journalists working on these news items, will add further insight by gathering more contextual information and adding further synthesis and analysis. This means that newspapers will often treat events reported by newswires the day before publication.

In summary, news items are the result of the activities of media industry participants as they edit, aggregate and distribute raw economic information. Media industry participants choose the degree that items are edited and aggregated in order to fit the medium's distribution frequency (i.e. continuously, daily, weekly, etc.) and distribution form (e.g. electronic or print). As a result, positive changes in firm specific newsflow, measured in the full cross-section of news sources, is a simple indicator reflecting the arrival of material unexpected information.

3.2. **Choice of indicators of economic information arrival**

Our choice of information arrival indicators is linked directly to the economic rationale inherent in our model describing the microeconomic sources of equity volatility, equation (3). This model states that unexpected returns may arise from changes in expected returns and changes in expected future cash flows. Our chosen variables are all considered indicators of the arrival of new economic information related to changes in expected returns or expected cashflows. We limit our choice of indicators of economic information arrival to a set of 83 subject categories presented in Table 1. These categories reflect a mutually exclusive set of corporate information categories covering an extensive set of corporate information events. In this way news response coefficients can be partially disentangled into economic content effects as suggested in Engelberg and Parsons (2011). The subject categories are, in a taxonomic sense, mutually exclusive, however this does not preclude that they may occur simultaneously or on the same day.

To control for the media attention effect, we include the 16 indicators displayed at the bottom of Table 1. These indicators measure the change in the level of news item counts in newspapers and other types of source formats. While the ideal measure would entail a measure of investor readership, we argue that our chosen variables control for the media attention effect, since they proxy for the distribution and proliferation of company specific news. The media attention hypothesis and the information processing hypothesis are not mutually exclusive and may well coexist. Hence we include proxies for media attention by measuring changes in the arrival of news. This allows us to explicitly test the importance of aggregate measures of media attention alongside proxies for different corporate information events.

Variation in the information content of news items suggests the relationship between market impact and news may vary depending on the nature of the news. We use the Dow Jones Intelligent Indexing categorization to distinguish between the content of news items. For example, one news item related to a corporate merger may have a larger impact than a news item related to the launch of a new product. The information typology used by Roll (1988) provides a starting point for thinking about news categorization. General and macroeconomic information may include facts about the state of consumer, industrial, and other markets as well as news related to politics, government policy, regulation and demographics. Industry specific information may include news about access to resources, mergers and acquisitions, and other industry information events. Firm specific news may include information about corporate issues and activities, such as: research and develop-

Table 1: Choice of News Variables

This table provides an overview of the subject categories from the Dow Jones Intelligent Indexing system that we use to create content specific indicators of new economic information arrival. When creating indicators for a given company, we include only indicator series that have news arrival on at least 1% of the days in the sample period. Dow Jones Factiva provides more information on their Dow Jones Intelligent Indexing system on the following webpage: http://www.factiva.com/content/indexing/indexing.asp.

Main DJII Code & Category Name	**# Selected Variables**	**List of subject category names**
GCAT-General / Political	12	Crime/Courts, Disasters/Accidents, Environmental News, Climate Change, Global/World Issues, Health, Int. Relations, Domestic Politics, Regional Politics, Science/Technology, Weather, Labor Issues, Demographics
ECAT-Economic	27	Economic Growth, Money Supply, Inflation/Prices, Personal Income/Average Earnings, Consumer Sentiment, Consumer Credit/Expenditure/Savings, Budget Account, Government Borrowing Requirement, Agricultural Production, Industrial Production, Capacity Utilization, Inventories, Factory Orders/Durable Goods, Employment Costs/ Productivity, Employment/Unemployment, Business Sentiment, Reserve Assets, Mortgage Applications/Refinancing, Car Registrations/Vehicle Sales, Bankruptcy Figures, Index of Leading Economic Indicators, Housing Starts/Construction Figures, Home Sales/Housing Affordability, Economic/Monetary Policy, Government Finance, Trade/External Payments, Euro Zone/Currency
MCAT-Financial / Commodity Markets	8	Money Markets, Foreign-Exchange News, Soft Commodity Markets, Metals Markets, Energy Markets, Fund Markets,
CCAT-Corporate / Industrial	36	Plans/Strategy, Corporate Crime/Legal/Judicial, Regulation/Government Policy, Annual Meetings, Dividends, Sales Figures, Earnings Surprises, Analyst Comment/Recommendation, Internal Audit, Bankruptcy, Share Capital, Corporate Debt Instruments, Financing Agreements, Corporate Credit Ratings, Acquisitions/Mergers/Takeovers, Divestitures/Asset Sales,Privatizations/Nationalization, Joint Ventures, Output/Production, New Products/Services, Research/Development, Capacity/Facilities, Information Technology, Product Safety, Marketing/Market Research, Government Contracts, Defense Contracts, Non-government Contracts, Licensing Agreements, Franchises, Outsourcing, Competition Issues, Management Issues, Labor Disputes, Lay-offs/Redundancies, Natural Reserves/Resources Discovery
Economic Information Indicators	*83*	
Media Attention Indicators	*16*	Dow Jones News Service, Reuters Newswire, Washington Post, New York Times, Wall Street Journal, Associated Press Newswires, USA Today, Financial Times, Aggregate Newspaper, Aggregate Newswire, Aggregate Industry Publication, Aggregate Newsletters, Aggregate Press Release Wire, Aggregate Other Wire, Aggregate General News and Business Publication, Aggregate Government and Politics Publication
Total	**99**	

ment, restructuring, litigation and arbitration, business/franchise performance, funding, as well as other corporate information events. With respect to Roll (1988)'s information typology the news in our dataset mainly covers industry and firm specific news items.

Since we aim to estimate the proportion of stock return volatility that is related to the arrival of economic information we are concerned with the issue of news item endogenity. Endogenous news are news items that include references to stock market information such as trading volume and price change alerts. We are only interested in news items that contain reference to some form of economic information related to the firm. In other words, we wish to filter out items that are automatically generated based on market activity. Other studies have approached this issue using filtering rules. Of all news considered to contain discussion of share price movements a subset also discusses events in the broad equity markets and the rest discuss events in other financial markets (steel, oil, natural gas, etc.). We therefore choose to remove all news that the Dow Jones Intelligent Indexing system labels as covering events in equity markets (M11), contains stock market pricing information (C1522) and does not contain any other type of economic information, in the form of a reference to other corporate information events. Our examination of market news items reveals that purely endogenous news items, created due to abnormal market activity, account for approximately 5% of the corporate newsflow and is more prevalent among a specific set of news sources. We find that news containing references to price movements can either be purely endogenous, triggered by trading activity, or include a summary or reference to corporate information events. The Dow Jones Intelligent Indexing system enables identification of the topics covered in an individual news item, hence we are able to filter out purely endogenous news and focus on news items that provide economic content.

3.3. **Data**

This subsection describes the characteristics of the data used in the investigation. We employ data from five databases covering the period from January 2001 through July 2009 for a sample of 28 large US companies. Table 2 provides an overview of the characteristics of our sample. Table 2 illustrates that the sample companies on average have 57.1% of revenues from their largest business unit, while 57.5% of revenues are from North American operations. In addition, the average market capitalization over the sample period across firms is $128.0 billion. These characteristics underscore the large size, business diversification and international activities of companies in the sample.

Intraday data for each company's common stock are extracted from The New York Stock Exchange Trade and Quote (TAQ) database. Firm specific news items are collected from the Dow Jones Factiva database based on the Dow Jones company code which we match with CUSIP identification codes as well as permanent company and security identification numbers from the Center for Research in Security Prices (CRSP) database. Daily returns and trading volume used for robustness investigations are from CRSP. For our robustness analysis we also use implied volatilities, these are from Bloomberg and based on 1 month at-the-money call options. Company specific information used in our discussion section comes from sources such as SEC fillings and other reference data extracted from

Standard & Poor's Capital IQ database. The sample's characteristics are summarized in Table 3 and Table 4. Table 3 presents the characteristics of our measure of firm specific realized variance while Table 4 describes characteristics of our news item dataset. Our sample starts after the decimalization occurred on the NYSE and NASDAQ exchanges. NYSE completed decimalization on the 29th of January 2001 while NASDAQ completed the transition on the 9th of April 2001. Alternatively the sample starts from when the last major corporate event occurred as marked by a name change, merger, initial public offering, or similar. As can be seen from Table 3, this results in 919 to 2,137 observations per company.

Table 3 also shows that our variable of interest, firm specific variance, generally accounts for between 58 and 77 percent of all intraday volatility in a firm's common stock. This is consistent with results from market models at lower frequencies, e.g. using daily data. For example, Roll (1988) finds that the market model's average R^2 across ninety-six large firms is 24%, corresponding to idiosyncratic return variability of 76%. The mean realized variance across stocks corresponds to annualized volatility in the range from 19.6 for Procter and Gamble (NYSE:PG) to 48.5 for Citigroup (NYSE:C). On the other hand, mean firm specific realized variance translates to annualized volatility in the range from 15.7 for PG to 38.0 for C.Mean realized betas correspond to expected magnitudes. We expect companies operating in cyclical industries to have a relatively high beta while companies in non-cyclical industries would have relatively lower betas. For example, we find that financial institutions such as Bank of America (NYSE:BAC), Citigroup (NYSE:C), and J.P. Morgan Chase (NYSE:JPM) have betas that generally lie in the top end of the range across stocks. However, these are the stocks with the most variability in their realized betas and realized variances which reflects the sample's inclusion of the 2007-2009 financial crisis. Looking across the remaining stocks we see that Intel has a high beta while Healthcare companies such as Pfizer (NYSE:PFE), Johnson & Johnson (NYSE:JNJ), and Merck (NYSE:MRK) have betas in the lower end of the range.

Table 4 contains descriptive statistics for our measure of aggregate firm specific news over the sample period. The table shows the number of daily news items for each company. Our dataset contains all news items identified by the Dow Jones Intelligent Indexing system as being related to the particular firm. It may also include news where the news agency has specifically marked news as being of importance to the firm. Our sample includes news items from 8,581 different news sources and more than 3.8 million news items for the 28 companies.

Table 2: Sample Description

The table describes some characteristics of the companies in our sample. We identify the largest business units by calculating mean percentage revenue contribution over the sample period. In addition, we present the percentage revenue contribution of North American or US revenues to total firm revenues excluding corporate adjustments. These two measures provide an overview of the activities which the individual firm is involved in. The table is based on SEC fillings obtained through Standard and Poors CapitalIQ. The last four columns correspond to: **(1)** Avg. % of Total Revenues of Largest Business Segment, **(2)** Name of the Largest Geographic Revenue Segment, **(3)** Avg. % of Total Revenues of the Largest Geographic Revenue Segment, and **(4)** Avg. Market Capitalization in $ billion over the years 2001 to 2009.

Company Name	**Exchange**	**Ticker**	**Headquarters**	**Largest Business Segment**	**(1)**	**(2)**	**(3)**	**(4)**
Exxon Mobil Corp.	NYSE	XOM	Irving, TX, US	Downstream - Non-United States	70	US	31	349.2
General Electric Co.	NYSE	GE	Fairfield, Connecticut, US	Infrastructure	34	US	55	310.8
Microsoft Corporation	NASDAQ	MSFT	Redmond, WA, US	Windows & Windows Live	30	US	65	285.4
Wal-Mart Stores Inc.	NYSE	WMT	Bentonville, AR, US	Walmart United States	67	US	79	214.9
Citigroup, Inc.	NYSE	C	New York, NY, US	Institutional Clients	51	North America	43	190.7
Pfizer Inc.	NYSE	PFE	New York, NY, US	Biopharmaceutical	91	US	54	187.9
Johnson & Johnson	NYSE	JNJ	New Brunswick, NJ, US	Pharmaceutical	43	US	56	177.3
Procter & Gamble Co.	NYSE	PG	Cincinnati, OH, US	Household, Fabric and Home Care	29	US	45	164.0
IBM Corp.	NYSE	IBM	Armonk, NY, US	Global Services	54	US	38	152.6
Bank of America Corp.	NYSE	BAC	Charlotte, NC, US	Consumer and Commercial Banking	57	US	92	146.8
Intel Corporation	NASDAQ	INTC	Santa Clara, CA, US	PC Client Group	73	US	20	143.3
AT&T, Inc.	NYSE	T	Dallas, TX, US	Wireline	78	US	100	134.7
Chevron Corporation	NYSE	CVX	San Ramon, CA, US	Downstream	85	US	51	128.9
JPMorgan Chase & Co.	NYSE	JPM	New York, NY, US	Investment Bank	34	US	74	118.8
The Coca-Cola Company	NYSE	KO	Atlanta, Georgia, US	Non-Alcoholic Beverages	100	North America	29	115.4
Verizon Comm. Inc.	NYSE	VZ	New York, New York, US	Wireline	54	US	99	105.9
Merck & Co. Inc.	NYSE	MRK	Whitehouse St., NJ, US	Pharmaceutical	92	US	59	100.2
Hewlett-Packard Comp.	NYSE	HPQ	Palo Alto, CA, US	Personal Systems Group	30	US	37	84.3
The Home Depot, Inc.	NYSE	HD	Atlanta, GA,US	Retail - Home Improvement	99	US	95	72.4
3M Co.	NYSE	MMM	St Paul, MN, US	Industrial and Transportation	28	US	40	55.5
American Express Comp.	NYSE	AXP	New York, NY, US	United States Card Services	47	US	72	54.9
United Technologies Corp.	NYSE	UTX	Hartford, CT, US	Carrier	28	US	53	52.0
Walt Disney Co.	NYSE	DIS	Burbank, CA, US	Media Networks	41	US and Canada	79	51.2
McDonald's Corp.	NYSE	MCD	Oak Brook, IL, US	Food Service Industry	100	US	35	47.8
Boeing Co.	NYSE	BA	Chicago, IL, US	Commercial Airplanes	48	US	65	44.3
EI DuPont de Nemours & Co.	NYSE	DD	Wilmington, DE, US	Agriculture & Nutrition	23	US	42	40.2
Caterpillar Inc.	NYSE	CAT	Peoria, IL, US	Machinery and Engines	78	North America	45	31.5
Alcoa, Inc.	NYSE	AA	Pittsburgh, PA, US	Primary Metals	36	US	58	24.1
				avg. =	57.1		57.5	128.0

Table 3: Characteristics of Firm Specific Realized Variance

This table illustrates descriptive statistics for our sample. The end date is July 31, 2009 for all stocks. In our analysis the variable of interest is idiosyncratic variance which captures firm specific price movements within the trading day. We call this measure Firm Specific Variance (FV) and suggest that it corresponds to a realized version of (7). We estimate it as

$$FV_{i,t} = RV_t^i - \beta_t^2 RV_t^{SPY}, \qquad \text{using} \quad \beta_t = \frac{RCov_t^{i,SPY}}{RV_t^{SPY}},$$

RV_t^{SPY} and $RCov_t^{i,SPY}$ represent the covariance matrix between the individual asset and the index as represented by the SPY exchange traded fund, ensuring FV_t is always positive. We interpret β_t as a realized beta and RV_t^{SPY} as the realized variance of the market index. Realized measures are computed by aggregating squared five minute returns within each trading day as done in Andersen et al. (2001). In addition, the realized covariance matrix between the individual asset and the market index is estimated using *Refresh Time* sampling as discussed in Barndorff-Nielsen et al. (2011). All realized measures are based on transaction prices according to the cleaning rules presented in Barndorff-Nielsen et al. (2009).

