MONASH University

MONASH BUSINESS SCHOOL

DEPARTMENT OF ECONOMETRCS AND BUSINESS STATISTICS

ISSN 1440-771X

WORKING PAPER SERIES

Late to the Party or Insider Trading? Exploring Channels for the Rise of Volatility Before News

Klaus Ackermann, Ashley Andrews, Bonsoo Koo, and Wei Wei

Working Paper No. 21/24
December 2024

Late to the Party or Insider Trading? Exploring Channels for the Rise of Volatility Before News*

Klaus Ackermann[a], Ashley Andrews[b], Bonsoo Koo[a], and Wei Wei[† a]

[a]Department of Econometrics and Business Statistics, Monash University, Australia
[b]European Stability Mechanism

December 13, 2024

Abstract

We introduce the tardy news hypothesis as a new channel that complements information leakage to explain why volatility rises before public news. Through a novel information processing framework, we show that both channels can lead to bias and inconsistency in the estimated relationship between news and volatility, and we establish bounds for the true relationship. Using a dataset that combines realized volatility with RavenPack news analytics for S&P 500 constituents, we find that the type of events a news item covers is related to the likelihood of a rise in pre-news volatility, and the size of the firm is related to the magnitude of the rise. This heterogeneity from event and firm types is consistent with the tardy news hypothesis.

Keywords: Information leakage, Tardy news, Public news, Realized volatility
JEL codes: G14, C51

*We thank Heather Anderson, Elliott Ash, Adam Clements, Catherine Forbes, and the participants at the QFFE 2023 conference, the MPSS 2024 workshop, and the ESAM 2024 Conference for helpful comments and feedback.

†corresponding author: wei.wei2@monash.edu, Department of Econometrics and Business Statistics, Monash University, 900 Dandenong Road, Caulfield East, Victoria 3145, Australia.

1 Introduction

A central tenet of market efficiency is that all information is instantly incorporated into prices. The information explosion in recent decades, however, challenges how we process and digest information. A recent and growing literature on rational inattention (e.g. Sims, 2003; Van Nieuwerburgh and Veldkamp, 2010; Maćkowiak et al., 2023) suggests that not all available information is internalized; attention is scarce, and investors can choose what information to pay attention to. Against this background, we revisit the relationship between public information and return variance in this paper.

Prevailing finance theories, for example, the mixture of distributions hypothesis (MDH) in Clark (1973) and Andersen (1996), or the no-arbitrage martingale approach in Ross (1989), posit that return variance is directly linked to the rate of information arrival. Early empirical work (e.g. French and Roll, 1986) finds that return variance is mostly driven by private information rather than public news, although macroeconomic news does play a role, especially in the jump variation (e.g. Evans, 2011). Recent developments in machine learning and text analytics allow researchers to identify news that is relevant for specific firms, and they find that public news contributes significantly to return variances (e.g. Groß-Klußmann and Hautsch, 2011; Boudoukh et al., 2019; Engle et al., 2021; Kerssenfischer and Schmeling, 2024).

While the aforementioned studies agree that volatility reacts to news arrival, they often treat the time of news coverage as the *true* event time based on the assumption that traders with no private information only become aware of an event—and thus can act—when it is made public through news coverage. In other words, news articles are assumed to be

timely information about events that may affect a firm's fundamentals. Under the *timely* assumption, econometric models generally consider the impact of news only upon arrival and during the post-news period. For instance, early work by Kalev et al. (2004) and Kalev and Duong (2011) and recent work by Engle et al. (2021) and Kerssenfischer and Schmeling (2024) use the news count on day t and its lags to explain firm-specific variance on day t. Yet, some notable empirical work that does not impose the timely assumption finds that volatility increases before the time of the news. Specifically, Groß-Klußmann and Hautsch (2011) find that volatility (and volume) increases at least 60 minutes before the arrival of *unscheduled* firm-specific news. Boudoukh et al. (2019) document the time-series pattern of stock return variances on identified news days to test the theory that public information should resolve uncertainty. Their results also show an increase in volatility prior to the news day, although their discussion focuses on the post-news days only.

Volatility, or trading activities in general, can rise prior to the news arrival through several channels. The first is information leakage. An informed trader may receive *private* signals about information before it becomes public. The trader can trade on leakage twice, once when they receive the signal and again when the information becomes public, thus increasing volatility both before and upon the news arrival (e.g. Brunnermeier, 2005). Information leakage is the primary explanation for abnormal trading activity prior to news in the existing literature (e.g. Irvine et al., 2007; Christophe et al., 2010; Hendershott et al., 2015; Park, 2022).

In this paper, we propose another channel called the tardy news hypothesis to explain the rise of volatility prior to news. The tardy news hypothesis consists of two conjectures. First, not all public information is published as news first. Throughout this paper, we define *news* as

articles that appear in traditional channels, which include wired news providers such as *Dow Jones Newswires* as well as web-based publishers such as *Benzinga*. With the rapid expansion of information sources and distribution platforms, events can become public information through non-traditional channels, such as blogs and social media, before they appear as news. In other words, news may be *tardy* in reporting public events. Second, we assume the existence of two types of traders: the limited attention traders, who acquire public information only from news, and the full attention traders, who acquire public information whenever it becomes public. Full attention traders can act on a public event before it appears as news, resulting in rises in pre-news volatility. The limited attention traders act on information available as news, resulting in contemporaneous and post-news volatility rises.[1]

A case study on Xilinx Inc., a semiconductor manufacturing company, offers an example of tardy news. On Nov 30, 2016, the CEO of Amazon Web Services (AWS), Andy Jassy, gave a keynote speech introducing AWS's latest products and services at their annual "re:Invent" conference held in Las Vegas, Nevada. One of the announcements Andy Jassy made was the use of Field Programmable Gate Arrays (FPGAs) to increase the computation power of the cloud instances provided by AWS. In an official AWS blog post accompanying the keynote speech, Jeff Barr, the Chief Evangelist of AWS, specified a Xilinx product as the building block for these FPGAs. The blog post was published on Nov 30 at 8:55 am Pacific Standard Time (PST), or 11:55 am East Standard Time (EST), while the keynote speech occurred between 9 am and 11 am PST[2]. The first news coverage for this event in the Ravenpack database, however, occurred on the following day (Dec 1) at 7:47 am EST, when "The Fly",

[1]We analyze the volatility response to the news at the daily level and characterize volatility changes as contemporaneous if they occur on the same trading day as the news.

[2]The blog publishing time is obtained from the html source code of the page: https://aws.amazon.com/blogs/aws/developer-preview-ec2-instances-f1-with-programmable-hardware/.

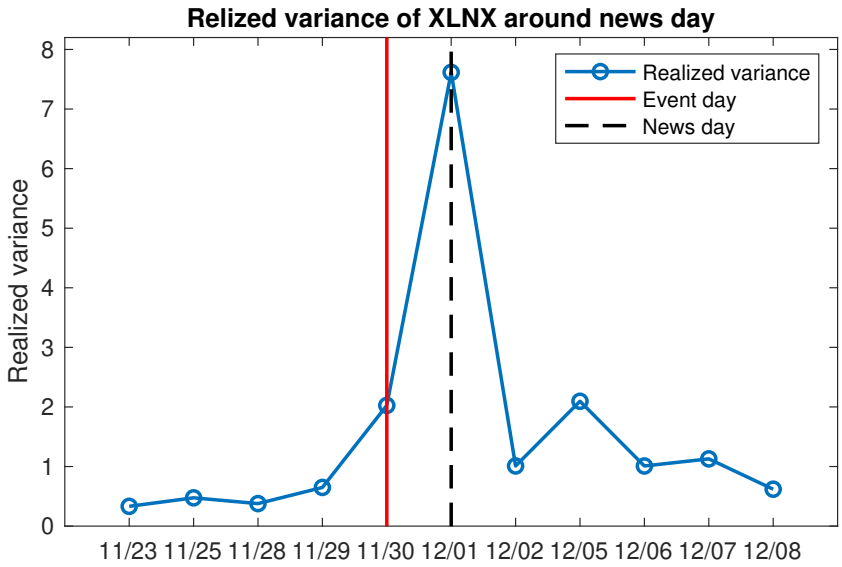

Figure 1: Realized variance of Xilinx (Ticker: XLNX) around the news day Dec 1, 2016 (black line). The event (Amazon announcing collaborations with Xilinx) occurs at 11:55 am, Nov 30. The first news related to this event was released at 7:47 am, Dec 1.

a popular web-based financial news provider, released an article about an analyst's "buy" recommendation for Xilinx due to Amazon's collaboration with the manufacturer.

Figure 1 depicts the daily realized variance of Xilinx around the news day. While volatility increases substantially on Dec 1, there is also a perceivable volatility rise on Nov 30. Note that there is no news item for Xilinx on Nov 29 or 30 that may offer an alternative explanation for the volatility rise on Nov 30.