	RV		**FV**		β		**SPY**		**FV/RV**		**Start**	**#**
Ticker	μ	σ^2	μ	σ^2	μ	σ^2	μ	σ^2	μ	σ^2	**Date**	**obs**
AA	5.84	114.85	3.73	30.78	1.08	.23	1.31	7.09	.75	.02	1/29/01	2,137
AXP	5.23	108.87	2.98	35.37	1.05	.21	1.31	7.09	.65	.03	1/29/01	2,136
BA	3.29	21.09	2.13	5.78	0.89	.09	1.31	7.09	.72	.03	1/29/01	2,136
BAC	7.04	635.13	3.70	161.56	1.00	.45	1.31	7.09	.64	.03	1/29/01	2,136
C	9.34	1308.2	5.74	530.02	1.18	.34	1.31	7.09	.62	.02	1/29/01	2,136
CAT	3.58	32.90	2.06	6.02	1.02	.14	1.31	7.09	.68	.03	1/29/01	2,136
CVX	2.57	28.04	1.38	3.00	0.84	.14	1.28	7.51	.69	.04	10/10/01	1,963
DD	2.97	20.75	1.64	3.24	0.94	.08	1.31	7.09	.65	.03	1/29/01	2,136
DIS	3.71	45.83	2.36	19.50	0.89	.10	1.31	7.09	.72	.03	1/29/01	2,136
GE	3.78	71.19	1.98	20.06	0.98	.10	1.31	7.09	.60	.03	1/29/01	2,136
HD	3.87	31.36	2.28	6.69	1.04	.12	1.31	7.09	.68	.02	1/29/01	2,136
HPQ	3.82	29.76	2.52	11.29	1.02	.11	1.31	8.06	.71	.02	5/06/02	1,822
IBM	2.28	13.98	1.24	2.73	0.87	.05	1.31	7.09	.62	.02	1/29/01	2,136
INTC	4.46	27.64	2.34	6.01	1.39	.18	1.29	7.17	.58	.02	04/09/01	2,087
JNJ	1.55	6.89	1.06	2.87	0.56	.06	1.31	7.09	.75	.02	1/29/01	2,136
JPM	6.14	194.08	3.39	54.18	1.16	.23	1.31	7.09	.63	.03	1/29/01	2,136
KO	1.66	5.99	1.09	1.62	0.62	.06	1.31	7.09	.74	.02	1/29/01	2,136
MCD	2.58	14.49	1.85	5.95	0.75	.09	1.31	7.09	.76	.02	1/29/01	2,136
MMM	2.10	11.66	1.22	2.74	0.81	.06	1.31	7.09	.65	.03	1/29/01	2,136
MRK	2.80	20.52	1.98	8.76	0.71	.11	1.31	7.09	.77	.02	1/29/01	2,136
MSFT	2.75	14.78	1.40	2.71	1.02	.10	1.29	7.17	.60	.02	04/09/01	2,087
PFE	2.48	11.37	1.65	3.55	0.77	.09	1.31	7.09	.74	.02	1/29/01	2,136
PG	1.53	7.32	0.98	1.61	0.63	.05	1.31	7.09	.72	.02	1/29/01	2,136
T	3.43	38.86	1.86	8.29	0.88	.10	1.74	14.10	.68	.03	12/1/05	919
UTX	2.55	18.68	1.51	5.05	0.86	.08	1.31	7.09	.67	.03	1/29/01	2,137
VZ	2.88	19.52	1.78	5.19	0.83	.09	1.31	7.09	.71	.03	1/29/01	2,137
WMT	2.23	10.36	1.34	3.25	0.84	.06	1.31	7.09	.66	.02	1/29/01	2,136
XOM	2.48	43.34	1.28	5.28	0.89	.11	1.31	7.09	.64	.03	1/29/01	2,136

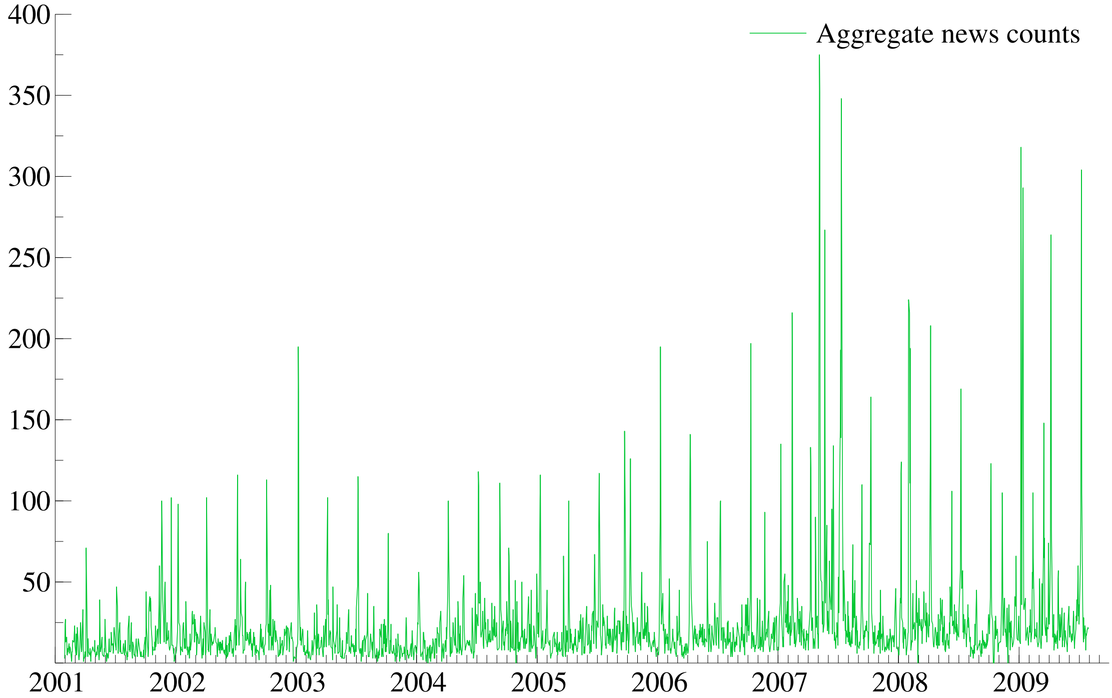


Figure 1: Aggregate News Counts for Alcoa (NYSE:AA) January 29th, 2001 to July 31, 2009. The chart illustrates the aggregate newsflow of Alcoa Inc in terms of number of news items registered in the Dow Jones Factiva database, using the Dow Jones Intelligent Indexing code for Alcoa.

Figure 1 illustrates the aggregate firm specific newsflow for Alcoa (NYSE:AA) across the full cross-section of news sources accessible from the Dow Jones Factiva Database. Each company's newsflow is composed of news items from a wide group of sources[4], between 1,700 and 4,700 individual sources per stock over the sample period[5]. A review of the top ten sources in terms of aggregate news counts, reveals three key newswires reappear consistently across the sample: Associated Press Newswires, Dow Jones News Service and Reuters News[6]. The remaining top news sources vary considerably, but tend to reflect characteristics of the particular company. The presence of several local news publications reflects the location of a company's corporate headquarters or a large local presence, examples include *SeattlePI.com* for Boeing[7], *Peoria Journal Star* for Caterpillar, *Charleston Gazette* for Dupont[8], *The Atlanta Journal* for Coca Cola, *The Cincinnati Post* for Procter & Gamble, and *The Arkansas Democrat Gazette* for Walmart.

Similarly, industry specific sources are also among the ten largest contributors to the firm specific newsflow: *Metal Bulletin News Alert Service* for Alcoa; *Moody's Ratings Delivery Service* for various financial institutions (AXP, BAC, C, JPM) and corporations with large financing arms (CAT, GE); *Health & Medicine Week*, *Biotech Week*, *Pharma*

[4]The database provides access to approximately 30'000 sources.

[5]Hansen (2012) illustrates how the number of sources change over time for each stock.

[6]The sourcecode identifiers are, APRS, DJ and LBA.

[7]Boeing was founded in and is one of the largest employers in the Seattle area.

[8]Dupont has a large manufacturing plant in Parkersburg, West Virginia.

Business Week and *Drug Week* for healthcare companies such as Pfizer, Johnson & Johnson and Merck; *Upstream* and *The Oil Daily* for Exxon and Chevron; *TR Daily* and *TR's State Newswire* for telecommunication services companies such as AT&T and Verizon; *FedBizOpps* for companies regularly engaging as contractors with the US Government such as Boeing and United Technologies Corporation; *The Grocer* for Walmart and *Just-Drinks* for Coca Cola.
Table 4 also provides insight on general timeliness characteristics of aggregate news items. While we observe a consistent pattern of less news on holidays and weekends as opposed to on trading days, the intensity of aggregate news varies substantially across companies. The language of the ten largest sources also varies across the stocks. Although most of the ten largest news sources are in English, there are sources in Chinese (Traditional and Simplified), French, German, Spanish, Swedish, Norwegian and Japanese[9].

In Table 4 we also include a measure of the contribution of the largest news sources to total newsflow. The measure is computed as a Gini coefficient using information on the news contribution of each individual source to the total newsflow over the sample period. A Gini coefficient of 0.85 for Alcoa (NYSE:AA) implies that 85% of the firm specific newsflow is contributed by the 15% largest news sources. We can think of the Gini coefficient as a concentration ratio, a high Gini coefficient implies that the firm specific newsflow is concentrated among relatively fewer sources. As a reference point, a Gini coefficient of 0.50 implies that each source contributes equally to the total newsflow.

Across the sample of companies the concentration ratio ranges from 80% to 90%. This pattern is partially due to the differences in distribution frequencies across the different sources. For example, there are relatively fewer newswires but they account for a larger proportion of the newsflow on an aggregate level. The pattern may be reinforced by the specialization of the different sources in terms of their coverage and content focus. For example, financial wires and industry publications are more likely to cover companies on a continuous basis than mainstream media. Using the Dow Jones Broadtape wire and the Wall Street Journal as the object of study, Thompson et al. (1987) provide a detailed description of how wire originated news is re-edited and distributed by newspapers. They suggest that newspapers screen corporate news based on general importance and interest, leading to lower news item frequencies than newswires.

[9] Hansen (2012) shows how news items are distributed in terms of languages and sources.

Table 4: News Dataset Characteristics

This table provides an overview of our news data. Number of sources are computed based on the unique number of source codes available in our sample. News items are computed based on the unique items in our sample identified by headline, date, timestamp, sourcecode and wordcount. Gini refers to the Gini coefficient computed based on the amount of unique news items per sourcecode for a given company. The gini coefficient measures concentration of news among observed sources and is calculated as $G = 1 - \frac{2}{n-1}\left(n - \frac{\sum_{i=1}^{n} i y_i}{\sum_{i=1}^{n} y_i}\right)$, where $y_i \leq y_{i+1}$. Trading, Holiday and Weekend refer to the mean daily number of news items over the sample period occurring on each type of day.

Ticker	**Start Date**	**# News Sources**	**# News Items**	**Gini (%)**	**Trading** μ (s.e.)	**Holiday** μ (s.e.)	**Weekend** μ (s.e.)
AA	1/29/01	1,922	54,056	85	14.1 (0.5)	8.5 (1.2)	2.4 (0.1)
AXP	1/29/01	2,799	62,755	80	15.4 (0.3)	9.0 (1.0)	3.7 (0.2)
BA	1/29/01	3,529	220,865	87	55.3 (1.0)	34.5 (3.0)	12.0 (0.5)
BAC	1/29/01	3,143	169,165	86	45.5 (1.3)	19.5 (2.3)	8.4 (0.4)
C	1/29/01	3,425	281,489	88	72.7 (1.5)	39.2 (3.9)	12.8 (0.5)
CAT	1/29/01	1,682	30,241	83	7.8 (0.2)	3.9 (0.4)	2.0 (0.1)
CVX	10/10/01	2,777	98,926	85	25.9 (0.5)	14.7 (1.6)	4.4 (0.2)
DD	1/29/01	2,292	41,076	81	10.5 (0.2)	6.9 (0.7)	1.9 (0.1)
DIS	1/29/01	3,618	163,030	87	38.8 (0.7)	27.4 (2.3)	14.5 (0.4)
GE	1/29/01	4,716	261,782	86	66.1 (1.1)	48.4 (4.4)	15.7 (0.5)
HD	1/29/01	1,668	45,668	83	11.2 (0.4)	5.8 (0.7)	3.1 (0.1)
HPQ	5/06/02	3,328	137,886	84	35.3 (0.8)	23.2 (2.5)	6.6 (0.3)
IBM	1/29/01	3,799	221,442	86	56.2 (1.0)	38.7 (3.6)	9.1 (0.4)
INTC	4/09/01	3,304	208,638	87	53.6 (1.1)	31.2 (3.1)	8.5 (0.3)
JNJ	1/29/01	3,130	112,468	86	27.9 (0.6)	20.9 (2.1)	6.3 (0.2)
JPM	1/29/01	3,437	316,957	90	80.1 (1.4)	50.5 (4.8)	14.1 (0.5)
KO	1/29/01	2,974	85,944	81	21.5 (0.4)	14.4 (1.4)	5.7 (0.2)
MCD	1/29/01	3,303	83,501	80	20.3 (0.4)	14.4 (1.5)	6.3 (0.2)
MMM	1/29/01	1,656	25,316	83	6.7 (0.2)	3.9 (0.5)	1.2 (0.1)
MRK	1/29/01	1,795	48,004	84	12.8 (0.4)	7.7 (0.8)	2.2 (0.1)
MSFT	1/29/01	4,493	407,804	87	102 (1.8)	63.6 (5.3)	22.6 (0.8)
PFE	1/29/01	2,806	114,728	86	29.4 (0.7)	18.7 (1.9)	6.1 (0.3)
PG	1/29/01	3,323	97,192	83	24.0 (0.6)	17.4 (1.7)	5.8 (0.4)
T	12/1/05	1,868	61,450	86	16.7 (0.5)	7.3 (1.3)	2.9 (0.2)
UTX	1/29/01	2,184	54,063	87	14.1 (0.3)	8.5 (0.8)	2.4 (0.1)
VZ	1/29/01	2,468	134,148	88	34.1 (0.7)	17.2 (1.6)	6.1 (0.2)
WMT	1/29/01	3,206	173,795	86	41.7 (0.8)	26.3 (2.2)	13.2 (0.4)
XOM	1/29/01	3,373	181,523	86	46.5 (0.9)	27.9 (2.6)	7.9 (0.3)

The timeliness of news items deserves a final remark before proceeding. We observe a general trend in the number of average daily news items and find that certain months have higher means of daily news items due to quarterly earnings announcements. This has implications for our time series regressions. We include a set of dummy variables and a time trend to account for effects from increasing overdispersion in aggregate news items over time[10].