Documenting how volatility rises prior to news and investigating the role the timeliness of news have three important implications. First, without accounting for the elevated volatility in the days preceding news, estimates on the contribution of public information to return variance will be biased. Second, without the channel of tardy news, a rise in volatility prior to news arrival may be wrongly attributed to information leakage, increasing the market's concern for insider trading. Third, analyzing how investors process information reveals empirical features of rational inattention and suggests new directions for developing economic

theories to better explain investor behavior.

To examine the pre-news volatility rise formally, we develop a novel information processing framework based on the information processing hypothesis (IPH) in Engle et al. (2021) and the mixture of distributions hypothesis (MDH) (see Clark (1973) and Andersen (1996)). Our framework allows private and public information to have separate effects on volatility as in IPH, and it relates volatility to private information using MDH. In contrast to MDH and IPH, we allow for information leakage by specifying a correlation between past private information and future public information. We also accommodate tardy news by allowing for volatility responses before a news article.

Under this new framework, we analyze how information leakage and tardy news affect the regressions commonly used for estimating the effect of public news on volatility. We show that although standard regression coefficient estimators are biased and inconsistent, we may analyze the direction of the bias. By combining regressions with upward-biased estimates and those with downward-biased estimates, we establish bounds on the effect of news.

We focus on firm-specific news in our empirical study as in Boudoukh et al. (2019). Using the equity news package by Ravenpack, which is considered the industry standard in news analytics, we identify highly relevant and novel stories regarding individual firms. To obtain dependent variables with higher precision, we compute realized variance for each stock using high-frequency transaction data from the New York Stock Exchange (NYSE) Trade and Quote (TAQ) database. Our dataset comprises the constituents of S&P 500 that traded continuously in our sample period from January 2015 to October 2020.

To investigate the reason behind pre-news volatility rises and examine competing hypotheses for explaining the dynamics of volatility around the news, we first differentiate between

scheduled news (also called anticipated news as in Huang et al. (2020)) and *unscheduled* news. Scheduled news, such as an earnings announcement, does not suffer from any "tardiness". Hence, only the pre-news volatility rise in *unscheduled* news may be attributable to the tardy news hypothesis. Furthermore, we group news according to the type of events it covers. We consider five event groups: *accounting*, which covers earnings/revenue/dividends reports and forecast; *analysis*, which covers third-party ratings or recommendations; *equity*, which covers changes in shareholder structure; *media*, which covers recent price movements; and *operations*, which covers issues related to the operation of a firm.

We find that news that directly reports fundamental information, such as *accounting* and *operations*, is associated with only contemporaneous volatility rises, whereas *Secondary* news that reports interpretations of or responses to fundamental events, such as the *analysis* and *media* news, experiences significant pre-news volatility rises. The *equity* news experiences moderate pre-news and contemporaneous volatility effects. As *secondary* news is more likely to report information that became public prior to being picked up by news outlets, the tardy news hypothesis predicts that secondary news is more likely to experience pre-news volatility rises. Our empirical result is consistent with this prediction.

Furthermore, the tardy news hypothesis predicts that firms with higher investor attention are less likely to experience pre-news volatility rises. We use firm size as a proxy for investor attention and separate the constituents of S&P 500 into three groups, namely large, which contains constituents of the Dow Jones Industrial Average (DJIA); medium, which contains constituents of the S&P100 index that are not in the DJIA; and small, the rest of the S&P 500 constituents. We find that larger firms experience smaller pre-news volatility rises, providing further support to the tardy news hypothesis.

The remainder of the paper is structured as follows. Section 2 describes our information processing framework, and Section 3 the regression models. Section 4 details the data sources and dataset construction. Section 5 presents the average effect of news, and Section 6 examines the pre-news volatility rises. Section 8 concludes.

2 Information processing with tardy news and information leakage

In this section, we develop a framework relating information arrival and asset pricing while incorporating the information leakage and the tardy news channel. Our framework is based on the mixture of distributions hypothesis (MDH) (see Clark (1973) and Andersen (1996)) and the information processing hypothesis (IPH) in Engle et al. (2021). The MDH postulates that the variance of the return in an arbitrary trading day is proportional to the intensity of information arrivals on that day and the random and positively correlated arrival process leads to an autoregressive stochastic volatility (SV) process for returns. The information processing hypothesis in Engle et al. (2021) postulates that the market processes private information differently from public information, resulting in distinct dynamics between the volatility due to public news and private news. Engle et al. (2021) model the variance due to private news by GARCH, and assume that the variance induced by public news is proportional to the number of news.

Let K_{it} denote the number of *private* information arrival for stock i on day t and V_{it}^{priv} the daily variance due to private information. From the MDH, the return over day t for stock

i is normally distributed conditional on K_{it}. In other words, V_{it}^{priv} is proportional to K_{it}. We assume that K_{it} is a stationary process that admits a moving average representation, and the average number of information arrivals per day is large, so K_{it} can be approximated by a continuous distribution. Then, V_{it}^{priv} has a moving average representation:

$$V_{it}^{\text{priv}} = \alpha_i + \sum_{j=0}^{\infty} \beta_j u_{i,t-j}, \tag{1}$$

where β_j measures the persistence of V_{it}^{priv} with $\sum_{j=0}^{\infty} |\beta_i| < \infty$, α_i measures the unconditional mean of V_{it}^{priv}, and $u_{i,t}$ is the innovation in the intensity of private information with $u_{i,t} \overset{i.i.d.}{\sim} (0, \sigma_u^2)$.[3] Note that if $\beta_j = \beta^j$ for $j \geq 0$, Equation (1) yields an AR(1) structure for V_{it}^{priv}.

Next, let V_{it}^{pub} denote the daily variance due to *public* information and n_{it} the number of public news for stock i on day t, we assume that

$$V_{it}^{\text{pub}} = \sum_{j=-\underline{J}}^{\overline{J}} \gamma_j n_{i,t-j}, \tag{2}$$

where γ_j measures the impact of public news arrived j periods ago/ahead, with $\sum_{j=-\underline{J}}^{\overline{J}} |\gamma_j| < \infty$.[4] Furthermore, we assume that n_{it} is a stationary process with no autocorrelation.

The specification in Equation (2) incorporates two salient features of information processing in the existing literature. First, we allow the public information to be processed at a different

[3] Equation (1) does not guarantee the positivity of V_{it}^{priv}. Nonetheless, we can define $\widetilde{V}_{it}^{\text{priv}} = \max(\alpha_i + \sum_{j=0}^{\infty} \beta_j u_{i,t-j}, \alpha_i)$, then $V_{it}^{\text{priv}} \approx \tilde{V}_{it}^{\text{priv}}$ given large α_i.

[4] While there is a growing literature on linking the jumps in returns to public news (e.g. Jeon et al. (2022), Timmermann et al. (2023) and Aït-Sahalia et al. (2024)), separating a price jump and a burst of volatility is difficult and many tests suffer from spurious detection (see e.g. Christensen et al. (2014) and Bajgrowicz et al. (2016)). Hence, we do not distinguish between the volatility due to jumps and the volatility due to the continuous component of the return process in this paper. Hence, $V_i t$ is a measure of the quadratic variation rather than integrated variance.

speed than private information, and hence V_{it}^{pub} possesses different persistence than V_{it}^{priv}. That is, $\gamma_j \neq \beta_j$. Second, we allow $\overline{J} > 0$ to reflect the growing empirical evidence that the private processing of public information may take time. For example, Tetlock et al. (2008) find that abnormal returns display delayed (up to two days) reactions to the Dow Jones News Service (DJNS) and Wall Street Journal (WSJ) news stories. Tetlock (2010) theorizes that informed traders have superior information but face a liquidity shock and that public news resolves asymmetric information. He finds evidence supporting this theory and that uninformed traders provide liquidity gradually (over a nine-day span) rather than instantly after the news.

The innovation in Equation (2) is the accommodation of the tardy news hypothesis by allowing $\underline{J} > 0$ and $\gamma_{-j} \neq 0$ for $j = 1, \ldots \underline{J}$. Under the tardy news hypothesis, when an event becomes public information on day $t - \underline{J}$, the full-attention traders start processing the information immediately, whereas the limited-attention traders only start acting when the news appears in print on day t. Hence, $n_{i,t}$ will be affecting volatility from day $t - \underline{J}$ to $t + \overline{J}$ with the impact measured by $\gamma_{-\underline{J}}, \ldots, \gamma_{\overline{J}}$.

Another important empirical feature in financial markets is information leakage—some traders may acquire firm-specific information in private before it is released to the public (e.g. Brunnermeier (2005), Irvine et al. (2007), and Indjejikian et al. (2014)). To accommodate this feature, we allow the number of public news, $n_{i,t}$, to be positively correlated with the intensity of *past* private news. Note that the presence of tardy news means that events become public knowledge prior to the news. Therefore, private information must be acquired prior to the event day rather than the news day. Suppose that private information can be acquired between one day to J^* days ahead of an event, and the news appears $\underline{J}$ days after

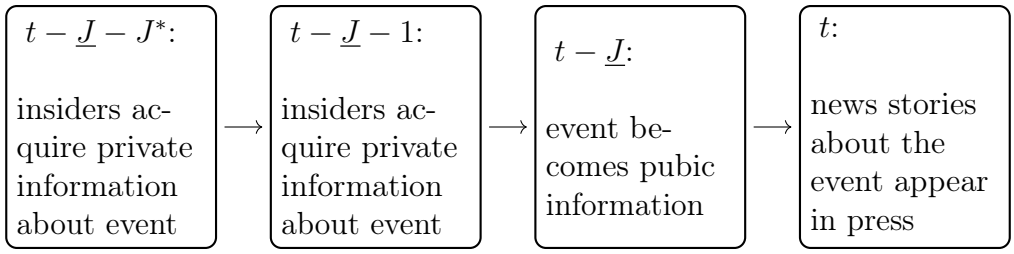

Figure 2: Timeline of information processing around one event

the event occurs. We define information leakage as $\mathsf{cov}(u_{i,t-\underline{J}-j}, n_{i,t}) > 0$ for $j = 1, \ldots, J^*$, while $\mathsf{cov}(u_{i,t-j}, n_{i,t}) = 0$ for $j \leq \underline{J}$. Figure 2 illustrates the timeline around one event.

Let r_{it} denote the idiosyncratic return for stock i, we assume that $r_{it} = r_{it}^{\text{priv}} + r_{it}^{\text{pub}}$, where $r_{it}^{\text{priv}} = \sqrt{V_{it}^{\text{priv}}}\eta_{it}^{\text{priv}}$ and $r_{it}^{\text{pub}} = \sqrt{V_{it}^{\text{pub}}}\eta_{it}^{\text{pub}}$, and that $\mathsf{cov}(\eta_{it}^{\text{priv}}, \eta_{it}^{\text{pub}}) = 0$. Then the idiosyncratic variance of stock i, denoted by V_{it}, is given by

$$V_{it} = V_{it}^{\text{priv}} + V_{it}^{\text{pub}}. \tag{3}$$

We are interested in using regressions to estimate the effect of public news, γ_j, in a setting where we observe V_{it} and public information flow n_{it}, but not the private information. We first consider the following two regressions:

$$V_{it} = \alpha_i + \sum_{j=-J}^{J} \theta_j^{(1)} n_{i,t-j} + e_{it}^{(1)}, \tag{4}$$

$$V_{it} = \alpha_i + \phi V_{i,t-1} + \sum_{j=-J}^{J} \theta_j^{(2)} n_{i,t-j} + e_{it}^{(2)}, \tag{5}$$

where $J \geq 0$. These regressions, or their restricted versions, are commonly used in the literature. For example, Boudoukh et al. (2019) use Equation (4) with $J = 0$ to quantify the effect of public news on volatility, and Equation (5) is used in Engle et al. (2021) and

Kerssenfischer and Schmeling (2024). In this section, we aim to analyze whether the OLS estimates $\hat{\theta}_j^{(1)}$ and $\hat{\theta}_j^{(2)}$ (or transformations of these estimates) yield unbiased/consistent estimators for γ_j as $T \to \infty$ in the presence of information leakage and tardy news.

We first consider the case with no information leakage, that is, when public news n_{it} and private information u_{it} are uncorrelated at all leads and lags. In this case, we show in the appendix $e_{it}^{(1)}$ in Regression (4) is uncorrelated with the regressors, and hence $\hat{\theta}_j^{(1)}$ from the OLS will be unbiased and consistent estimator of γ_j for $j = -J, \ldots, J$.

For Regression (5), we start by assuming that $\beta_j = \beta^j$, i.e., the private variance follows an AR(1) process, and that $J \geq \max\{\underline{J}, \overline{J} + 1\}$. In this case, we can construct estimators for γ_j by

$$\hat{\gamma}_j = \begin{cases} \hat{\theta}_j^{(2)}, & j = -J, \\ \hat{\theta}_j^{(2)} + \hat{\phi}\hat{\gamma}_{j-1}, & j = -J+1, \ldots, J. \end{cases} \tag{6}$$

We show in the appendix that $\hat{\gamma}_j \xrightarrow{p} \gamma_j$ for $j = -J, \ldots, J$.

If we believe that the dynamics of private variance are better captured by a different structure, for example, if V_{it}^{priv} follows an AR(p) process,

$$V_{it}^{\text{priv}} = \alpha_i + \psi_1 V_{i,t-1}^{\text{priv}} + \cdots + \psi_p V_{i,t-p}^{\text{priv}} + u_{it},$$

we can modify Regression (5) to

$$V_{it} = \alpha_i + \phi_1 V_{i,t-1} + \cdots + \phi_p V_{i,t-p} + \sum_{j=-J}^{J} \theta_j^{(2)} n_{i,t-j} + e_{it}^{(2)}. \tag{7}$$

As long as the autoregressive part of Regression (7) captures the dynamics of V_{it}^{priv} and

$J \geq \max\{\underline{J}, \overline{J} + p\}$, transformations similar to Equation (6) can be applied to obtain an estimator $\hat{\gamma}_j$ that converges to γ_j for $j = -J, \ldots, J$. In summary, we have the following proposition.

Proposition 1. *If there is no information leakage, the OLS estimator $\hat{\theta}_j^{(1)}$ from Regression (4) is a consistent estimator for γ_j for $j = -J, \ldots, J$. For Regression (5) (or Regression (7)), if the dynamics for private variance is correctly specified and $J \geq \max\{\underline{J}, \overline{J} + 1\}$ (or $J \geq \max\{\underline{J}, \overline{J} + p\}$ for Regression (7)), the transformation in Equation (6) (or a transformation similar to Equation (6)) yields a consistent estimator for γ_j.*

Proof. See appendix. □

Remark. *In the absence of information leakage, the pre-news rise in volatility is truly the effect of public news.*

Remark. *The regressors $n_{i,t-j}$ are assumed to be independent from each other, so omitting $n_{i,t-j}$ has no effect on the properties of $\theta_{j^*}^{(1)}$, for any $j^* \neq j$. For the consistency of $\theta_j^{(2)}$, however, J needs to be larger than p to eliminate the correlation between lagged $V_{i,t}$ and $e_{it}^{(2)}$.*

Next, we consider the case of information leakage. Specifically, we assume that $\mathsf{cov}(u_{i,t-\underline{J}-j}, n_{i,t}) > 0$, for $j = 1, \ldots, J^*$. In this case, $e_{it}^{(1)}$ in Regression (4) is correlated with n_{t+j} for any $j \leq \underline{J}+J^*$, and standard OLS yields the following proposition.

Proposition 2. *In the presence of information leakage, $\hat{\theta}_j^{(1)}$ from Regression (4) will be biased and inconsistent for any $j \geq -\underline{J} - J^*$. Specifically, let $\rho_j = \mathsf{cov}(n_{i,t}u_{i,t-\underline{J}-j}) > 0$ for $j = 1, \ldots, J^*$ and set $\beta_j = 0$ for $j < 0$, we have*

$$\hat{\theta}_j^{(1)} \xrightarrow{p} \gamma_j + \ var(n_{i,t})^{-1} \sum_{k=1}^{J^*} \rho_k \beta_{j+\underline{J}+k}, \tag{8}$$

for any $j \geq -\underline{J} - J^*$.

Proof. See appendix. □

Remark. *The bias in* $\hat{\theta}_j^{(1)}$ *does not depend on what leads and lags other than* j *are included in the regression.*

Remark. *Empirical studies on variance dynamic suggest that* β_j *is typically positive for* $j \geq 1$. *Hence, the second term in (8) is positive, and* $\theta_j^{(1)}$ *is an upward-biased estimator of* γ_j. *Moreover, if* $\beta_j > \beta_{j-1}$ *and* $\rho_j > \rho_{j-1}$ *for any* j, *the bias is most severe when* $j = -\underline{J} - J^*$, *and it decreases as* j *increases.*

For Regression (5), the effect of information leakage depends on $\underline{J}$. Suppose that the variance dynamics are correctly specified and we know the value of $\underline{J}$, we can set $\hat{\theta}^{(2)}_{-\underline{J}-j} = 0$ for $j > 0$ and obtain consistent estimator of γ_j. Note that this is equivalent to omitting any $n_{i,t-j}$ if $j < -\underline{J}$. For example, suppose that there is no tardy news ($\underline{J} = 0$) and $\beta_j = \beta^j$, the following regression

$$V_{it} = \alpha_i + \phi V_{i,t-1} + \sum_{j=0}^{J} \theta_j^{(3)} n_{i,t-j} + e_{it}^{(3)}, \tag{9}$$

and the transformation

$$\hat{\gamma}_j = \begin{cases} 0, & \text{if } j < 0 \\ \hat{\theta}_j^{(3)}, & \text{if } j = 0 \\ \hat{\theta}_j^{(3)} + \hat{\phi}\hat{\gamma}_{j-1}, & \text{if } j = 1, \ldots, J \end{cases} \tag{10}$$

yields consistent estimator as $\hat{\gamma}_j \xrightarrow{p} \gamma_j$ for $j = -J, \ldots, J$ and $\hat{\phi} \xrightarrow{p} \beta$.

In practice, we do not know the value of $\underline{J}$. We propose to use regression (9) with transformation (10) and summarize the effect of omitting $n_{i,t+1}, \ldots, n_{i,t+\underline{J}}$ in the following proposition.