[10]These variables are rarely significant when news is included in the model. This is consistent with the results in Andersen and Bollerslev (1998), suggesting that day of the week dummies imperfectly substitute for announcement effects such as unanticipated news.

3.4. **Model selection**

With our chosen set of explanatory variables there are a vast range of possible models we could settle upon. To ensure model selection is free from our subjective judgement, we resort to the general-to-specific model selection procedure as surveyed in Campos, Ericsson, and Hendry (2005) and implemented in Doornik (2009). The approach builds on the work of Hoover and Perez (1999), Krolzig and Hendry (2001), and Hendry and Krolzig (2005) in response to the experiments carried out by Lovell (1983).

Our use of Doornik (2009)'s Autometrics procedure is an effort to keep our model selection procedure as objective as possible[11].

Doornik (2009)'s approach, Autometrics, uses a path search through the space of possible models determined by the variables introduced in the initial model specification. The starting point for the model reduction procedure is a general unrestricted model with the full information set. The model has been specified in order to be a statistically well behaved process. Each insignificant variable in the initial model is a possible reduction path. The first path corresponds to removing the variable with the lowest absolute t-value. The process continues in this way until we reach a terminal model. Along the way, each model is subjected to a series of encompassing and diagnostic tests. If model reduction fails the current model is considered a terminal model, and the model selection procedure continues. The procedure may arrive at multiple terminal models in which case we use an information criterion to choose the model that best fits the data. Doornik (2009)'s approach is focused on efficiency and therefore uses various approaches for reducing the number of models that need to be estimated and evaluated, so that it is unnecessary to consider every possible path.

It is important to explain the way we conduct estimation of the various sub-models we compare in this paper. Whenever we estimate a particular specification Autometrics gets to select the model all over. This implies that the terminal models that we end up with may not be nested, even though the initial models was needed. This explains why it can happen that the R-squared may decrease when going from a smaller to a larger nesting initial model.

3.5. **Preliminary Analysis of the Relation Between News and Price Changes**

In this subsection we study the unconditional relationship between changes in volatility and different types of news variables.

Using our measure of realized firm specific variance we compute sample cross-correlations between the changes in log transformed firm specific variance and positive changes in the level of news counts. In order to reduce the dimensions of this analysis we summarize the results across stocks. Figure 2 presents autocorrelation functions of the levels of aggregate newswires, aggregate newspapers, and log transformed firm specific realized variance, $ln(FV_t)$. In addition Figure 2, in the panels above and below the diagonal, illustrates

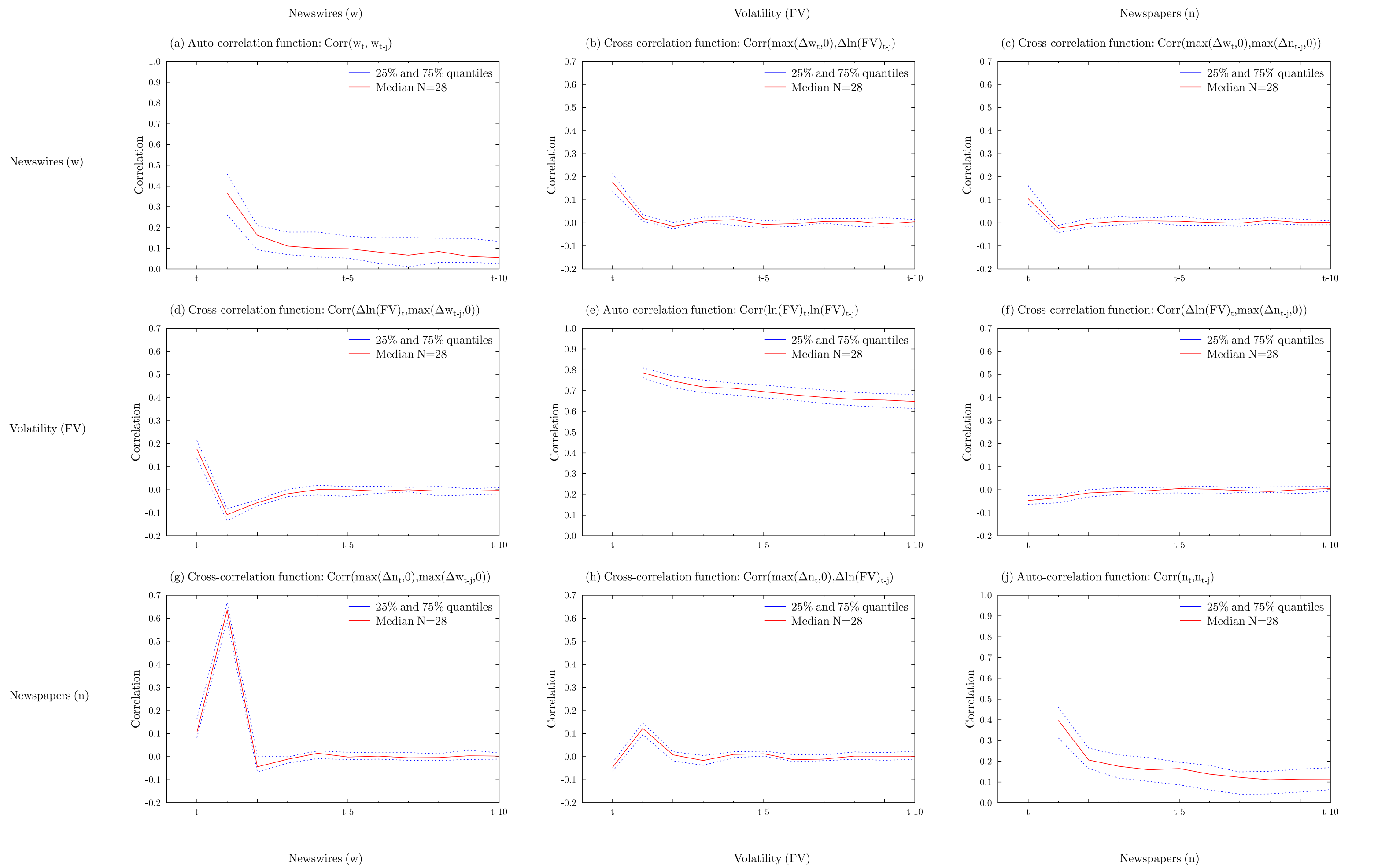
[11] We have also approached the estimation procedure using the Least Absolute Shrinkage and Selection Operator (LASSO) implemented using the Least Angle Regression algorithm (LARS). This approach reached similar results.

the six cross-correlation functions for the three series. Figure 3 provides more detail by presenting sample cross-correlations between changes in $ln(FV_t)$ and a set of selected newspapers and financial wire services previously used in empirical studies. For the Dow Jones News Service and Reuters Newswire we observe the largest relation, median correlations of 0.16 and 0.14, on a contemporaneous basis. It is interesting to contrast this finding with the results of Groß-Klußmann and Hautsch (2011). They examine the impact of news on high frequency returns and other measures of market activity, they find that a sentiment measure, based on Reuters news items, is related to significant unconditional reactions in returns, volatility and liquidity. Their study suggests that the impact occurs immediately after information arrival and that the volatility response is persistent, lasting until the fifth minute after the event. For the Deutsche Mark-US Dollar exchange rate, Andersen and Bollerslev (1998) find that macroeconomic announcements impact intraday volatility between 60 to 160 minutes after the announcement on the Reuters newswire. The results in this investigation suggest that the arrival of unexpected firm specific news is related to changes in realized variance up to several days after arrival.

Figure 3 also suggests lagged information effects may exist. The lagged effects for both newswires is approximately half the size of contemporaneous correlations but opposite in sign. There are interesting parallels between this result and the findings of Ederington and Lee (1993) and Patell and Wolfson (1984). Ederington and Lee (1993) examine the impact of scheduled macroeconomic news announcements on interest rate futures and foreign exchange futures. They find that the main price adjustment occurs within the first minute and that volatility is above normal for approximately 15 minutes and slightly elevated for several hours. The speed of price adjustment is faster than what Patell and Wolfson (1984) find in equity markets, where price change variance remained elevated into the next trading day. Ederington and Lee (1993) argue that volatility persistence may arise either from continuous trading based on the initial information, as the market works out its implications, or from price reactions to details of the information release as these are discovered by market participants. In other words, Ederington and Lee (1993) suggest that a possible explanation of lagged information effects, as we observe for newswires in panel (d) in Figure 2 and panels (e) and (f) in Figure 3, is information processing behavior by investors in their efforts to determine the full implication of new public information.

In addition, Figure 2 and Figure 3 illustrate that positive changes in the number of newspaper items correlate with the previous day's change in realized firm specific variance. In Figure 2 this is observed by looking at panel (h), which illustrates a spike in cross-correlation between aggregate newspapers at time t and the changes in firm specific variance at time $t-1$. In Figure 3 this can be seen by examining panels (a), (b), (c) and (d), each presenting an increase in cross-correlation between leading newspapers at time $t+1$ and changes in volatility at time t. A reason for this pattern is proposed by Thompson et al. (1987), they find that newspaper items that are first distributed by wire predominantly appear with a one-day delay, while items that first appear in the newspaper, usually are transmitted on the same day. 45.6% of the later items are labeled as containing *forecasts* or *analysis*, emphasizing that newspapers focus on providing in depth

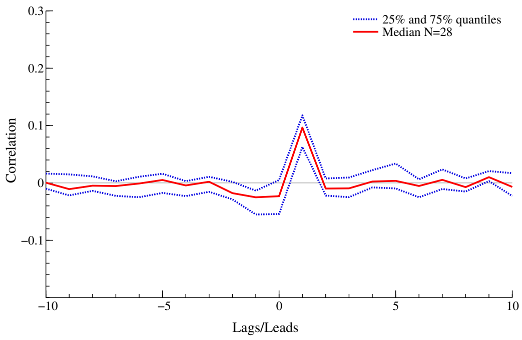
Figure 2: Auto and cross-correlation functions between firm specific return variance and news counts



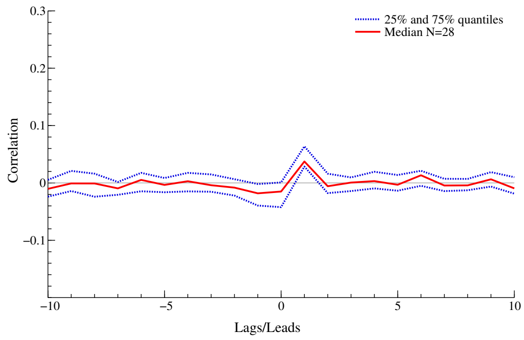
Notes: This panel illustrates auto and cross-correlation functions between changes in Firm Specific Realized Variance (FV) and two indicators of news arrival. The diagonal presents the autocorrelation functions of the levels (panel a, e, and j) while the upper and lower parts of the panel illustrate cross-correlation functions for changes in log-transformed FV ($\Delta ln(FV_t)$) and the positive change in news counts for aggregate newswires, $max(\Delta w_t, 0)$, and aggregate newspapers, $max(\Delta n_t, 0)$.

Figure 3: Cross-correlations between firm specific return variance and news counts

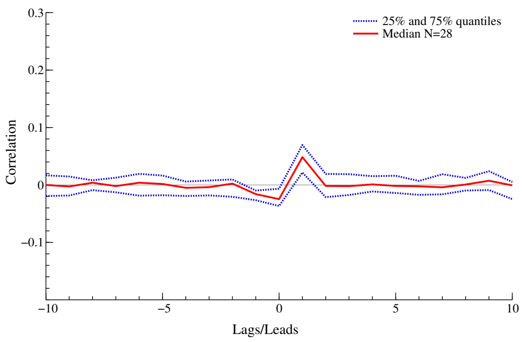
(a) The Wall Street Journal



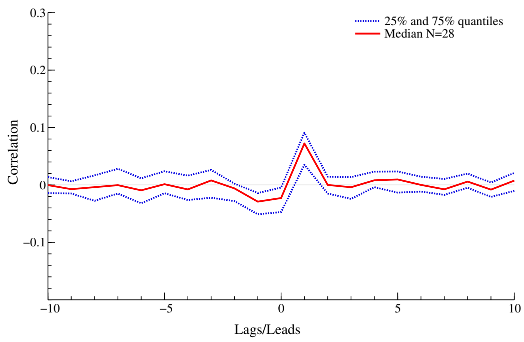
(b) The Washington Post



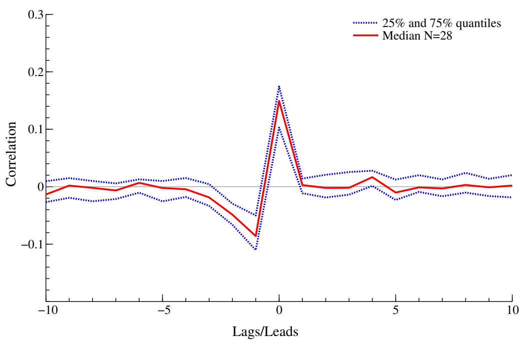
(c) USA Today



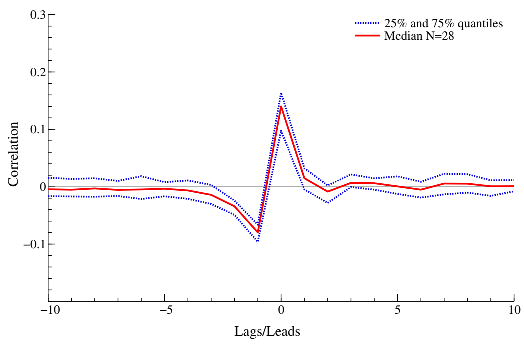
(d) The New York Times



(e) Dow Jones Newswire



(f) Reuters Newswire



Notes: This panel illustrates cross-correlations between first differenced log transformed firm specific variance and first differenced news counts. FV_t is the idiosyncratic variance for a given stock and $n_{i,t}$ are the individual time series of news counts. FV_t is estimated using intraday data. The correlations correspond to

$$Corr(\Delta ln(FV_t), n_{i,t+s}), \text{ where } n_{i,t+s} \equiv max(\Delta c_{i,t+s}, 0)$$

where i denotes a specific news source. $n_{i,t}$ gives positive changes in the number of news items. s denotes the time of the news indicator and corresponds to the value on the x-axis in the above charts. The sample includes 28 large US stocks. Sample length differs from stock to stock but generally spans from January 29, 2001 to July 31, 2009. We use the following Factiva source codes to identify the respective publications: J (The Wall Street Journal), WP (The Washington Post), USAT (USA Today), NYTF (The New York Times), DJ (Dow Jones Newswire), LBA (Reuters).

analysis of corporate information events. Similarly, Mitchell and Mulherin (1994) find that 44.3% of Dow Jones items are also reported in the Wall Street Journal, and 36.3% of items are only transmitted via the Dow Jones wire, suggesting that these news items are considered to contain less important information. The pattern we observe in Figure 2 and Figure 3 is consistent with the description of the news disclosure process by previous empirical research suggesting that newspapers generally operate with a delay with respect to newswires. In fact, panel (g) in Figure 2 provides direct evidence of this. Panel (g) in Figure 2 presents a large positive cross-correlation between aggregate newspapers at time t and aggregate newswires at time $t-1$.