Proposition 3. *Suppose that $\beta_j = \beta^j$. In the presence of information leakage and tardy news, Regression* (9) *with transformation* (10) *yields*

$$\hat{\gamma}_j \xrightarrow{p} \gamma_j - \beta^{j+1}\gamma_{-1},$$

for $j \geq 0$.

Proof. See appendix. □

Remark. *If the tardy news has a positive effect on volatility ($\gamma_{-1} > 0$), the estimators from (9) and (10) will be downward-biased relative to the true effects of news. Moreover, the bias decreases as j increases.*

If the private variance is better captured by an AR(p) process, we can include p lags of V_{it} in (9) and use a transformation similar to (10) to obtain consistent estimator of γ_j when $\underline{J} = 0$. When $\underline{J} > 0$, the estimator will still be downward-biased if $\gamma_{-j} > 0$ for $j = 1, \ldots, \underline{J}$.

3 Regression model construction

The daily idiosyncratic variance V_{it} is not directly observable, and squared residuals from regressing returns on the market factor are commonly used as a proxy (e.g. Boudoukh et al. (2019)). We adopt panel data models with interactive fixed effects by Bai (2009) which

estimates the market factor from the data and obtain the news coefficients simultaneously. Furthermore, we use the realized variance (RV) as the dependent variable as it provides a less noisy measurement of the variance than squared daily returns.

3.1 Realized variance

Suppose that the continuous-time logarithm price for firm i is sampled at M points over the one-day interval $[t-1, t]$. Let r_{i,t_j} denote the continuously compounded return during the interval $[t-1+\frac{j-1}{M}, t-1+\frac{j}{M}]$, where $j = 1, \ldots, M$, the realized variance for day t is defined by

$$RV_{it} = \sum_{j=1}^{M} r_{i,t_j}^2. \tag{11}$$

If there is no market microstructure noise, RV converges to the quadratic variation of the continuous-time logarithm price when M goes to infinity (Barndorff-Nielsen and Shephard (2002), Andersen et al. (2001)). Empirically, coarser intervals are adopted to mitigate the effect of microstructure noises.

3.2 News grouping

Different news may have different impacts on volatility. For example, Jiang et al. (2012) find that scheduled news resolves information uncertainty and decreases volatility, while unscheduled news creates uncertainty and increases volatility. We group news by their characteristics and allow different groups of news to have heterogeneous effects on volatility. Specifically, let $\mathcal{K}$ define a set of K disjoint news groups. We define $n_{k,i,t}$ as the number

of news in group $k \in \mathcal{K}$ that company i has on day t. From our information processing framework, the true effect of public news is given by

$$V_{it}^{\text{pub}} = \sum_{k=1}^{K} \sum_{j=-\underline{J}}^{\overline{J}} \gamma_{k,j} n_{k,i,t-j}, \tag{12}$$

where $\underline{J} > 0$ indicates the presence of tardy news as in Equation (2). Information leakage for news group k is defined by $\mathsf{cov}(u_{i,t-\underline{J}-j}, n_{k,i,t}) > 0$ for $j = 1, \ldots, J^*$.

3.3 Panel regression models

In the capital asset pricing model, the return of asset i is linearly related to the return of the market portfolio, usually approximated by the S&P500 index, through its beta. With time-varying variance, the day t variation of asset i can be decomposed into the idiosyncratic variance V_{it} and a market component $\lambda_i V_{mt}$, where V_{mt} is the day t variation of the market portfolio, and λ_i equals the squared beta (see e.g. Engle et al. (2021)). When the number of stocks is large, we can estimate the common variance factor of the stocks under examination directly from data using the principle components analysis (PCA). This is akin to the asset pricing models with latent, rather than observable, factors (see e.g. Connor and Korajczyk (1988)). To estimate both the factors and news impacts, we specify an interactive fixed effect (IFE) model as follows:

$$\text{IFE: } RV_{it} = \alpha_i + \lambda_i f_t + \sum_{k=1}^{K} \sum_{j=-J}^{J} \theta_{k,j} n_{k,i,t-j} + \epsilon_{i,t}, \tag{13}$$

where α_i is the individual fixed effect that does not change over time, f_t is a common factor in variance, and λ_i the factor loading. We allow one factor in this paper, but the number of factors can be easily extended to more. Note that α_i, λ_i and f_t are not separately identifiable when f_t is not observed. We achieve identification by setting $\sum_{t=1}^{T} f_t = 0$ and $\sum_{t=1}^{T} f_t^2 = 1$. Under our identification strategy, α_i is equal to the unconditional mean of RV_{it} in the absence of news, i.e., $\alpha_i = E(RV_{it}|\text{no news})$.

The IFE panel model nests the commonly used two-way additive fixed effects (AFE) models as a special case. For comparison, we also consider the AFE regression:

$$\text{AFE: } RV_{it} = \alpha_i + f_t + \sum_{k=1}^{K} \sum_{j=-J}^{J} \theta_{k,j} n_{k,i,t-j} + \epsilon_{i,t}, \tag{14}$$

with $\sum_{t=1}^{T} f_t = 0$. In the absence of information leakage, there exists a consistent estimator $\hat{\theta}_{k,j}$ for $\theta_{k,j}$ from either equation (13) or (14), which will provide consistent estimator for $\gamma_{k,j}$ as both N and T tend to infinity. In the presence of information leakage, however, the estimators from both (13) and (14) will be upward-biased, and the severity of bias depends on both the correlation between public and private information, $E(n_{i,t} u_{i,t-\underline{J}-j})$, and the autocorrelation structure of V_{it}^{priv} as shown in Proposition 2.

To model the dynamic of V_{it}^{priv}, we use a restricted AR structure motivated by the HAR model in Corsi (2009). Specifically, we include lagged averages of weekly RV, in addition to lagged daily RV, to account for long-range dependence parsimoniously. Our interactive fixed

effect (IFE) with a restricted AR model is specified as follows:

$$\text{IFE-AR: } RV_{it} = \alpha_i + \lambda_i f_t + \phi_1 RV_{i,t-1} + \frac{\phi_2}{5}\sum_{j=1}^{5} RV_{i,t-j} + \sum_{k=1}^{K}\sum_{j=0}^{J}\theta_{k,j} n_{k,i,t-j} + \epsilon_{i,t}. \tag{15}$$

Similarly, the two-way fixed effect with a restricted AR model is given by

$$\text{AFE-AR: } RV_{it} = \alpha_i + f_t + \phi_1 RV_{i,t-1} + \frac{\phi_2}{5}\sum_{j=1}^{5} RV_{i,t-j} + \sum_{k=1}^{K}\sum_{j=0}^{J}\theta_{k,j} n_{k,i,t-j} + \epsilon_{i,t}. \tag{16}$$

Transformations similar to (10) will provide a consistent estimator of $\hat{\gamma}_{k,j}$ if there is no tardy news. First, let $\eta_1 = \phi_1 + \frac{1}{5}\phi_2$ and $\eta_2 = \frac{1}{5}\phi_2$. Then

$$\hat{\gamma}_{k,j} = \begin{cases} 0, & \text{if } j < 0 \\ \hat{\theta}^{(3)}_{k,j}, & \text{if } j = 0 \\ \hat{\theta}^{(3)}_{k,j} + \hat{\eta_1}\hat{\gamma}_{k,j-1} + \hat{\eta_2}\sum_{l=2}^{5}\hat{\gamma}_{k,j-l}, & \text{if } j = 1, \ldots, J \end{cases} \tag{17}$$

In the presence of tardy news, $\hat{\gamma}_{k,j}$ from either regression (15) or (16) and transformation (17) will be downward-biased, as shown in Proposition 3.

4 Data and Summary Statistics

We consider stocks that are members of the S&P 500 as of January 2020[5]. Our sample period is from January 2015 to October 2020. To have a balanced panel, we filter out firms that do not have trading activities in any of the trading days within our sample period. The final

[5]Constituents of the S&P 500 were identified using the Bloomberg Terminal along with their GICS taxonomy, CUSIP IDs, and ISIN IDs. The CUSIP and ISIN IDs are used to create a lookup table that links Ravenpack's entity IDs (RP_ENTITY_ID) to each of the constituents.

sample contains 426 out of the total S&P 500 firms, each with 1510 daily observations.

4.1 News Data - Ravenpack

We use the news analytics data from *RavenPack - Full edition*, which aggregates news from three products: the Dow Jones (DJ) product, the Web Edition (WEB), and the Press Release Edition (PR). The DJ product contains news from premium subscription services including *Dow Jones Newswires*, *The Wall Street Journal* and *Marketwatch*. The WEB edition tracks web publishers, such as *Benzinga* and *Ticker Report*, and local news, which are free or charge a small subscription fee. The PR edition includes news from press releases and regulatory disclosures.

We filter the news according to their relevance following Boudoukh et al. (2019), who highlight the importance of relevant news in the news-volatility relationship. *RavenPack* provides a relevance score that is measured based on the location and frequency with which an entity is referred to in the news text. We only retain the news with a relevance score greater than 90, which indicates that the entity is mentioned in the headline of the news item. Furthermore, we filter out the news that can not be classified into any event category.

To mitigate the impact of stale information or "news cluster" as those discussed in Tetlock (2011) and Huang et al. (2020), we only use the *first* news reporting the same event in a 24-hour window by retaining news with a Global Event Novelty Score (G_ENS) equal to 100. After filtering, we obtain 1,396,644 news items that are novel and highly relevant to the 426 stocks under consideration during our sample period.

We assume that limited attention traders acquire information through Ravenpack or

similar providers. Hence, examining the nature of the information available in Ravenpack is important. We first look at the sources of news. Similar to Huang et al. (2020), we find that news from *Dow Jones Newswires* only comprises a small percentage (21.7%) of our dataset. The top contributor is *Ticker Report*, a web-based publisher, who provided 24% of news. We list the top ten news sources with their respective news counts and percentages in Appendix B.

Next, we group news according to whether the news is scheduled (*sch*) or unscheduled (*unsch*); whether the news is timestamped within the trading hours (*trdng*) or overnight (*ovrnt*); whether the news sentiment is positive (*pos*), neutral (*ntrl*), or negative (*neg*); and which event category the news belongs to. For the time stamp, we consider news that is timestamped between 16:00 on day $t-1$ and 9:30 on day t as overnight news for day t, while those timestamped between 9:30 and 16:00 on day t as trading hour news for day t[6]. We use Event Sentiment Score (ESS) provided by Ravenpack for the news sentiment. The ESS score ranges from 0 to 100, with a value less than 50 indicating negative sentiment, larger than 50 indicating positive sentiment, and equal to 50 indicating neutrality.

RavenPack classifies news items into more than 37 categories based on the type of event the news is covering. Appendix B lists the 20 most common event categories, along with the number of news in each category and descriptions of typical events falling under that category. The *RavenPack* event categories contain topics that are similar to each other, and we classify the 37 event categories into five broader groups, namely *accounting* , *analysis*, *equity*, *media*, and *operations*, based on topic similarity. Table 1 summarizes our groupings

[6]The RavenPack timestamp is converted to New York standard time to match the transaction data from the TAQ database.

Table 1
Description of Event Categories

Event group	Ravenpack event category	Description
Accounting	earnings, revenues, dividends	News about the announcements or the forecasts of a firm's key accounting quantities such as earnings, revenue and dividends.
Analysis	analyst-ratings, technical-analysis , price-targets, credit-ratings.	News about financial analysts or institutions advising on buy/hold/sell a firm's stock, predicting its future price, or rating a firm's bond.
Equity	equity-actions, acquisitions-mergers, insider-trading	News about a firm's shareholder structure such as acquisitions & mergers or shareholders changing positions.
Media	investor-relations, stock-prices	News related to media exposure. "Investor-relations" are usually press-releases and earnings conference calls; "stock-prices" are news about recent price movements.
Operations	assets, credit, labor-issues, legal, products-services, regulatory, and others	News that do not fall into the prior four groups. The news under this category primarily relate to the operations of a firm. For example, "assets" relate to the sale/purchase of facilities, "labor-issues" relate to change of personnel such as CEOs.

Note: This table summarizes the five event groups in our study based on topics. The second column details the Ravenpack event categories that are assigned to each of our groups, while the third column provides a more detailed description of the topic in each category.

and the *RavenPack* categories forming the five groups.

After forming 60 news groups based on the event type, the sentiment, the timing and the scheduling, we truncate the number of news within the same group for the same firm/day combination to one. The truncation further alleviates possible stale news that repeats information from previous articles, and it allows at most 60 news items for each firm/day observation if they are of different event types or have different timing or scheduling or sentiment.

After truncation, the total number of news in our sample is 1,009,444, among which 82.0% are unscheduled, and 63.8% occur overnight. Positive sentiment stories comprise 48.5% of news items, while negative and neutral news make up 33.4% and 18.05%, respectively. News from the *equity* category occurs most frequently (35.3% of all news), followed by *analysis* (25.7%), *operations*(15.7%), *accounting*(13.6%), and *media*(9.7%).

The news stories are unevenly distributed across firms or days. Out of the 643,260 firm/day observations, 31.3% have no news, 29.4% have one piece of news, and 39.3% have at least two pieces of news. Table 2 reports the percentage of news days, defined as the percentage of firm/day observations with a given group of news, for each group.

From Table 2, 52 out of the 60 news groups have non-zero observations. News in the *accounting* category, which includes earnings and revenues reports and forecasts, is more likely to be scheduled and occur overnight. The *analysis* category only has unscheduled news, as third-party ratings or recommendations are usually released around earnings announcements but do not have a set time frame. The sentiment of *analysis* news is usually non-neutral. *Equity* and *operation* related news is mostly unscheduled, more likely to occur overnight, and carry non-neutral sentiments. Scheduled *media* news is mostly neutral as it is press releases, whereas unscheduled *media* news usually carries a non-neutral sentiment as it is about large movements of the firm's stock prices.

4.2 Realized Variance - NYSE TAQ

We use high-frequency trade data from the New York Stock Exchange (NYSE) Trade and Quote (TAQ) database. Following the data-cleaning procedures outlined in Barndorff-Nielsen

Table 2
Percentage of news days

		ovrnt		*trdng*	
		sch	*unsch*	*sch*	*unsch*
accounting	*pos*	4.1%	0.5%	1.2%	0.1%
	neg	2.4%	0.2%	0.6%	0.0%
	ntrl	7.9%	1.4%	2.8%	0.2%
analysis	*pos*		11.7%		9.0%
	neg		7.1%		6.8%
	ntrl		3.2%		2.4%
equity	*pos*	0.2%	17.6%	0.1%	7.8%
	neg	0.3%	17.8%	0.1%	6.9%
	ntrl	0.1%	1.1%	0.0%	3.5%
media	*pos*	0.0%	3.1%	0.0%	2.9%
	neg		2.5%		2.4%
	ntrl	1.7%	1.7%	0.8%	0.0%
operations	*pos*	2.7%	8.5%	1.6%	4.9%
	neg	0.1%	3.2%	0.1%	2.1%
	ntrl	0.8%	0.1%	0.6%	0.1%

Notes: This table reports the percentage of news days (number of days with a given group of news divided by the total number of firm/day observations) for each news group. The groups are formed based on whether the news is released during the overnight period (*ovrnt*) or trading hours (*trdng*), whether the sentiment is positive (*pos*), negative (*neg*), or neutral (*ntrl*), whether the news is scheduled (*sch*) or unscheduled (*unsch*), and which event category the news belongs to (see Table 1 for the definitions of each category). In total, there are 60 groups by which we separate news.

et al. (2009), we retrieve trades during normal market hours, between 9:30 and 16:00, and remove instances of corrected trades, abnormal sale conditions, and zero prices. Further, instead of using the data at the millisecond level, we prepare the data by aggregating prices to the second and taking the median price. Each stock has trades from more than one exchange. We choose between New York Stock Exchange (NYSE) and the National Association of Security Dealers Automated Quotation (NASDAQ) the exchange that records the most trades for each firm across our sample period.

To compute RV, we sample the prices using equidistant 5-minute intervals and the previous tick method, and then compute daily RV for firm i using Equation (11). The 5-minute RV is shown to be a good proxy in the US equity market (Liu et al., 2015). The constructed RV measures the open-to-close variance for each stock.

The average RV across all firm/day observations is 2.64, which translates to an annualized volatility of 25.8 percent. To examine the firm heterogeneity, we present histograms of the mean RV for firm i, denoted by $\overline{RV}_i$, in Figure 3(a). The 5% and 95% percentiles of $\overline{RV}_i$ are 1.41 and 5.58, respectively. In other words, the mean annualized volatility is between 18.9 and 37.5 percent for 90% of firms. In Figure 3(b), we present the histogram of the first-order autocorrelation, $Corr(RV_{it}, RV_{i,t-1})$. In 294 out of 426 firms, the first-order autocorrelation of RV exceeds 0.7, suggesting a high level of volatility persistence for most firms.