From Figure 3 we also observe that newspaper based information indicators are less correlated with changes in realized variance than leading newswires. We argue that this result reflects the role of newswires and newspapers in the media industry. Newswires focus on providing continuous coverage of issues as they unfold while newspapers focus on providing in-depth analysis. The print publication format of newspapers may add further constraints on their coverage of corporate information events. Thompson et al. (1987) suggest that newspapers have materiality limits implying that some stories transmitted to newswires never make it to newspapers. Thompson et al. (1987) cite an example of defense contracts, where contracts below a certain dollar amount are not included in the newspaper edition of the Wall Street Journal. Constraints on news item distribution may therefore impose materiality limits on the news included in newspapers.

4. Evidence of Information Processing Components in Equity Volatility

This section summarizes our estimation results and contrasts these to previous research. First, we summarize general measures of fit for each estimated model. We proceed to show results on the relationship between volatility clustering and news. Next, we provide an overview of the type of economic information included in the estimated time series regressions across stocks. We conclude by reviewing the role of media attention indicators.

Table 5 presents the incremental R^2 from time series regressions where measures of information are included. Starting from the left and moving to the right, each column shows how the R^2 measure of fit changes when different subsets of news indicators are allowed to enter the model. R^2_{Total} denotes a model where all three components are included in the form of an autoregressive part, contemporaneous news and lagged news. The next column, R^2_{Max}, denotes a model with all three components, where no model reduction procedure has been performed. The last three columns present Wald statistics for restrictions for groups of news indicators.

Table 5: The additional R^2 of firm specific indicators of news arrival

This table illustrates the incremental R^2 from adding contemporaneous and lagged news indicators. The model estimated is:

$$\Delta ln(FV_t) = \omega + \underbrace{\sum_{k=1}^{K} \rho_k ln(FV_{t-k})}_{\text{A}} + \underbrace{\sum_{i=1}^{M} \gamma_i n_{i,t}}_{\text{N}} + \underbrace{\sum_{i=1}^{M} \sum_{j=1}^{J} \tilde{\gamma}_{i,j} n_{i,t-j}}_{\text{L}} + \epsilon_t,$$

where FV_t is the firm specific realized variance and $n_{i,t}$ are indicators of new information arrival. A, N, and L denote what groups of variables are included in a model. A is the autoregressive part, N is contemporaneous news, and L is lagged news. Wald statistics (W_{+N},W_{+L},W_{+NL}) correspond to tests of linear restrictions on a subset of parameters, significance is marked by * based on p-values below levels of 0.01. Sample length generally spans from 29 Jan 2001 to 31 Jul 2009. CVX, HPQ, INTC, T, MSFT, start at the following dates: Oct 10, 2001; Apr 9, 2011; May 6, 2002; Dec 1, 2005; Apr 9, 2001. R^2_{Max} denotes a model where no reduction was performed, and all contending variables are included.

Company	R^2_A	Δ_{+N}	R^2_{AN}	Δ_{+L}	R^2_{AL}	Δ_{+NL}	R^2_{Total}	R^2_{Max}	W_{+N}	W_{+L}	W_{+NL}
AA	31	5*	36	5*	36	10*	40	49	191	177	344
AXP	27	5*	32	3*	29	7*	34	43	220	81	237
BA	29	8*	37	3*	32	10*	39	49	354	109	416
BAC	23	5*	28	4*	27	8*	31	43	222	132	328
C	22	6*	28	6*	28	11*	33	44	238	230	401
CAT	30	7*	37	3*	33	9*	39	47	263	102	330
CVX	31	2*	33	4*	35	5*	36	45	83	128	169
DD	29	5*	34	3*	32	8*	38	46	168	115	299
DIS	26	3*	29	3*	29	7*	33	44	119	132	260
GE	24	6*	30	7*	31	11*	35	45	188	249	385
HD	27	3*	30	3*	30	7*	34	42	134	154	254
HPQ	27	6*	32	5*	32	9*	36	48	198	154	280
IBM	28	4*	32	4*	32	7*	35	45	132	161	243
INTC	25	3*	29	3*	29	6*	32	42	127	111	201
JNJ	27	7*	34	4*	31	10*	37	47	227	170	369
JPM	23	5*	28	6*	29	10*	33	44	220	216	344
KO	27	5*	32	4*	31	7*	35	45	173	113	266
MCD	29	11*	40	5*	34	13*	42	50	459	192	501
MMM	27	10*	36	3*	30	12*	39	45	351	111	428
MRK	29	3*	31	3*	32	6*	34	43	93	106	186
MSFT	29	3*	32	3*	32	5*	35	45	103	116	204
PFE	28	8*	36	3*	31	9*	37	47	297	88	306
PG	29	6*	34	2*	31	8*	36	46	198	73	273
T	27	8*	35	12*	38	21*	47	59	141	203	368
UTX	27	5*	32	3*	30	7*	34	42	195	108	234
VZ	27	5*	32	3*	30	8*	35	44	208	156	267
WMT	29	5*	35	3*	32	7*	37	46	209	155	290
XOM	32	1*	33	3*	35	6*	38	46	62	146	230

The Wald statistics in Table 5 provide evidence that, for all 28 stocks, both the public and private information processing components are statistically significant. This result suggests that the arrival of unexpected public information triggers information processing activities leading to private information being incorporated sequentially. A close examination of Table 5 reveals that both the contemporaneous news component and the lagged news component are significant by themselves as well as together. Suggesting that public and private information effects from new economic information arrival appear to be equally important in accounting for changes in firm specific realized variance.

Our finding, that news is related to increases in volatility, is of particular interest since prior research using measures of news arrival found small or insignificant relationships with return volatility. Roll (1984) examined the variability in orange juice futures prices and found a substantial amount of unexplained volatility. While he found higher volatility on days with Wall Street Journal articles covering *oranges* than on days without, he argued that news was of little importance given the low frequency of such information events. Both Mitchell and Mulherin (1994) and Berry and Howe (1994) examine the relationship between public information arrival and aggregate stock market volatility, and conclude that the relationship is small and insignificant. On a daily basis, Mitchell and Mulherin (1994) examine the relationship between all the news items published by Dow Jones[12] and absolute returns on value-weighted indeces across the NYSE, AMEX and the over-the-counter markets. They find that news is positively and significantly related to the absolute value of market returns, although the size of the relationship is small. On an intraday basis, Berry and Howe (1994) examine the relationship between the number of news items released by the Reuters's *North American Securities News* wire service, and return volatility of the S&P500 index. The relationship is insignificant for all 13 one-half-hour periods considered.

The idea that public information arrival may generate private information through private processing of new public information is probably related to Roll (1988)'s finding that the probability of information arrival is higher on days without public news arrival. Roll (1988) describes the prevailing paradigm about changes in asset prices and suggests that through observation and measurement of unanticipated economic information a large proportion of changes in stock prices should be explainable. Using market model regressions on days with and without presence of firm specific news in the Wall Street Journal and the Dow Jones News Service, he finds slight evidence that public news reduces the explanatory power of systematic factors. He argues that the average probability of news is higher in the censored samples than in full samples and suggests this result is due to the financial press missing a great deal of relevant information generated privately, but that the volatility from private information generally is lower than that related to big newsworthy events.

In this paper, by exploiting the full cross-section of news sources, we have sought a measure of firm specific news that is as collectively exhaustive as possible. As presented in the seventh column (Δ_{+NL}) of Table 5, we find that 5 to 20% of changes in firm specific realized variance are related to information arrival. Considering that 58 to 77% of realized variance is firm specific[13] and the high persistence[14] in firm specific realized variance, we argue that our ability to, ex-post, relate 5 to 20% of changes in variance to crude measures of information arrival is of economic importance. Even so, our conclusion is similar to Roll (1988) since we find that public news cannot account for all variation in returns. Interestingly, our study shows that lagged news effects, which were not accounted for in Roll (1988), are significant alongside contemporaneous effects. To summarize, our results favor the paradigm that changes in volatility can be related to the arrival of unanticipated economic information.

Across stocks, we find that 5 to 20% of changes in firm specific variance are related to the arrival of unexpected firm specific information. This result contributes to the discussion

[12] Five news wire services, one newspaper (The Wall Street Journal) and one magazine (Barron's)

[13] See Table 3

[14] See Table 7

raised by French and Roll (1986). French and Roll (1986) use the distinction between trading and non-trading periods as a proxy for the rate of information arrival in order to study information processing in financial markets. They suggest three hypotheses for patterns in stock return variances: Public information is more likely to arrive during trading hours, private information is incorporated into prices only through trading by informed investors during trading hours, and pricing errors occur during trading hours. They conclude that variation in the information flow is the most likely determinant of volatility, and that private information is likely to be the largest component. We measure the rate of information arrival directly, our results suggest that the relationship between public and private information is more complex and, based on the relative size of the incremental R^2 in column seven (Δ_{+NL}) of Table 5 for regressions including indicators of information arrival, that public information and related processing of public information is of economic importance for stock return variances.

This view is in stark contrast to the results of Cutler et al. (1989). They estimate the fraction of aggregate stock returns that can be attributed to various types of economic news. They find that the arrival of new information about the performance of the economy can explain only 20 to 33% of the variation in stock returns and that a review of headlines from the New York Times was unable to account for large returns. Similarly, Shiller (1981) and Grossman and Shiller (1981) argue that stock prices vary more than can be expected given the relatively lower variance of corporate dividends. Similarly, Schwert (1989) investigates the relationship between stock market volatility and measures of economic volatility and is unable to find a strong relationship. We argue that the Wald statistics in the last three columns of Table 5 are evidence for the view that information arrival is related to changes in asset prices. In addition, since 5 to 20% of changes in firm specific return volatility can be explained by such a crude measure of information arrival it is likely that news is an important missing piece in Shiller (1981)'s volatility puzzle, which states that stock price volatility is too high to be accounted for by new information about the economic performance of the firm.

If the autoregressive properties of firm specific return volatility are due to either public news clustering or private information arrival, then Table 6 supports the view that most return volatility is due to information arrival. Table 6 presents R^2's for the explained variation in Realized Variance, RV_t. Table 6 considers how much of the entire RV_t is explained, therefore this table is different from Table 5 which considers how much of changes in log-transformed firm specific variance can be explained, $\Delta ln(FV_t)$. From Table 6 we see that across stocks the systematic component explain between 35.6 to 84.0 percent of the variation in Realized Variance while models with an autoregressive component and news indicators explains between 74.6 and 94.9 percent.

Table 6: Comparison of R^2 for Realized Volatility for different models of with and without news arrival indicators

Company	S&P500	News	Lags	News & Lags	AR	AR & News	AR & Lags	AR, News & Lags
AA	61.1	81.1	85.2	85.6	88.0	88.3	89.0	**89.3**
AXP	59.4	69.5	66.0	71.6	84.9	85.0	**85.2**	85.0
BA	51.2	72.2	82.4	69.7	88.2	84.0	**88.7**	83.1
BAC	72.4	85.7	87.1	86.0	87.8	90.9	88.4	**91.2**
C	57.0	68.4	75.3	80.7	77.6	82.9	85.2	**89.4**
CAT	68.9	89.4	88.6	89.6	92.4	93.8	92.6	**94.2**
CVX	82.6	90.8	90.8	91.2	94.4	94.5	**94.9**	94.7
DD	71.6	89.2	90.0	90.0	93.1	93.9	93.5	**94.3**
DIS	45.5	62.4	63.1	61.5	72.1	72.8	73.1	**76.8**
GE	66.3	82.9	85.1	87.7	86.6	89.7	88.8	**90.7**
HD	62.2	85.8	85.0	85.4	91.2	91.3	91.3	**91.4**
HPQ	41.0	73.9	81.6	77.0	87.9	88.1	88.4	**88.7**
IBM	69.5	87.8	88.4	87.7	93.2	92.9	**93.6**	93.1
INTC	58.9	86.3	88.3	88.3	91.5	92.0	92.1	**92.5**
JNJ	42.0	64.1	63.7	70.6	76.6	**83.6**	76.6	81.0
JPM	66.2	84.4	84.7	84.7	86.8	87.4	**89.5**	88.7
KO	53.2	82.4	85.1	84.3	90.1	89.6	**90.5**	90.2
MCD	35.6	58.3	56.5	61.7	71.6	75.7	72.8	**76.7**
MMM	64.1	79.9	79.8	81.0	88.3	89.3	88.4	**89.6**
MRK	38.2	43.4	50.4	50.4	70.2	73.2	71.8	**74.6**
MSFT	68.4	88.2	88.4	88.4	92.0	92.3	92.4	**92.5**
PFE	45.0	70.6	70.6	70.4	82.0	84.9	84.0	**86.6**
PG	65.3	81.5	80.6	81.8	88.9	89.3	88.9	**89.4**
T	69.7	81.1	88.1	90.2	89.1	89.3	93.4	**94.6**
UTX	60.8	74.7	76.5	74.3	85.2	86.3	85.5	**86.5**
VZ	57.7	79.2	77.2	78.7	87.2	87.9	87.3	**88.1**
WMT	51.5	72.5	69.5	72.3	82.8	83.3	82.9	**83.6**
XOM	84.0	82.9	83.6	83.3	91.4	91.3	**91.5**	91.4

Notes: Each column represents the respective R^2 for each model. For all columns the R^2 corresponds to:

$$R^2 = 1 - \frac{\sum_{t=1}^{T}(RV_t - \widehat{RV_t})^2}{\sum_{t=1}^{T}(RV_t - \overline{RV_t})^2}$$

although $\widehat{RV_t}$ differs according to the specification for firm specific variance, $E[FV_t]$, that is estimated from (8) in such a fashion that $E[FV_t] = \exp(E[\Delta ln(FV_t)] + \frac{1}{2}\sigma_{\epsilon t} + ln(FV_{t-1}))$. For the second column, **S&P500**, $E[FV_t] = 0$. For the rest, $E[FV_t]$ differs depending on which components are included (**AR**, **News**, **Lags**).

4.1. Volatility Clustering

This investigation allows for a direct test of the mixture of distributions hypothesis. As mentioned in Section 2, the mixture of distributions hypothesis captures the prevailing theories and hypotheses about the sources of volatility and volatility clustering. Table 7 answers the question of how the persistence in volatility is affected by adding measures of information arrival. It compares non-nested models with one or more of the following parts: An autoregressive component (A), contemporaneous news (N), and lagged news (L). Table 7 allows for a comparison of the persistence in volatility in models with and without indicators of information arrival. The first row for each model illustrates the respective R^2. The second row corresponds to the estimated half-life computed from the impulse response function of the model. The third row presents the percentage change in half-life when compared to the base model, a model only containing the autoregressive component, $\mathcal{M}_U^5(A)$. From the third row we can see that the relative percentage change in half-life is generally negative as we move from models without news to models with news. Looking across the 28 stocks, a 95% confidence interval for the percentage change in half-life ranges from 0.4 to 11.1 for models with contemporaneous news, 1.6 to 15.5 for models with lagged news effects, and 3.3 to 18.3 for models with both contemporaneous and lagged news.

Table 7: Comparison of persistence as measured by the half-life of shocks in the impulse response function accross different models of firm specific volatility with and without news arrival indicators

	Model										
	$\mathcal{M}_U^5(A)$		$\mathcal{M}_U^6(AN)$			$\mathcal{M}_U^7(AL)$			$\mathcal{M}_U^8(ANL)$		
Ticker	R^2	HF	R^2	HF	%ΔHF	R^2	HF	%ΔHF	R^2	HF	%ΔHF
AA	31	17	36	15	-12	36	12	-29	40	12	-29
AXP	27	39	32	33	-15	29	51	31	34	37	-5
BA	29	14	37	12	-14	32	13	-7	39	13	-7
BAC	23	47	28	30	-36	27	44	-6	31	30	-36
C	23	51	28	39	-24	28	25	-51	33	23	-55
CAT	30	14	37	16	14	33	15	7	39	19	36
CVX	31	11	33	9	-18	35	10	-9	36	8	-27
DD	29	15	34	13	-13	32	15	0	38	14	-7
DIS	26	16	29	15	-6	29	13	-19	33	12	-25
GE	24	27	30	19	-30	31	18	-33	35	16	-41
HD	27	16	30	18	13	30	14	-13	34	15	-6
HPQ	27	16	32	14	-13	32	14	-13	36	16	0
IBM	29	19	32	18	-5	32	16	-16	35	16	-16
INTC	25	17	29	17	0	29	14	-18	32	17	0
JNJ	27	14	34	18	29	32	14	0	37	14	0
JPM	23	35	28	25	-29	29	19	-46	33	18	-49
KO	27	18	32	18	0	31	17	-6	35	17	-6
MCD	29	7	40	7	0	34	8	14	42	8	14
MMM	27	11	36	10	-9	30	12	9	39	11	0
MRK	29	7	31	7	0	32	7	0	34	7	0
MSFT	29	19	32	17	-11	32	17	-11	35	14	-26
PFE	28	11	36	11	0	31	8	-27	37	8	-27
PG	29	13	34	13	0	31	16	23	36	15	15
T	27	5	35	5	0	38	5	0	47	5	0
UTX	27	12	32	12	0	30	14	17	34	13	8
VZ	27	15	32	14	-7	30	13	-13	35	14	-7
WMT	29	12	35	13	8	32	11	-8	37	12	0
XOM	32	13	34	15	-15	35	11	-15	38	12	-8

Notes: 1st column for each model illustrates the respective R^2. Variable subsets are denoted by a letter in the parenthesis of the model notation. For example, notation $\mathcal{M}_R^8(ANL)$ refers to the unrestricted version of a model with autoregressive components (A), contemporaneous news (N), and lagged news (L) indicators.
The 2nd column, HF, is the half-life computed as the time it takes for half of the total accumulated impulse response to volatility to pass. So the half-life measures the time until half of the long run effect of a shock to volatility has taken place. A 3rd column exists for all but the first model and represents the percentage change in half-life from $\mathcal{M}_U^5(A)$ to the respective models,$\mathcal{M}_U^6(AN)$, $\mathcal{M}_U^7(AL)$ and $\mathcal{M}_U^8(ANL)$. The 3rd row therefore represents the relative percentage change in half-life relative to the half-life in $\mathcal{M}_U^5(A)$. For example, -12 for $\mathcal{M}_U^6(AN)$ of Alcoa, implies that the half-life has decreased by -12 % with respect to $\mathcal{M}_U^5(A)$.

The message is clear, including measures of information arrival reduces the amount of volatility clustering explained by the autoregressive component. While the size of this effect may appear small, its implications are of economic importance. Consider the case of Alcoa (NYSE:AA). AA has a half-life[15] of shocks (an increase in volatility with respect to the previous day) to the volatility process of approximately 17 days. In models with both contemporaneous and lagged news, the half-life corresponds to approximately 12 days. Across the stocks, this effect is most substantial for Citigroup (NYSE:C), for which

[15]The time required for volatility to fall to half its value.

the half-life of shocks to the volatility process declines from 51 to 23 days when news is included. Generally, across stocks including contemporaneous and lagged news leads to a decrease in the half-life of volatility shocks of between 1 to 6 days, as measured by a 95% confidence interval across stocks.

The results in Table 7 suggest that information arrival is related to volatility clustering and that the persistence in the volatility process in standard volatility models may overestimate persistence in the underlying process when news effects are not included. These findings are consistent with the results of Goodhart, Hall, Henry, and Pesaran (1993) and Andersen and Bollerslev (1998). Both studies examine exchange rates, respectively the Sterling-Dollar and Deutsche Mark-Dollar rates, and model volatility while accounting for news effects. Goodhart et al. (1993) finds that the parameters of a GARCH(1,1) model for the variance decrease substantially, implying that including news effects allows for a more accurate estimate of the persistence in volatility dynamics. While news reduces the size of the parameters, Goodhart et al. (1993) finds that news effects cannot account for all serial correlation in squared returns. In the Deutsche Mark-Dollar setting Andersen and Bollerslev (1998) estimate three different components of the volatility and find that the interdaily variance component is by far the component of largest economic significance, since intradaily calendar effects disappear when aggregated to daily frequency. Similarly, intraday announcement effects, although significant, account for a small proportion of explanatory power. Andersen and Bollerslev (1998) conclude that intraday calendar effects and intraday announcement effects cannot account for the interdaily autoregressive component. In comparison with our model, it should be noted that the model approach in Andersen and Bollerslev (1998) does not allow for interdaily effects of announcements, such as the lagged news effects in our model.

The results in Table 5 and Table 7 also have implications for a large body of literature exploring different methods to estimate the underlying information flow of financial assets by using observed returns and trading volume. Past studies have proposed models that explain characteristics of the return generating process best described as autoregressive properties in the variance of returns. Mandelbrot (1963) and Fama (1965) were the first to note that large returns tend to follow large returns and small returns tend to follow small returns. The mixture of distributions hypothesis is one theoretical explanation for this phenomena. The hypothesis suggests that a serially correlated mixing variable, measured as the rate of information arrival, causes the autoregressive properties of the variance of returns of financial assets. This theory has been proposed in several variations (Press 1967, Clark (1973), Epps and Epps (1976), Tauchen and Pitts (1983), Ross (1989) and Andersen (1996)). A recent study by Maheu and McCurdy (2004) puts forth a version of the mixture of distributions hypothesis where the return generating process is assumed to be directed by a latent news process. The conditional variance of returns is specified to have a smoothly evolving component, related to the diffusion of past news arrival, and a component related to the contemporaneous information arrival process that generates jumps. Our empirical approach can be considered a direct test of the mixture of distributions hypothesis, where measures of new economic information are related to changes in the rate of price informativeness. Again, our results support mixture model specifications of the return generating process and in particular mixture models with lagged effects of information arrival such as the diffusion of past information arrival in Maheu and McCurdy (2004).

4.2. **Economic Information**

We now turn to characterizing the firm specific news that we find is related to changes in stock return volatility. Table 8 shows that news items categorized as covering *Acquisitions/Mergers/Takeovers* are most commonly found in regressions explaining changes in firm specific realized variance. In time series regressions that only include contemporaneous news series, our model selection procedure includes it 8 times, while in regressions where we also consider lagged information indicators, it is included 10 times across our sample of 28 stocks. Alcoa (NYSE:AA) is one of the stocks for which this is the case. Alcoa's newsflow for this subject category includes 6,564 news items. 1,433 mention Alcan, including Alcoa's attempted acquisition efforts initiated on May 7, 2007. 727 items refer to Rio Tinto, one of Alcoa's competitors. 326 are related to Alcoa's efforts to acquire the Norwegian aluminium producer Elkem. 278 discuss BHP Billiton, another competitor. 107 mention the Chinese aluminium corporation, Chalco. In other words, this subject category is dominated by industry related news items mentioning other industry participants and their acquisition, merger and takeover activities.

Table 8 also shows that *Earnings Surprise* is the second most common information arrival indicator across stocks. This series includes earnings announcements deemed by news agencies and the Dow Jones Intelligent Indexing system to convey material change in performance. Newsflow in this category often contains headlines mentioning earnings or profitability measures and such words as *decline*, *disappoint*, *drop*, *loss*, *plunge*, *jump*, *strong*, *beat*, and *boost*. American Express (NYSE:AXP) is one of the companies where *Earnings Surprise* newsflow is included. The news items often refer specifically to changes in the performance of individual business units and end-markets. For example, trends in card issuance and delinquencies associated with American Express's credit card business. Moreover, it is not uncommon to find news items discussing earnings surprises for competitors and companies in the same lines of business.

Earlier work by Patell and Wolfson (1984), explored the intraday price adjustment speed after earnings and dividend announcements. Using intraday data from 1976 to 1977 for 96 firms and news from the Dow Jones News Service, they found that earnings and dividend announcements are related to disturbances and persistence in price change variance for several hours after the announcement, extending into the next trading day.

Table 8: Parameter Overview - Corporate and Industrial Indicators

This table summarizes what indicators of economic information arrival generally are included in the time series regressions in the cross section of our sample of 28 large US stocks. The *Number* column presents the number of stocks for which the respective time series of a given information arrival indicator was included in a model describing the changes in firm specific realized variance of a given stock. The model estimated is:

$$\Delta ln(FV_t) = \omega + \sum_{k=1}^{K} \rho_k ln(FV_{t-k}) + \sum_{i=1}^{M} \gamma_i n_{i,t} + \sum_{i=1}^{M} \sum_{j=1}^{J} \tilde{\gamma}_{i,j} n_{i,t-j} + \epsilon_t,$$

where FV_t is the firm specific realized variance for a given stock and $n_{i,t}$ are the individual time series indicating the arrival of new information. K is 5, corresponding to five lags of the log-level realized firm specific variance, while M is 99 corresponding to 83 economic information indicators and 16 media attention indicators. J is 3 and corresponds to inclusion of up to 3 lags of the corresponding information arrival indicator. FV_t is estimated using intraday data.

	Number				
	News	News & Lagged News			
Indicator	γ	γ	$\tilde{\gamma}_{t-1}$	$\tilde{\gamma}_{t-2}$	$\tilde{\gamma}_{t-3}$
Acquisitions/Mergers/Takeovers	8	10	2	4	7
Earnings Surprises	8	8	5	10	8
Divestitures/Asset Sales	7	3	3	1	1
Sales Figures	6	6	4	6	6
Corporate Crime/Legal/Judicial	5	10	1	4	4
Management Issues	5	7	5	3	6
Natural Reserves/Resources Discovery	5	4	2	5	2
Privatizations/Nationalization	5	3	4	3	4
Analyst Comment/Recommendation	4	7	2	3	3
Dividends	4	6	6	4	4
Franchises	4	5	4	3	4
Research/Development	4	5	3	5	2
Defense Contracts	4	5	2	2	5
Share Capital	4	4	5	3	5
Lay-offs/Redundancies	4	4	3	5	6
Financing Agreements	4	4	3	1	5
Corporate Credit Ratings	3	6	3	4	3
Joint Ventures	3	5	5	6	8
Regulation/Government Policy	3	4	6	3	1
Product Safety	3	4	4	4	1
Information Technology	3	4	4	3	4
Licensing Agreements	3	2	3	2	6
Bankruptcy	3	0	3	1	2
Internal Audit	2	4	10	5	6
Non-governmental Contracts	2	4	4	5	4
Government Contracts	2	4	4	2	0
Corporate Debt Instruments	2	3	3	2	3
Competition Issues	2	3	3	2	1
Capacity/Facilities	2	1	2	5	6
Outsourcing	2	0	4	5	3
Annual Meetings	1	3	3	5	2
Labor Disputes	1	2	6	2	7
Output/Production	1	2	2	6	0
Marketing/Market Research	1	1	5	4	1
Plans/Strategy	1	1	4	2	4
New Products/Services	0	4	3	4	5

The results found in Table 8 are in line with Patell and Wolfson (1984) since we find that *Earnings Surprise* is the information arrival indicator most likely to have a lagged relationship with firm specific realized variance. This can be observed by examining the number of lagged indicators for the *Earnings Surprise* category that are used in stock specific regressions. We see that the contemporaneous *Earnings Surprise* indicator is included for 8 stocks, while between 5 and 10 of the stock specific regressions include lagged effects of this indicator. This pattern suggests that the release of *Earnings Surprise* information is more likely to provide information-based investors with an advantage from information processing activities, since there are information-related effects up to three days after a shock to the *Earnings Surprise* newsflow.

An interesting result in Table 8 is that no category is included across more than one third of our 28 stocks. An examination of the underlying newsflow significant for each stock also reveals material differences in the type of issues surfacing within any single newsflow category. For example, we examine the *Corporate Crime/Legal/Judicial* category which is significant for both IBM (NYSE:IBM) and Citigroup (NYSE:C). For IBM, an information technology products and services provider, themes in the newsflow include a series of lawsuits including a lawsuit by Compuware for misappropriation of source code used in mainframe testing, a civil lawsuit by the SCO Group related to licensing rights to the UNIX computer operating system, as well as investigations into IBM's accounting practices by the U.S. SEC. For Citigroup, a financial services provider, the same newsflow category covers an investigation by the German financial services regulator into Citigroup's alleged manipulation of European bond markets as well as litigation with respect to underwriting activities for WorldCom, Enron and Parmalat. These examples illustrate the diversity of the events covered in the news for different companies.

Many of the corporate information events covered by our indicators have previously been studied. In fact, the idea that new economic information drives changes in asset prices is supported by a large body of literature using event study methodologies to explore how different corporate events are related to changes in stock prices. McWilliams and Siegel (1997) examine the use of event studies in management research. They find the event study methodology has been used to analyze the effects of events endogenous and exogenous to the firm. Endogenous corporate events studied include changes in corporate control, corporate refocusing, CEO turnover, use of affirmative action programs, layoffs, plant closures, corporate illegalities, product recalls, customer service changes, diversification programs, strategic investment decisions, and the formation of joint ventures. Exogenous events studied include enactment of major legislation, the appointment of top executives to cabinet positions and the deaths of CEOs. Although the event study literature is extensive and far reaching in scope, McWilliams and Siegel (1997) highlight the disparity in the empirical implementation of the studies and shed light on the importance of controlling for confounding events when using the methodology. A common concern in event studies is the ability to control for confounding events. As Table 3 illustrates, on average the stocks in our sample have between 7 and 102 news items each per trading day. In fact, it is seldom that these stocks a have a day where they are not mentioned in the media. Despite the presence of multiple news items, our model selection procedure still identifies certain news subjects as significant. Results in Table 8 confirm the importance of several corporate events studied in prior research while controlling for the vast number of simultaneously occurring news items available for a given company.