To gauge the persistence of the idiosyncratic variance, we apply the fixed effects models in Equation (13) and Equation (14) to RV_{it} without using news as covariates. The residuals from these regressions measure idiosyncratic variance subject to a constant level shift. Panel (c) and (d) in Figure 3 present the first-order autocorrelation of the idiosyncratic variance obtained using interactive fixed effects and additive fixed effects, respectively. From panel

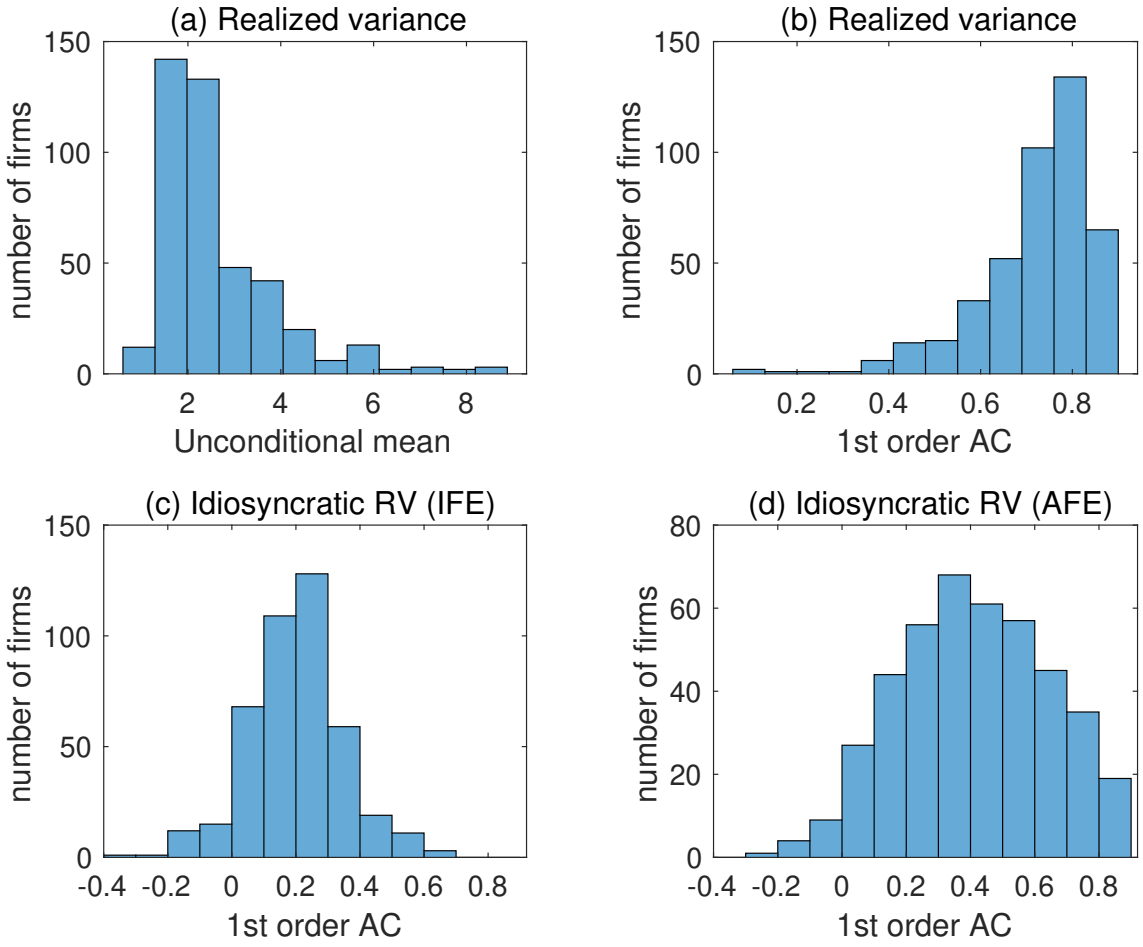

Figure 3: Descriptive statistics of realized variances and idiosyncratic RV. The idiosyncratic RV is obtained as the residuals from either the interactive fixed effects (IFE) regression $RV_{it} = \alpha_i + \lambda_i f_t + e_{it}$ or the additive fixed effects (AFE) regression $RV_{it} = \alpha_i + f_t + e_{it}$.

(c), the interactive fixed effects reduce the first-order autocorrelation substantially to around 0.2. In comparison, the first-order autocorrelation in panel (d) is centered around 0.4, with 54 firms still having autocorrelation exceeding 0.7.

In summary, the common factor, which can be interpreted as the market variance, plays an important role in the persistence of RV. Moreover, the interactive fixed effects model is better at separating the market variance from the idiosyncratic variance than the additive fixed effects model.

5 Bounding the contribution of news to return variance

In this section, we focus on quantifying the return variance attributable to news. Specifically, we define the news variance of stock i, denoted by V_i^{pub}, as the time average contribution of public news to stock i's daily variance, i.e., $V_i^{\text{pub}} \equiv \text{E}(V_{it}^{\text{pub}})$. From the modified information processing theory in Section 2, we have $V_i^{\text{pub}} = \sum_{k=1}^{K} \sum_{j=-J}^{J} \theta_{k,j} \pi_{k,i}$, where $\pi_{k,i} = \text{E}(n_{k,i,t})$ is the expected number of news from group k occurring for stock i. Hence, the news variance is determined by the effect of each type of news, which is homogeneous across stocks, and the probability of the news occurring, which is heterogeneous across stocks. The estimated value of V_i^{pub}, denoted by $\hat{V}_i^{\text{pub}}$, is computed using $\frac{1}{T} \sum_{t=J}^{T-J} \sum_{k=1}^{K} \sum_{j=-J}^{J} \hat{\theta}_{k,j} n_{k,i,t-j}$ for Regression IFE and AFE, or $\frac{1}{T} \sum_{t=J}^{T-J} \sum_{k=1}^{K} \sum_{j=0}^{J} \hat{\gamma}_{k,j} n_{k,i,t-j}$ for Regression IFE-AR and AFE-AR.

In the presence of information leakage, Proposition 2 suggests that Regression IFE and AFE yield upward-biased estimates of $\gamma_{k,j}$, which lead to the upward-biased estimates of V_i^{pub}. In the presence of tardy news, Proposition 3 suggests that Regression IFE-AR and AFE-AR yield downward-biased estimates of $\gamma_{k,j}$ and V_i^{pub}. Hence, setting sampling errors aside, estimates from the static regressions can be treated as an upper bound to the true news variance, whereas estimates from the AR regressions provide a lower bound. Moreover, the gap between the upper and the lower bound gauges the severity of information leakage or tardy news, although we can not distinguish between the two channels. If there is neither information leakage nor tardy news, both the upper and lower bounds measure the true effect of news.

Figure 4 plots the estimated news variance, $\hat{V}_i^{\text{pub}}$, using Regression IFE (blue solid line) and IFE-AR (red dashed line) and two news groups, scheduled and unscheduled. For better

visualization, we order the stocks by $\hat{V}_i^{\text{pub}}$ from Regression IFE-AR, so that the red dashed line is monotonically decreasing as i increases. The difference between $\hat{V}_i^{\text{pub}}$ estimated by the IFE and the IFE-AR is small across all stocks, suggesting that the effect of information leakage or tardy news is small compared to the overall effect of news.

We plot $\hat{V}_i^{\text{pub}}$ using Regression AFE (blue solid line) and AFE-AR (red dashed line) in Figure 5. The gap between the upward-biased estimates (AFE) and downward-biased estimates (AFE-AR) is a lot more substantial in Figure 5 compared to Figure 4. In other words, IFE and IFE-AR regressions provide tighter bounds for V_i^{pub} than the AFE and AFE-AR regressions. Recall from Figure 3 that IFE type regressions capture the market factor more effectively, thus reducing the autocorrelation in individual stock volatility compared to the additive factor structure. From Propositions 2 and 3, the severity of bias depends on the autocorrelation in V_{it}. Hence, the IFE-type regressions suffer less from bias, and they are more robust than AFE-type regressions in the presence of information leakage and/or tardy news. Therefore, in what follows, our analysis focuses on the IFE type.

6 The pre-news volatility rise

In this section, we focus on the volatility impact of news up to five days prior to and five days after the news day. Figure 6 plots the dynamic variance response θ_j, with $j = -5, \ldots, 5$, for aggregated scheduled versus aggregated unscheduled news using the IFE regression. From Figure 6, volatility experiences no significant increase before the release of scheduled news, and it stays elevated one day after the news. On the contrary, for unscheduled news, volatility starts to increase two days before the news day and returns to a normal level on the post-news

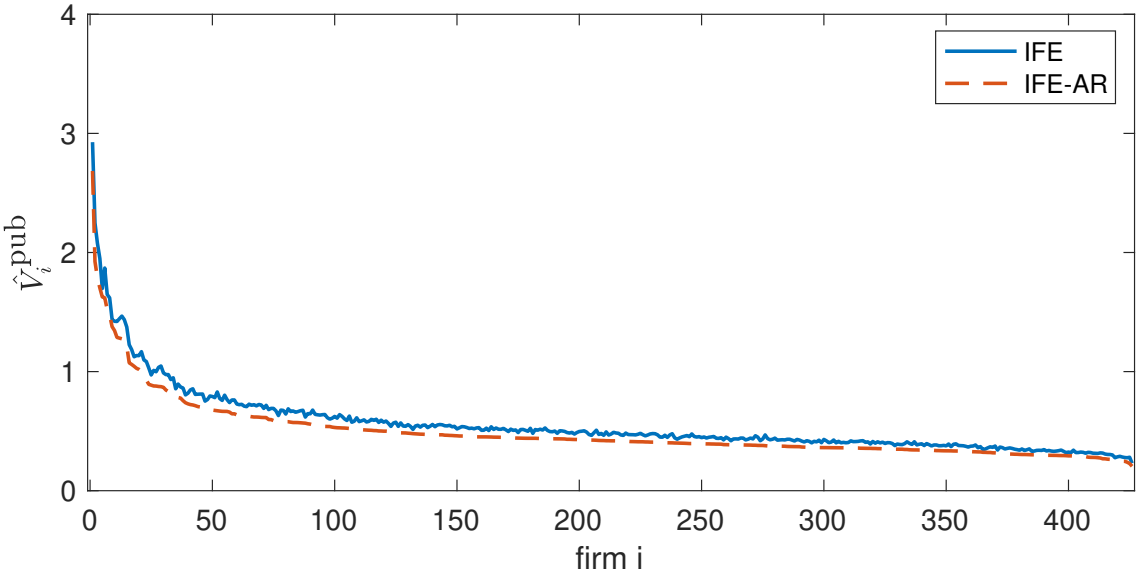

Figure 4: Average effect of news for individual stocks estimated from interactive fixed effect models. The blue solid line corresponds to Regression IFE specified in Equation (13), and the red dashed line corresponds to IFE-AR specified in Equation (15). Stocks are ordered by $\hat{V}_i^{\text{pub}}$ from Regression IFE-AR.