4.3. **Media Attention**

Similar to Table 8, Table 9 presents an overview of the media attention indicators found significant across our sample of stocks. The indicators in this table have all been included in order to simultaneously control for the media attention hypothesis put forth by studies such as Huberman and Regev (2001), Barber and Odean (2008) and Engelberg and Parsons (2011). Engelberg and Parsons (2011) define the media attention hypothesis as follows:

> ...the effects we identify at the local level should apply generally, that is, to national media outlets with audiences large enough to meaningfully impact capital allocation. [p. 96]
>
> We find that extreme earnings surprises are related to the volume of retail trade. However, the media effect we identify is several times larger than this information effect no matter how we define our earnings surprise. Simply put, in our setting, the media is at least, and sometimes more, likely to drive trade than information. If these generalize (even partially) to the aggregate level, they easily are capable of influencing prices and allocations. [p. 96]

Table 9 seems to say that newswire media attention is important when explaining changes in firm specific realized variance. This result appears to be at odds with the results of Engelberg and Parsons (2011). Engelberg and Parsons (2011) compare the behavior of investors with access to different media coverage of the same information event. They find that local media coverage predicts local trading while controlling for the characteristics of earnings announcements, investors and individual newspapers. They suggest that their results would also apply to large media outlets, such as newspapers with substantial circulation, and that the media attention relationship is more likely to drive trade than unanticipated economic information.

Our investigation has tested the information processing hypothesis while controlling for the media effect. From Table 8 and Table 9 we conclude that, while several economic information categories are relevant for explaining changes in firm specific variance, the media attention hypothesis cannot be ruled out. However, the main channel for the media attention effect in our study appears to be the aggregate attention from newswires as opposed to leading newspapers with large circulation, such as *USA Today*. We argue that the newspaper media attention effect is more plausible when thought of as a channel for the resolution of asymmetric information between informed and non-informed traders as modeled in Foster and Viswanathan (1990) and as examined theoretically and empirically in Tetlock (2010). On the other hand, the significance of the newswire media attention indicator may arise from the indicator's ability to proxy for the number of informed traders. For example, Holden and Subrahmanyam (1992) show that price change variance increases with the number of informed traders.

Table 9 confirms the evidence in Tetlock (2010) that the number of newswire items has predictive power for news day price changes. Tetlock (2010) uses 2.2 million articles from the Wall Street Journal and the Dow Jones News Service to examine if public news eliminates information asymmetry from two types of traders for a sample of 13,842 firms from 1979 to 2007. He suggests that public news levels the playing field for other investors by resolving asymmetric information. In addition, he finds that the number of newswire messages subsumes the predictive power of news day returns but finds a small positive relationship between contemporaneous public information and absolute returns.

Table 9: Parameter Overview - Media Attention Indicators

This table summarizes what indicators of economic information arrival generally are included in the time series regressions in the cross section of our sample of 28 large US stocks. The *Number* column presents the number of stocks for which the respective time series of a given information arrival indicator was included in a model describing the changes in firm specific realized variance of a given stock. The model estimated is:

$$\Delta ln(FV_t) = \omega + \sum_{k=1}^{K} \rho_k ln(FV_{t-k}) + \sum_{i=1}^{M} \gamma_i n_{i,t} + \sum_{i=1}^{M} \sum_{j=1}^{J} \tilde{\gamma}_{i,j} n_{i,t-j} + \epsilon_t,$$

where FV_t is the firm specific realized variance for a given stock and $n_{i,t}$ are the individual time series indicating the arrival of new information. K is 5, corresponding to five lags of the log-level realized firm specific variance, while M is 99 corresponding to 83 economic information indicators and 16 media attention indicators. J is 3 and corresponds to inclusion of up to 3 lags of the corresponding information arrival indicator. FV_t is estimated using intraday data. The sample includes 28 large US stocks. Sample length differs from stock to stock but generally spans from 29 January 2001 to 31 July 2009. CVX, HPQ, INTC, MSFT, T start at the following dates respectively: October 10, 2001; April 9, 2011; May 6, 2002; December 1, 2005; April 9, 2001.

	Number				
	News	News & Lagged News			
Indicator	γ	γ	$\tilde{\gamma}_{t-1}$	$\tilde{\gamma}_{t-2}$	$\tilde{\gamma}_{t-3}$
News wires	22	22	6	5	6
Dow Jones News Service	11	11	4	2	4
Reuters News	10	10	5	8	3
Press Release wires	5	6	3	3	2
Associated Press Newswires	5	6	3	2	4
Newsletter wires	4	6	7	4	3
News and Business Publications	4	6	3	2	3
The Washington Post	4	4	4	2	1
Other Publications	4	2	3	5	6
The Wall Street Journal	3	6	5	4	4
Industry Publications	3	3	4	5	2
The New York Times	3	3	3	3	4
USA Today	3	3	2	5	4
Newspaper Publications	2	0	4	2	4
Financial Times	1	3	3	3	2
PR Newswire (U.S.)	1	2	2	2	1
Government Politics Publications	0	3	5	5	0

The results in Table 8 and Table 9 differ in one key way from Tetlock (2010), we find a significant relationship between contemporaneous information arrival and changes in realized variance. We think that Tetlock (2010)'s finding that public information cannot account for news day returns may be due to the omission of sources other than the Dow Jones newswires and the Wall Street Journal, in order to measure content materiality better. For example, as presented in Table 9, measures of increases in aggregate newswire attention are included in the time series regressions of 22 out of 28 companies. Moreover, we find that the main Dow Jones and Reuters news wires are often included, 11 and 10 times respectively. In addition, the Wall Street Journal is included between 3 to 6 times across stocks. This may reflect the fact that it is focused on in depth analysis of corporate information events originally disclosed by Dow Jones's, the Wall Street Journal's corporate parent, family of newswires. In other words, we find support for the hypothesis of asymmetric information resolution (Foster and Viswanathan 1990; Tetlock 2010) due to the inclusion of several media attention indicators for leading newspapers, however we suggest that, by exploiting the cross-section of news sources, our empirical approach more accurately measures firm specific news arrival.

Our view, that lagged information effects are related to the processing of public information, is supported by a study by Engelberg, Reed, and Ringgenberg (2010). Engelberg et al. (2010) examine the trading pattern of short sellers when news from the Dow Jones newswires and the Wall Street Journal arrives. Using a sample covering 2005 to 2007 they find a significant increase in short selling after news events and that the most informative trades appear to be on days following news arrival. In addition, the most profitable trades are made on days where trades arrive later than those of other investors. They suggest that informed traders do not predict information arrival but rather gain their information advantage from processing publicly available information. They conclude that public news arrival creates trading opportunities for skilled information processors.

5. **Robustness: Switch, Mix and Split**

To investigate the robustness of our results we employ a three legged approach consisting of: switching the dependent variable with alternate measures, mixing the firm specific newsflow across the 28 stocks, and splitting the sample into non-overlapping periods. The results of our robustness checks provide suggestions for further research.

First, we use dependent variables that differ from realized firm specific variance. We employ trading volume and squared returns from CRSP as well as implied volatility from 1 month at-the-money call options. Our empirical approach accommodates the switch. When using trading volume, squared close-to-close returns, and implied volatility, we aim to check the relationship of news with other measures, linked to the firm's common stock, that we expect are related to the arrival of new information. Several empirical studies have shown that trading volume, squared returns and implied volatility are related to information arrival. Berry and Howe (1994) and Mitchell and Mulherin (1994) use aggregated news measures and document a positive relationship between broad market activity and the number of news items from the Reuters and Dow Jones news wires, respectively. Tetlock (2010) finds that the contemporaneous cross-correlation between absolute returns and abnormal trading volume is temporarily higher by 3.5% on days with Dow Jones or Wall Street Journal news items. At high-frequency, Groß-Klußmann and Hautsch (2011) find that volatility and trading volume are most sensitive to news arrival. With the media attention hypothesis in mind, Engelberg and Parsons (2011) are able to explain roughly 30% of the variation in log-transformed trading volume by using proxies for local media attention and indicators of earnings announcements. These studies suggest that reasonable proxies for information arrival and media attention should also be related to excess trading volume and excess volatility in close-to-close returns. Table 10 provides evidence that our indicators of information arrival and changes in media attention are able to explain significant amounts of changes in these measures.

Trading volume and squared returns are directly related to realized variance through the trading process, this is not the case for implied volatility from options on a given common stock. To add further flavor to our robustness check, we examine the relationship between information arrival and changes in implied volatility. Beyond the obvious connection, previous research suggests that the options market reflects the characteristics of the realized volatility process. Investigations by Patell and Wolfson (1979; 1981) examined the behavior of implied volatility around anticipated information releases such as earnings announcements. They find that option prices reflect investor's anticipation of earnings announcements and that implied volatility decreases upon announcement. Similarly, Donders and Vorst (1996) examine the impact of scheduled firm specific news events on implied volatility. Donders and Vorst (1996) find that implied volatility increases during the pre-event period but then drops following the announcement. Using non-scheduled events, Levy and Yoder (1993) investigate the behavior of implied volatility around M&A announcements. Around such events, implied volatility of targets increase up to three days prior to announcement. In a different setting, Ederington and Lee (1996) study the impact of macroeconomic information releases, such as the employment report and the producer price index, on T-bond, Euro dollar and Dollar/Deutsche Mark rates. They find that implied volatility increases following unscheduled announcements and decreases following scheduled announcements. More recently, Fornari and Mele (2001) investigate the impact of scheduled and non-scheduled news, from the Italian financial newspaper *Il Sole 24 ore*, on the implied volatility of long term Italian bond rates. Beyond at-the-money options, they also consider the deepest in-the-money and out-of-the-money options and confirm the results of Ederington and Lee (1996). With a focus on volatility dynamics, Engle and Mustafa (1992) show that the strong persistence of conditional variance from daily returns of common stocks is also reflected in the prices of the respective options. In other words, prior empirical research suggests that implied volatility should be related to measures of firm specific news.

Figure 4 illustrates lead and lag correlations between our indicator of *Earnings Surprise* news arrival and each of the four measures considered. Figure 4 provides a first indication of whether our indicators of information arrival are related to each measure in a similar fashion. It can be seen that firm specific realized variance, squared close-to-close returns and trading volume react similarly to the arrival of new information about earnings surprises. Interestingly, implied volatility reacts differently with a negative contemporaneous correlation, confirming the results of Patell and Wolfson (1979; 1981) and Ederington and Lee (1996).

Table 10: Model results when using different dependent variables - Trading Volume, Squared Close-to-Close returns and Implied Volatility

This table shows R^2's from models using trading volume (TV) from CRSP, squared close to close returns (SQ) from CRSP, and implied volatility from 1 month at-the-money call options from Bloomberg (IV):

$$\Delta ln(TV_t) = \omega + \alpha \Delta ln(SQ_{SP500,t}) + \sum_{k=1}^{K} \rho_k ln(TV_{t-k}) + \sum_{i=1}^{M} \gamma_i n_{i,t} + \sum_{i=1}^{M} \sum_{j=1}^{J} \tilde{\gamma}_{i,j} n_{i,t-j} + \epsilon_t,$$

$$\Delta ln(SQ_t) = \omega + \alpha \Delta ln(SQ_{SP500,t}) + \sum_{k=1}^{K} \rho_k ln(SQ_{t-k}) + \sum_{i=1}^{M} \gamma_i n_{i,t} + \sum_{i=1}^{M} \sum_{j=1}^{J} \tilde{\gamma}_{i,j} n_{i,t-j} + \epsilon_t,$$

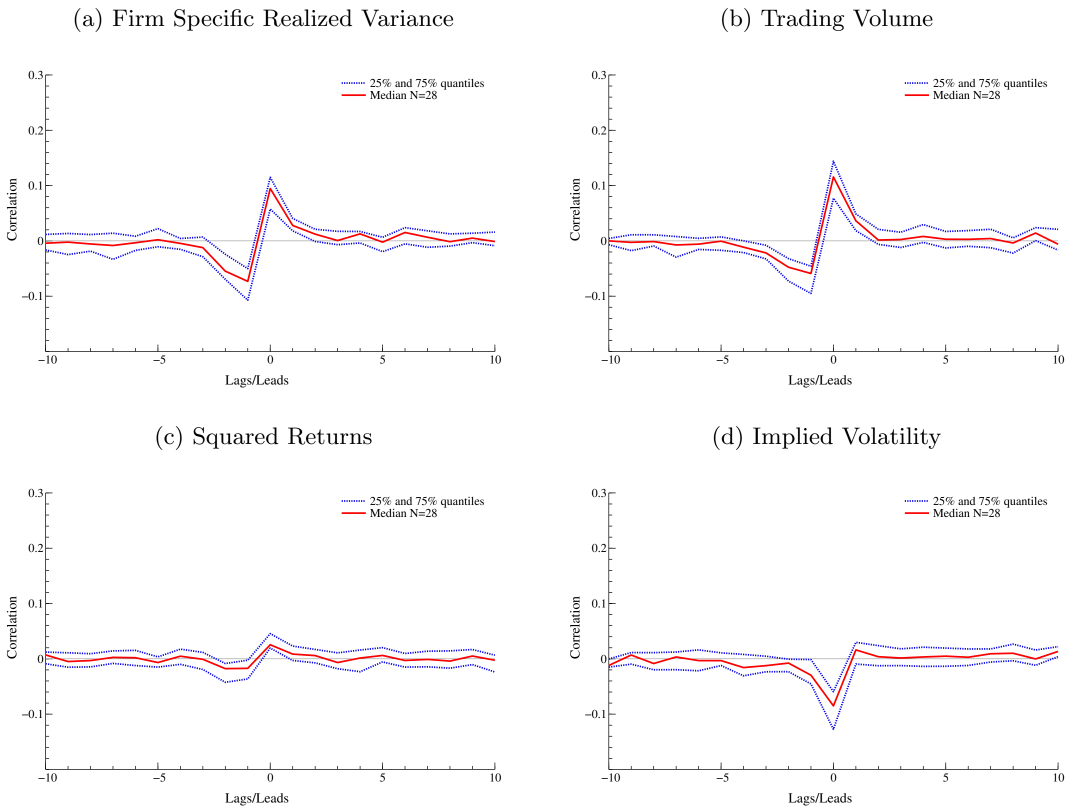
$$\Delta ln(IV_t) = \omega + \alpha \Delta ln(IV_{SP500,t}) + \sum_{k=1}^{K} \rho_k ln(IV_{t-k}) + \sum_{i=1}^{M} \gamma_i n_{i,t} + \sum_{i=1}^{M} \sum_{j=1}^{J} \tilde{\gamma}_{i,j} n_{i,t-j} + \epsilon_t,$$

Likelihood ratio test significance is marked by * based on p-values below levels of 0.01.ΔR^2_{+N} refers to the R-squared from models with an autoregressive component and contemporaneous news. ΔR^2_{+L} corresponds to models with all three components: An autoregressive part, as well as contemporaneous and lagged news.

	Trading Volume			Squared Returns			Implied Volatility		
Company	R^2	ΔR^2_{+N}	ΔR^2_{+L}	R^2	ΔR^2_{+N}	ΔR^2_{+L}	R^2	ΔR^2_{+N}	ΔR^2_{+L}
AA	28	15*	6*	49	4*	5*	14	4*	10*
AXP	27	9*	1*	52	2*	3*	23	2*	9*
BA	26	11*	3*	49	2*	3*	21	3*	7*
BAC	23	8*	8*	49	1*	7*	25	5*	9*
C	24	9*	4*	50	1*	3*	27	5*	5*
CAT	29	16*	7*	54	3*	5*	19	9*	7*
CVX	27	8*	6*	48	3*	6*	19	1*	4*
DD	32	7*	6*	52	3*	3*	25	3*	7*
DIS	25	11*	5*	46	3*	5*	21	3*	7*
GE	26	8*	8*	50	2*	5*	26	1*	1*
HD	23	14*	6*	46	4*	6*	25	5*	9*
HPQ	30	21*	3*	47	3*	3*	16	10*	7*
IBM	25	12*	5*	50	2*	4*	20	14*	8*
INTC	33	15*	4*	52	3*	5*	22	7*	7*
JNJ	28	12*	8*	48	2*	5*	18	5*	4*
JPM	27	9*	3*	51	2*	3*	28	3*	2**
KO	28	9*	6*	48	2*	6*	18	4*	6*
MCD	27	16*	4*	48	3*	4*	16	2*	5*
MMM	28	17*	5*	50	3*	4*	20	8*	4*
MRK	24	2*	6*	46	2*	6*	15	3*	6*
MSFT	27	12*	3*	51	2*	2*	21	4*	1*
PFE	25	18*	5*	48	3*	6*	19	3*	10*
PG	27	13*	2*	46	2*	4*	19	2*	7*
T	28	16*	21*	51	4*	14*	28	10*	14*
UTX	24	8*	4*	46	2*	4*	21	2*	8*
VZ	30	7*	8*	43	1*	5*	22	3*	2*
WMT	31	11*	3*	49	3*	1*	24	4*	4*
XOM	31	6*	4*	48	5*	4*	30	1*	6*

Figure 4 and the significant Wald statistics summarized in Table 10 confirm the evidence found in prior research suggesting that implied volatility is related to the arrival of firm specific information. We argue that this robustness check illustrates the strength of the evidence in Table 5 and supports our proposition that positive changes in newsflow are indicators of the arrival of relevant information.

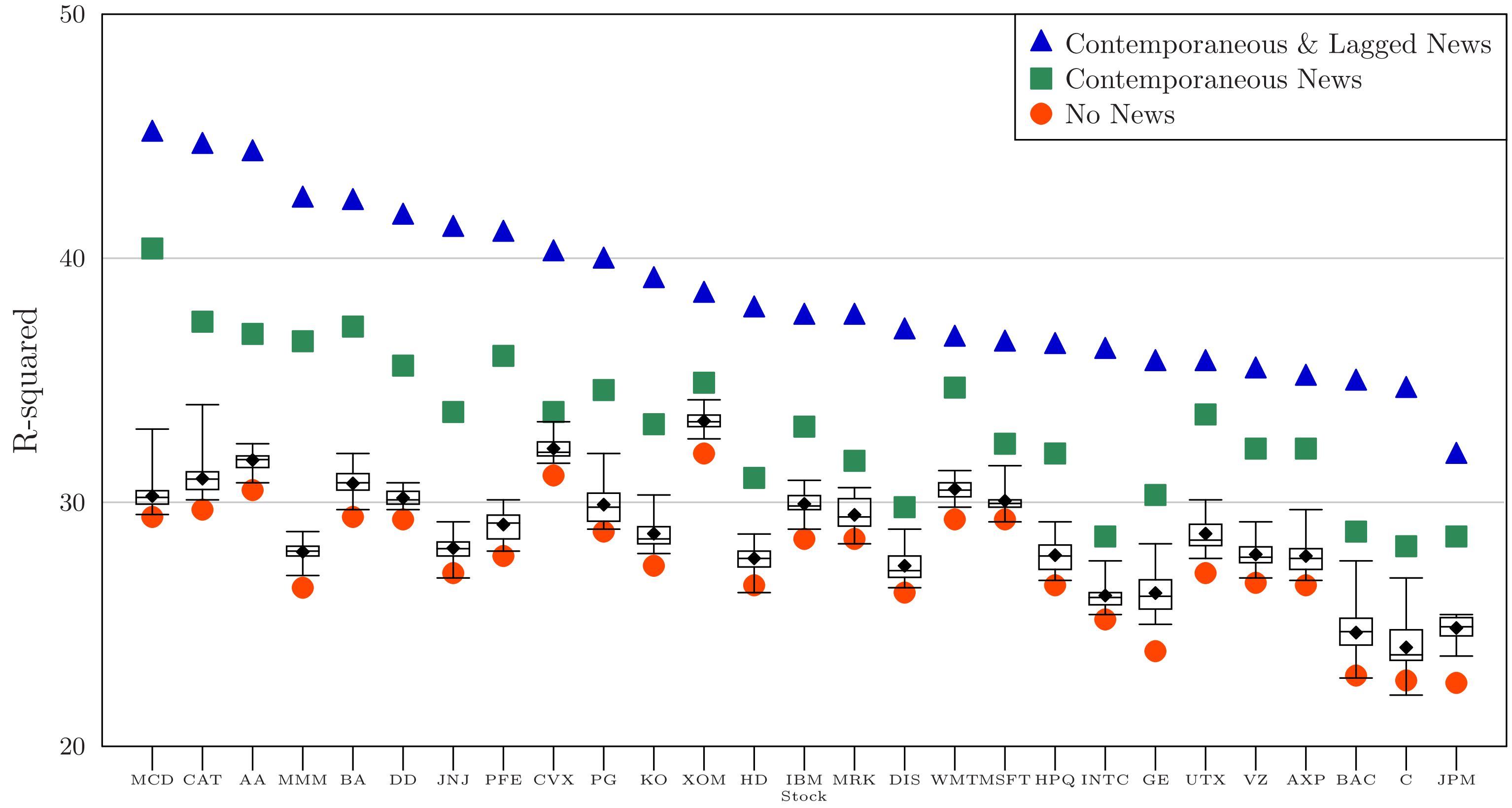
Figure 4: Cross-correlations between *Earnings Surprise* Indicators and Market Activity



Notes: This panel illustrates cross-correlations between first differenced log transformed firm specific variance and first differenced news counts. FV_t is the idiosyncratic variance for a given stock and $n_{i,t}$ are the individual time series of indicators of news arrival. TV_t is trading volume and SQ_t are squared close to close returns, both from CRSP. IV_t is implied volatility from 1-month at the money call options extracted from Bloomberg. The correlations correspond to Panel (a): $Corr(\Delta ln(FV_t), n_{i,t+s})$, (b): $Corr(\Delta ln(TV_t), n_{i,t+s})$, (c): $Corr(\Delta ln(SQ_t), n_{i,t+s})$, and d: $Corr(\Delta ln(IV_t), n_{i,t+s})$ where $n_{i,t} \equiv max(\Delta c_{i,t}, 0)$. So i denotes a specific categorization of news items, in this case based on a given subject. $n_{i,t}$ gives positive changes in the number of news items. s denotes the time of the news indicator and corresponds to the value on the x-axis in the above charts. The sample includes the 28 stocks. We use the following Factiva subject code: pC1514 - Earnings Surprise.

Second, we proceed in our robustness check efforts by using the original dependent variable, realized firm specific variance and mix the newsflow across our sample of stocks. In other words, we use firm specific newsflow for other companies than the company whose firm specific realized variance we are using as the dependent variable. Figure 5 shows that for all stocks in our sample, using the correctly matched newsflow always yields higher R^2s than using the incorrect newsflow. Generally, using the incorrectly matched realized variance and newsflow yields R^2 only slightly above a basic model without news indicators. These results suggest that our measures of information arrival and media attention are firm specific.

Figure 5: Robustness check: R^2 when mixing newsflow across stocks



Notes: Blue points indicate the R^2 from models where both contemporaneous and lagged news is included. Green points indicate R^2 from a model where a company's Firm Specific contemporaneous newsflow is used to explain changes in Firm Specific Realized Variance (FV). Red points denote a model for changes in FV using only past levels of the FV. The box plot for each stock ticker corresponds to the distribution of R^2 for models using the newsflow related to other companies. Each boxplot consists of 26 model estimations using newsflow for each of the other companies in the sample. AT&T (NYSE:T) is excluded due to its substantially smaller sample.

Table 11: Results from split sample estimation

This table shows R^2's from models using non-overlapping sub-samples. The model estimated is:

$$\Delta ln(FV_t) = \omega + \sum_{k=1}^{K} \rho_k ln(FV_{t-k}) + \sum_{i=1}^{M} \gamma_i n_{i,t} + \sum_{i=1}^{M} \sum_{j=1}^{J} \tilde{\gamma}_{i,j} n_{i,t-j} + \epsilon_t,$$

where FV_t is the firm specific realized variance for a given stock and $n_{i,t}$ are the individual time series of indicating the arrival of new information. K is 5, corresponding to five lags of the log-level realized firm specific variance, while M is 101 corresponding to 83 economic information indicators and 16 media attention indicators. J is 3 and corresponds to inclusion of up to 3 lags of the corresponding information arrival indicator. FV_t is estimated using intraday data. Sample length differs from stock to stock but generally spans from 29 January 2001 to 31 July 2009. CVX, HPQ, INTC, and MSFT start at the following dates respectively: October 10, 2001; April 9, 2011; May 6, 2002; April 9, 2001. Statistical significance is marked by p-values next to the incremental R-square measure, * denotes a significance level at 0.01. ΔR^2_{+L} corresponds to models with all three components: An autoregressive part, as well as contemporaneous and lagged news.

	2001 to 2003			2004 to 2006			2007 to 2009		
Company	R^2	ΔR^2_{+N}	ΔR^2_{+L}	R^2	ΔR^2_{+N}	ΔR^2_{+L}	R^2	ΔR^2_{+N}	ΔR^2_{+L}
AA	37	6*	13*	35	7*	20*	29	12*	23*
AXP	29	6*	13*	32	13*	15*	28	8*	23*
BA	29	11*	21*	36	8*	15*	34	13*	22*
BAC	28	6*	21*	33	7*	3*	21	6*	-4
C	27	6*	20*	31	5*	4*	22	13*	0
CAT	33	10*	12*	34	12*	19*	32	13*	10*
CVX	28	10*	14*	38	5*	18*	29	4*	24*
DD	36	7*	17*	37	6*	17*	28	10*	21*
DIS	33	6*	13*	29	6*	26*	28	10*	21*
GE	26	7*	20*	36	7*	19*	24	12*	5*
HD	31	8*	14*	29	12*	15*	28	12*	23*
HPQ	29	12*	16*	33	11*	15*	28	8*	23*
IBM	27	3*	20*	37	8*	15*	28	7*	22*
INTC	26	9*	13*	32	8*	14*	25	6*	28*
JNJ	30	11*	16*	33	12*	18*	25	12*	26*
JPM	27	8*	21*	32	5*	19*	18	10*	27*
KO	31	8*	16*	36	12*	17*	25	12*	21*
MCD	31	20*	12*	35	13*	13*	28	15*	20*
MMM	30	9*	14*	36	14*	9*	26	15*	19*
MRK	29	5*	20*	33	4*	15*	30	15*	24*
MSFT	29	3*	5*	36	6*	18*	31	11*	19*
PFE	32	8*	18*	34	17*	14*	28	9*	23*
PG	31	8*	6*	35	12*	19*	29	10*	24*
UTX	31	4*	13*	34	8*	15*	25	14*	22*
VZ	29	8*	24*	34	12*	16*	24	13*	17*
WMT	29	4*	15*	37	7*	15*	30	10*	20*
XOM	30	3*	11*	41	7*	14*	29	4*	24*

Our third robustness analysis consists of splitting our sample into different periods and subsequently investigating the sensitivity of our results to this approach. We split our sample into three non-overlapping sub-samples corresponding to the ranges 2001 to 2003, 2004 to 2006 and 2007 to 2009. Comparing Table 11 with Table 5 it appears that in most cases the R^2 measures are somewhat higher. This result could be due to a true data generating process where new information has a time varying relationship with volatility, for example, the impact of similar corporate events, e.g. a new product launch, may be different depending on the state of the macroeconomic environment. For example, Boudoukh et al. (2007) find evidence that a state dependent model explains almost 50% of the return of orange juice futures on days where the temperature drops below the freezing point. While the robustness checks reveal that there is room for further improvement in the measurement and modeling framework this should not come as any surprise. The model and its relationships are unlikely to be stable through time as multiple aspects of the data generating process are changing continuously. We are investigating the relationship between economic information arrival and price movements, and instead of observing economic information we are observing a proxy, the positive change in the number of news items. As previously described, news items are the result of media industry participants creating, editing, aggregating and distributing information. Therefore, news item counts will be an imperfect measure. To partially infer economic information from the corporate newsflow, we utilize three strategies which are subject to measurement error.

First, we apply a corporate event taxonomy as implemented by the Dow Jones Intelligent Indexing system. This taxonomy has not been created with the specific purpose of identifying economic information used by investors for valuation purposes. In addition, there still exists uncertainty related to what amount of firm specific news that the indexing system has not identified and coded, as such there may be systematic exclusion across our sample that we are unaware of. Improving corporate event and information taxonomies, and documenting characteristics of corporate newsflow are clear steps forward in resolving such measurement issues.

Second, we rely on the positive change in the number of news items for a given subject category to observe the economic importance of the underlying information. The change in the level of news items may be determined by other factors than simply the arrival of new economic information. Several studies have suggested mechanisms by which corporations attempt to actively manage the corporate newsflow (Ahern and Sosyura 2011). Similarly, some studies suggest that actions by the media may be driven by factors other than the arrival of new economic information. For example, Veldkamp (2006) suggests that the low replication and distribution costs of information may induce media frenzies since market participants attempt to meet fluctuating information demands driven by movements in asset prices.

Finally, while we have attempted to gather as collectively exhaustive a dataset of firm specific newsflow as possible, the database that we use, while extensive, does not include all news sources. For example, any news published by Bloomberg's news agency, the fourth largest news agency by number of employees[16], is not completely included in our dataset. Bloomberg news items are only represented when other publications or newswires import content created by Bloomberg's news agency. Like the Dow Jones Factiva database, Bloomberg provides access to a wide range of news sources. The extent and impact of missing news sources has been examined and reveals that there are significant overlaps in the stories reported by leading news agencies. On a more general note, both Bloomberg and Factiva report that they provide access to more than 20,000 sources, suggesting a high degree of overlap between the two databases. As a final remark, we highlight our focus on news items as the source for measuring economic information arrival. For example, fillings with the U.S. Securities and Exchange Commission and other corporate disclosure suggests that industry supply and demand for aluminum should affect the economic performance

[16] As of May 2011.

of Alcoa (NYSE:AA). We have not included such sources of economic information. However, it should be noted that the firm specific news flow contains news items describing events related to aluminum and other commodities. Our results control for firm specific newsflow covering General & Political, Macroeconomic, and Financial Market subject categories to the extent news agencies and the Dow Jones Intelligent Indexing system mark such information as being relevant to the specific company. We think that any omitted sources of economic information arrival will be independent of the information identified and measured in this investigation.