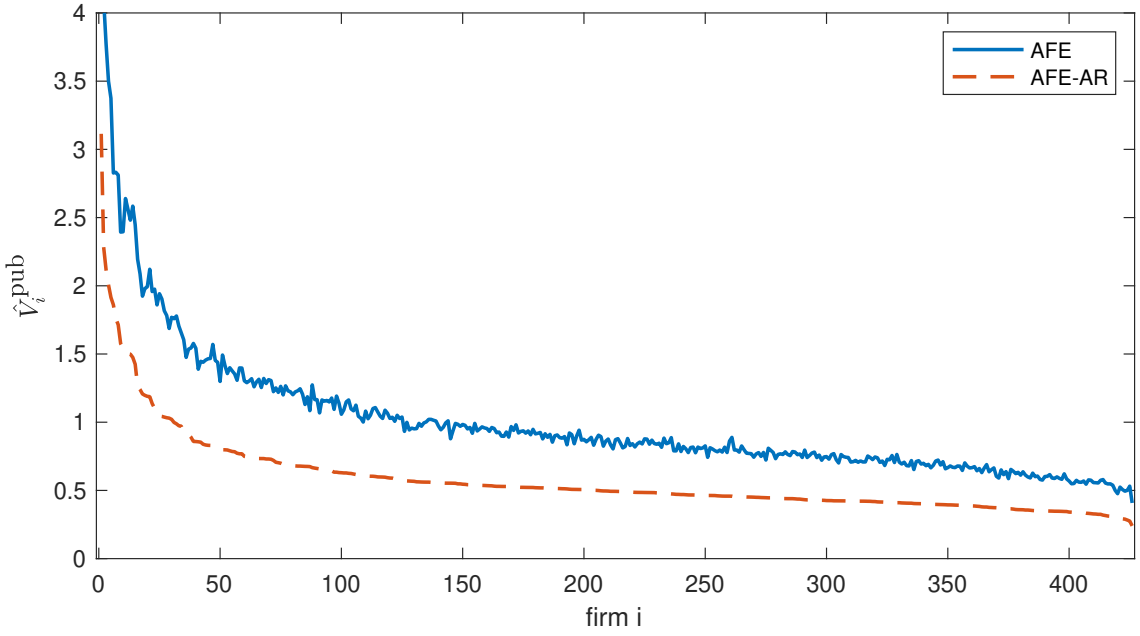

Figure 5: Average effect of news for individual stocks estimated from additive fixed effect models. The blue solid line corresponds to Regression AFE specified in Equation (14), and the red dashed line corresponds to AFE-AR specified in Equation (16). Stocks are ordered by $\hat{V}_i^{\text{pub}}$ from Regression AFE-AR.

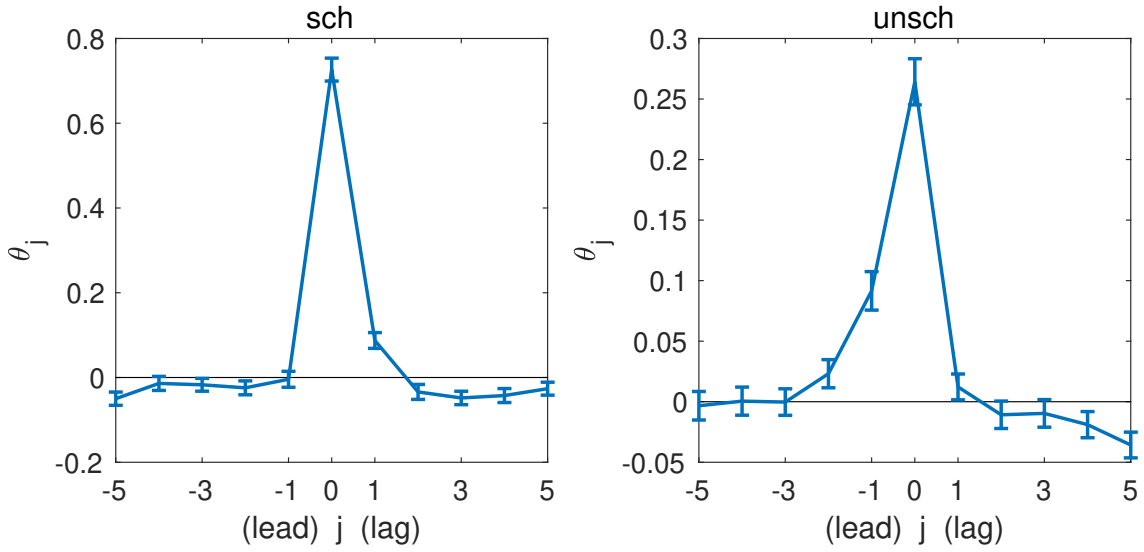

Figure 6: The dynamics of volatility responses to scheduled and unscheduled news. The coefficients are estimated using the IFE regression with $K = 2$. The bars around the point estimates indicate the 95% confidence interval.

day.

6.1 Effects of news grouping

To further investigate why volatility rises before unscheduled news, we run the IFE regression using the 52 news groups that have non-zero observations as listed in Table 2. Figure 7 presents the dynamic variance response for unscheduled news of different event types, sentiment and timing. Coefficients associated with neutral sentiments are omitted as they are mostly insignificant.

From Figure 7, the *accounting* group of news has a strong contemporaneous impact on volatility, but it does not have any pre-news or post-news volatility impact. This group comprises news about earnings, revenues, and dividends, which are pertinent to the fundamental value of a firm. The significance and size of the contemporaneous impact from our estimates align with Engle et al. (2021), who find that news with economic information