6. **Summary and Conclusion**

Using an extensive collection of firm specific news we have confronted the paradigm that changes in stock prices are related to the arrival of new economic information. Our suggested return specification is a simple mixture model with two information processing components. The public information processing component is defined by the contemporaneous relationship with public information while the private information processing component is specified as a general autoregressive process corresponding to the sequential price discovery mechanism of investors as private processing of public information is incorporated into prices. Our empirical approach consists of estimating time-series regressions that approximate the components in our mixture model.

Our results show that changes in return volatility are related to public information arrival. For all 28 stocks in our sample, adding contemporaneous and lagged firm specific news explains a significant proportion of changes in firm specific return volatility. Firm specific volatility generally accounts for 58 to 77 percent of all variation in 5 minute returns within the trading day. Including contemporaneous and lagged indicators, of both economic information arrival and changes in media attention, explains between 5 to 20 percent of changes in firm specific volatility. In addition, we show that volatility persistence decreases between -.5 to -1.4% for models with both contemporaneous and lagged news indicators. This corresponds to a decrease in the half-life of volatility shocks of between 1/2 to 3 days. Robustness checks confirm that our measures of information arrival are indeed firm specific and capture relevant information related to the firm. Furthermore, robustness checks underscore the time varying relationship between volatility and firm specific news arrival.

Our primary contribution is to sample firm specific news in an extensive cross-section of news sources. Our approach highlights the challenge inherent in correctly measuring economic information arrival once all relevant news items are observed. In this respect, our investigation brings forth more questions than it provides answers. How should news arrival be measured? What is the relevant information set? How can we disentangle economic content from media attention effects? Theory provides little guidance on which specific information events are relevant for asset pricing.

In contrast to prior financial information research (Shiller 1981; Roll 1988; Cutler et al. 1989; Schwert 1989), this investigation supports the paradigm that public news arrival is related to changes in asset prices. Our study suggests that news arrival triggers some form of information processing, since clustering in public information only partly explains the degree of clustering in volatility, and lagged public information effects are significant. The question of whether private information processing effects are due to the sequential arrival of private information, resolution of information asymmetry, diminishing differences in opinion, or other information related sources of volatility clustering, is left unresolved.

References

Admati, A. and P. Pfleiderer (1988). A theory of intraday patterns: Volume and price variability. *Review of Financial Studies 1*(1), 3–40.

Ahern, K. and D. Sosyura (2011). Who Writes the News? Corporate Press Releases During Merger Negotiations. *Unpublished Working Paper*.

Andersen, T., T. Bollerslev, F. Diebold, and H. Ebens (2001). The distribution of realized stock return volatility. *Journal of Financial Economics 61*(1), 43–76.

Andersen, T. G. (1996). Return volatility and trading volume: An information flow interpretation of stochastic volatility. *Journal of Finance 51*(1), 169–204.

Andersen, T. G. and T. Bollerslev (1998). Deutsche mark-dollar volatility: Intraday activity patterns, macroeconomic announcements, and longer run dependencies. *Journal of Finance 53*(1), 219–265.

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

Barndorff-Nielsen, O., P. Hansen, A. Lunde, and N. Shephard (2009). Realized kernels in practice: Trades and quotes. *Econometrics Journal 12*(3), C1–C32.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2011). Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics 162*(2), 149 – 169.

Berry, T. D. and K. M. Howe (1994). Public information arrival. *Journal of Finance 49*(4), 1331–1346.

Black, F. (1976). Studies of stock price volatility changes. In *Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economics Statistics Section*, pp. 177–181.

Black, F. (1986). Noise. *Journal of Finance 41*(3), 529–543.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*(3), 307–327.

Boudoukh, J., M. Richardson, Y. Shen, and R. Whitelaw (2007). Do asset prices reflect fundamentals? Freshly squeezed evidence from the OJ market. *Journal of Financial Economics 83*(2), 397–412.

Campbell, J. (1991). A variance decomposition for stock returns. *Economic Journal 101*(405), 157–179.

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

Campos, J., N. R. Ericsson, and D. F. Hendry (2005). General-to-specific modeling: An overview and selected bibliography. International Finance Discussion Papers 838, Board of Governors of the Federal Reserve System (U.S.).

Christie, A. (1982). The stochastic behavior of common stock variances: Value, leverage and interest rate effects. *Journal of Financial Economics 10*(4), 407–432.

Clark, P. K. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica 41*(1), 135–156.

Copeland, T. (1976). A model of asset trading under the assumption of sequential information arrival. *Journal of Finance 31*(4), 1149–1168.

Copeland, T. and D. Friedman (1987). The effect of sequential information arrival on asset prices: An experimental study. *Journal of Finance*, 763–797.

Cutler, D., J. Poterba, and L. Summers (1989). What moves stock prices? *Journal of Portfolio Management 15*(3), 4–12.

DeGennaro, R. and R. Shrieves (1997). Public information releases, private information arrival and volatility in the foreign exchange market. *Journal of Empirical Finance 4*(4), 295–315.

Donders, M. and T. Vorst (1996). The impact of firm specific news on implied volatilities. *Journal of Banking & Finance 20*(9), 1447–1461.

Doornik, J. (2009). Autometrics. In *The Methodology and Practice of Econometrics*, Volume 1, Chapter 4, pp. 88–122. Oxford Scholarship Online Monographs.

Ederington, L. H. and J. H. Lee (1993). How markets process information: News releases and volatility. *Journal of Finance 48*(4), 1161–1191.

Ederington, L. H. and J. H. Lee (1996). The creation and resolution of market uncertainty: The impact of information releases on implied volatility. *Journal of Financial and Quantitative Analysis 31*(4), 513–539.

Engelberg, J. and C. Parsons (2011). The causal impact of media in financial markets. *Journal of Finance 66*(1), 67–97.

Engelberg, J., A. Reed, and M. Ringgenberg (2010). How are shorts informed? Short sellers, news, and information processing. *Unpublished Working Paper*.

Engle, R. (2004). Risk and volatility: Econometric models and financial practice. *American Economic Review 94*(3), 405–420.

Engle, R. and T. Bollerslev (1986). Modelling the persistence of conditional variances. *Econometric Reviews 5*(1), 1–50.

Engle, R. and C. Mustafa (1992). Implied ARCH models from options prices. *Journal of Econometrics 52*(1-2), 289–311.

Epps, T. W. and M. L. Epps (1976). The stochastic dependence of security price changes and transaction volumes: Implications for the mixture-of-distributions hypothesis. *Econometrica 44*(2), 305–322.

Fama, E. F. (1965). The behavior of stock-market prices. *Journal of Business 38*(1), 34–105.

Fang, L. and J. Peress (2009). Media coverage and the cross-section of stock returns. *Journal of Finance 64*(5), 2023–2052.

Fornari, F. and A. Mele (2001). Volatility smiles and the information content of news. *Applied Financial Economics 11*(2), 179–186.

Foster, F. and S. Viswanathan (1990). A theory of the interday variations in volume, variance, and trading costs in securities markets. *Review of Financial Studies 3*(4), 593–624.

French, K. (2008). Presidential address: The cost of active investing. *Journal of Finance 63*(4), 1537–1573.

French, K. and R. Roll (1986). Stock return variances: The arrival of information and the reaction of traders. *Journal of Financial Economics 17*(1), 5–26.

Goodhart, C., S. Hall, S. Henry, and B. Pesaran (1993). News effects in a high-frequency model of the sterling-dollar exchange rate. *Journal of Applied Econometrics 8*(1), 1–13.

Groß-Klußmann, A. and N. Hautsch (2011). When machines read the news: Using automated text analytics to quantify high frequency news-implied market reactions. *Journal of Empirical Finance 18*(2), 321 – 340.

Grossman, S. and R. Shiller (1981). The determinants of the variability of stock market prices. *American Economic Review 71*(2), 222–227.

Hansen, M. (2012). A look into the ant farm: The market for firm specific news. *Unpublished Working Paper, CREATES*.

Hendry, D. and H. Krolzig (2005). The Properties of Automatic GETS Modelling. *Economic Journal 115*(502), C32–C61.

Holden, C. and A. Subrahmanyam (1992). Long-lived private information and imperfect competition. *Journal of Finance*, 247–270.

Hoover, K. and S. Perez (1999). Data mining reconsidered: Encompassing and the general-to-specific approach to specification search. *Econometrics Journal 2*(2), 167–191.

Huberman, G. and T. Regev (2001). Contagious speculation and a cure for cancer: A nonevent that made stock prices soar. *Journal of Finance 56*(1), 387–396.

Kalev, P. S., W.-M. Liu, P. K. Pham, and E. Jarnecic (2004). Public information arrival and volatility of intraday stock returns. *Journal of Banking & Finance 28*(6), 1441–1467.

Kandel, E. and N. Pearson (1995). Differential interpretation of public signals and trade in speculative markets. *Journal of Political Economy 103*(4), 831–872.

Krolzig, H. and D. Hendry (2001). Computer automation of general-to-specific model selection procedures. *Journal of Economic Dynamics and Control 25*(6-7), 831–866.

Levy, H. and J. Yoder (1993). The behavior of option implied standard deviations around merger and acquisition announcements. *Financial Review 28*(2), 261–272.

Lovell, M. C. (1983). Data mining. *Review of Economics and Statistics 65*(1), 1–12.

Lumsdaine, R. (2009). What the market watched: Bloomberg news stories and bank returns as the financial crisis unfolded. *Unpublished Working Paper*.

Maheu, J. M. and T. H. McCurdy (2004). News arrival, jump dynamics, and volatility components for individual stock returns. *Journal of Finance 59*(2), 755–793.

Mandelbrot, B. (1963). The variation of certain speculative prices. *Journal of Business 36*(4), 394–419.

McWilliams, A. and D. Siegel (1997). Event studies in management research: Theoretical and empirical issues. *Academy of Management Journal 40*(3), 626–657.

Merton, R. (1974). On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance 29*(2), 449–470.

Mitchell, M. L. and J. H. Mulherin (1994). The impact of public information on the stock market. *Journal of Finance 49*(3), 923–950.

Patell, J. and M. Wolfson (1979). Anticipated information releases reflected in call option prices. *Journal of Accounting and Economics 1*(2), 117–140.

Patell, J. and M. Wolfson (1981). The ex ante and ex post price effects of quarterly earnings announcements reflected in option and stock prices. *Journal of Accounting Research 19*(2), 434–458.

Patell, J. and M. Wolfson (1984). The intraday speed of adjustment of stock prices to earnings and dividend announcements. *Journal of Financial Economics 13*(2), 223–252.

Press, S. J. (1967). A compound events model for security prices. *Journal of Business 40*(3), 317–335.

Reuters (2006, January). Reuters foundation reporter's handbook.

Roll, R. (1984). Orange juice and weather. *American Economic Review 74*(5), 861–880.

Roll, R. (1988). R2 (R squared). *Journal of Finance 43*(3), 541–566.

Ross, S. (1989). Information and volatility: The no-arbitrage martingale approach to timing and resolution irrelevancy. *Journal of Finance*, 1–17.

Schwert, G. W. (1989). Why does stock market volatility change over time? *Journal of Finance 44*(5), 1115–1153.

Shiller, R. (1981). Do stock prices move too much to be justified by subsequent changes in dividends? *American Economic Review 71*(3), 421–436.

Tauchen, G. E. and M. Pitts (1983). The price variability-volume relationship on speculative markets. *Econometrica 51*(2), 485–505.

Tetlock, P. (2010). Does public financial news resolve asymmetric information? *Review of Financial Studies 23*(9), 3520.

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

Thompson, R. B., C. Olsen, and J. R. Dietrich (1987). Attributes of news about firms: An analysis of firm-specific news reported in the wall street journal index. *Journal of Accounting Research 25*(2), 245.

Veldkamp, L. L. (2006). Media frenzies in markets for financial information. *American Economic Review 96*(3), 577–601.

Winkler, M. (2011). *The Bloomberg Way: a guide for reporters and editors.* Bloomberg Press.

Research Papers 2012

CREATES

Center for Research in Econometric Analysis of Time Series

2012-40: Mark Podolskij, Christian Schmidt and Johanna Fasciati Ziegel: Limit theorems for non-degenerate U-statistics of continuous semimartingales

2012-41: Eric Hillebrand, Tae-Hwy Lee and Marcelo C. Medeiros: Let's Do It Again: Bagging Equity Premium Predictors

2012-42: Stig V. Møller and Jesper Rangvid: End-of-the-year economic growth and time-varying expected returns

2012-43: Peter Reinhard Hansen and Allan Timmermann: Choice of Sample Split in Out-of-Sample Forecast Evaluation

2012-44: Peter Reinhard Hansen and Zhuo Huang: Exponential GARCH Modeling with Realized Measures of Volatility Statistics

2012-45: Peter Reinhard Hansen and Allan Timmermann: Equivalence Between Out-of-Sample Forecast Comparisons and Wald

2012-46: Søren Johansen, Marco Riani and Anthony C. Atkinson: The Selection of ARIMA Models with or without Regressors

2012-47: Søren Johansen and Morten Ørregaard Nielsen: The role of initial values in nonstationary fractional time series models

2012-48: Peter Christoffersen, Vihang Errunza, Kris Jacobs and Hugues Langlois: Is the Potential for International Diversi…cation Disappearing? A Dynamic Copula Approach

2012-49: Peter Christoffersen, Christian Dorion , Kris Jacobs and Lotfi Karoui: Nonlinear Kalman Filtering in Affine Term Structure Models

2012-50: Peter Christoffersen, Kris Jacobs and Chayawat Ornthanalai: GARCH Option Valuation: Theory and Evidence

2012-51: Tim Bollerslev, Lai Xu and Hao Zhou: Stock Return and Cash Flow Predictability: The Role of Volatility Risk

2012-52: José Manuel Corcuera, Emil Hedevang, Mikko S. Pakkanen and Mark Podolskij: Asymptotic theory for Brownian semi-stationary processes with application to turbulence

2012-53: Rasmus Søndergaard Pedersen and Anders Rahbek: Multivariate Variance Targeting in the BEKK-GARCH Model

2012-54: Matthew T. Holt and Timo Teräsvirta: Global Hemispheric Temperature Trends and Co-Shifting: A Shifting Mean Vector Autoregressive Analysis

2012-55: Daniel J. Nordman, Helle Bunzel and Soumendra N. Lahiri: A Non-standard Empirical Likelihood for Time Series

2012-56: Robert F. Engle, Martin Klint Hansen and Asger Lunde: And Now, The Rest of the News: Volatility and Firm Specific News Arrival