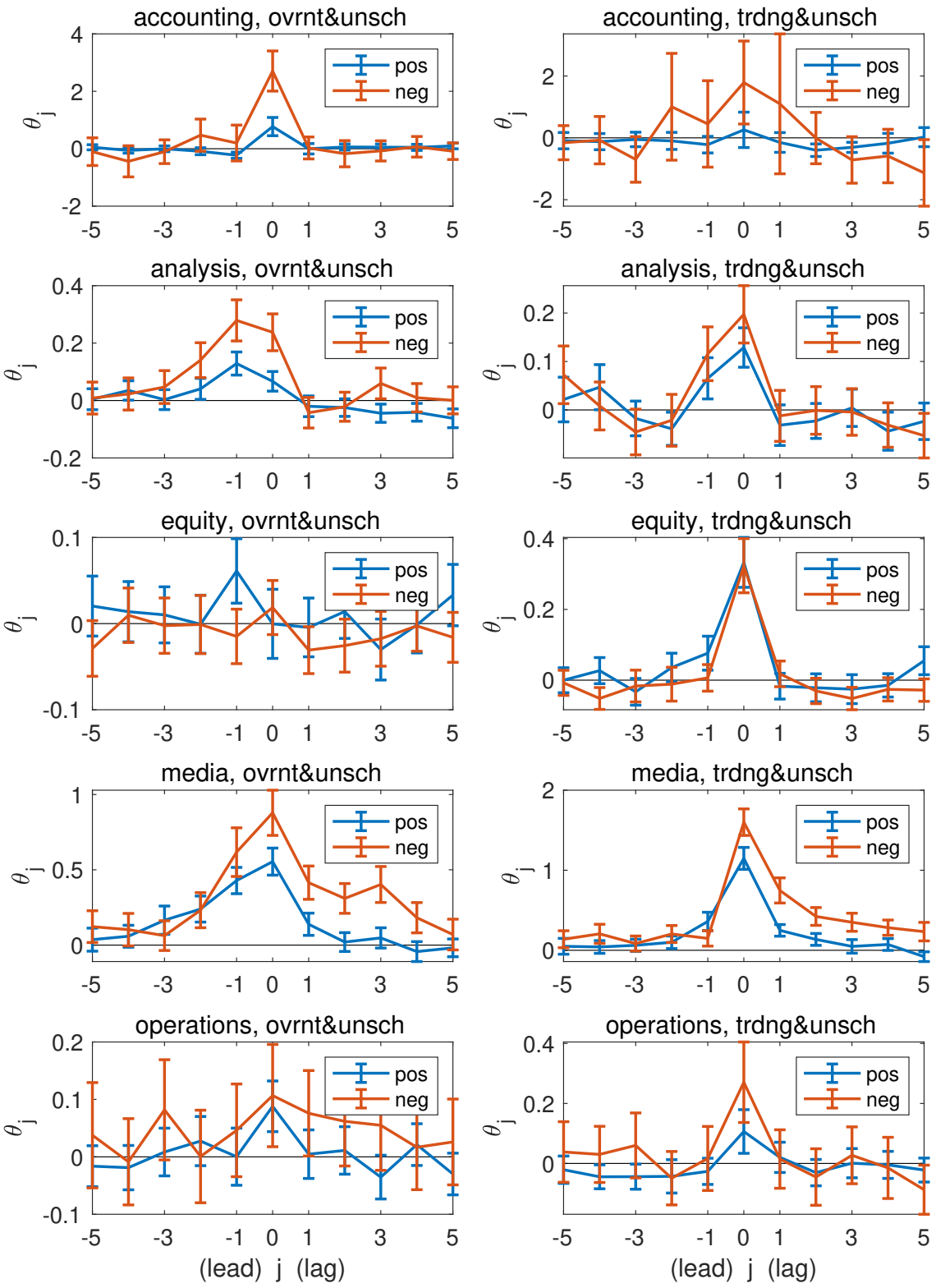

Figure 7: The dynamics of volatility responses to unscheduled news. The coefficients are estimated using the IFE regression with 52 news groups listed in Table 2. The bars around the point estimates indicate the 95% confidence interval.

explains a significant proportion of the changes in the firm-specific variance, and Dubinsky et al. (2019), who find that earning announcements are accompanied by quantitatively large volatility. The lack of pre-news or post-news volatility impact suggests that the market is efficient at processing this type of information. This supports Timmermann et al. (2023) who find that the market absorbs information in earning announcements at warp speed. It does not support, however, the model implications of Hirshleifer et al. (2011) where investors choose not to pay attention to accounting news.

The *analysis* group news, which consists of stock recommendations by financial analysts and institutions, exhibits strong pre-news and contemporaneous volatility rises. The overnight (trading hour) *analysis* news is associated with significant volatility rises up to two days (one day) prior to and on the day of publishing, and negative news has a larger volatility effect than positive news.

The volatility dynamic associated with the *analysis* group of news is consistent with the tardy news hypothesis. Assuming that financial analysts and institutions are not privy to private information, *analysis* news, such as a buy recommendation, is based on some public information. Full attention traders act on this public information immediately. For limited attention traders, analysts have expertise in distilling public information, and *analysis* news provides a less costly channel to acquire information. Furthermore, financial analysts face attention constraints themselves (e.g. Choi and Gupta-Mukherjee, 2022), leading to a delay between the time of the public information and the time of a recommendation. As a result, the trading activities of full attention traders occur before the release of the *analysis* news, whereas limited attention traders act after the news, causing contemporaneous and post-news volatility rises.

Consider the case of Xilinx outlined in the introduction as an example. When Amazon announced its collaboration with Xilinx at their annual conference and blog, full attention traders perceived this as a positive signal to Xilinx's value and bought Xilinx immediately. The limited attention traders, however, would only react to this information on the following day, when an analyst recommended "buy" for Xilinx due to its collaboration with Amazon and this recommendation appeared on a financial news website. Hence, we observe volatility rise both prior to and on the day of the analysis news.

It is worth noting that the pre-news rise is also consistent with information leakage as in Irvine et al. (2007), who find abnormal trading activity by institutional investors prior to the release of analysts' recommendations. Yet, the results in Irvine et al. (2007) could also be explained by the tardy news hypothesis: institutional investors are more likely to be full-attention traders who acquire information simultaneously as analysts and act before the release of analysts recommendations. Individual investors, on the other hand, are more likely to be limited attention traders who follow analysts' recommendations. The tardy news hypothesis also supports the findings in Huang et al. (2020), who find that institutional investors do not predicatively trade on the news but are fast at processing information when it first becomes available.

The groups of unscheduled *media* news experience both pre-news and post-news volatility rises. The unscheduled *media* news reports recent price changes, reflecting information that is already incorporated in the market. A news item that is only about price changes can serve as a proxy for investor sentiment, defined in Tetlock (2007) as the difference in the beliefs about future dividends between noise traders who hold random beliefs and rational arbitrageurs who hold Bayesian beliefs. Our results support the conclusion in Tetlock (2007),

who finds that media pessimism predicts pessimistic investor sentiment. We differ from Tetlock (2007) in that we use the type of news to identify investor sentiment, whereas Tetlock (2007) uses patterns of return reversals to identify investor sentiment. Specifically, our results suggest that news about the stock price going up (down) a large percentage induces optimistic (pessimistic) sentiments in noise traders, resulting in them buying from (selling to) arbitrageurs. Furthermore, while both positive and negative news affect investor sentiment, the effect of negative news is larger and more prolonged than positive news.

The investor sentiment theory only explains the rise in contemporaneous/post-news volatility, but not pre-news volatility. Furthermore, it is not uncommon for the news about large price changes to also include explanations or speculations as to why the price changes occur. For example, the headline of stock loss news about DaVita Inc., a prominent provider of kidney dialysis services, reads as follows: "DaVita shares are down 4.2% premarket after the California Assembly passed a dialysis bill that would cap payments at lower Medicare rates for certain providers." This news item was released at 8:53 am on Aug 30, 2018, and there is no news concerning DaVita on Aug 29. Yet, the volatility of DaVita increased on both Aug 29 and Aug 30. Through an internet search, we found that the dialysis bill was passed on Aug 29, explaining the volatility increase one day before the news report. On Aug 30, investors reacted further to the news about stock loss, and the volatility increased again.

The volatility rises that happen prior to unscheduled *media* news with more than just price change information are consistent with the tardy news hypothesis. In a nutshell, full attention traders react to events that are relevant to fundamental values but are not covered in news; actions of the full attention traders cause large price movement that becomes news; limited attention traders become aware of the fundamental events from the news on price

change, causing a further increase in volatility. Note that the tardy news and the investor sentiment hypotheses do not exclude each other; both could be at work in financial markets.

The unscheduled overnight *equity* news with negative or positive sentiments are the two largest groups in terms of news frequency, occurring to 17.8% or 17.6% of all firm/day observations (see Table 2). Yet from Figure 7, they are associated with insignificant contemporaneous volatility changes and a small amount of pre-news volatility rise (positive sentiment news only). The trading day *equity* news significantly impacts contemporaneous volatility and, to a small degree, the pre-news volatility. In the *operations* group, the effects of news only exist contemporaneously, and they are small in magnitude.

We consider *analysis*, *equity* and *operation* as *primary* news as they directly report events that are relevant to the fundamental values of a firm. In contrast, *analysis* and *media* are *secondary* news as they report interpretations of or responses to events that change the firm fundamentals. Our results suggest that secondary news is more likely to experience pre-news volatility rises than primary news. This differentiation is evidence favoring the tardy news hypothesis over the information leakage theory, as "insiders" are more likely to acquire private information prior to primary news, while secondary news is more likely to experience a delay in reporting.

6.2 Effects of firm characteristics

Another important testable implication of the tardy news hypothesis is that firms with more investor attention will be less likely to experience pre-news volatility rise. To test this implication, we divide the constituents of the SP500 into three groups. Group one

Table 3
Pre-news volatility changes

		ovrnt&unsch			trdng&unsch		
		large	*medium*	*small*	*large*	*medium*	*small*
analysis	*pos*	0.059*	0.114**	0.123***	0.079**	0.046	0.078***
		(0.033)	(0.046)	(0.025)	(0.039)	(0.038)	(0.027)
	neg	0.046	0.178***	0.325***	0.163***	0.132**	0.100***
		(0.037)	(0.062)	(0.046)	(0.057)	(0.066)	(0.033)
media	*pos*	0.204***	0.352***	0.486***	0.064	0.385***	0.377***
		(0.052)	(0.070)	(0.061)	(0.079)	(0.114)	(0.079)
	neg	0.309***	0.430***	0.748***	0.103	0.212***	0.187***
		(0.066)	(0.104)	(0.118)	(0.094)	(0.072)	(0.069)

Notes: This table reports the volatility effect of news one day prior to the news day ($\theta_{k,-1}$) from the IFE regression. The stocks are separated into three groups: large (DIJA constituents), medium (rest of S&P 100 constituents), and small (rest of S&P 500 constituents). The asterisks ***,**, and * denote p-values lower than 1%, 5%, and 10%, respectively.

contains the constituents of the Dow Jones Industrial Average (DJIA), group two includes the constituents of the S&P100 index that are not in the DJIA, and group three contains the rest. Group one (three) includes the largest (smallest) and most (least) established firms, hence stocks in this group are likely to attract more (less) attention from limited-attention traders.

Table 3 reports the changes in volatility one day prior to the news belonging to the *analysis* and the *media* groups with different timing and sentiment. In six out of the eight news groups, the largest stocks have the smallest pre-news volatility rise. The tardy news hypothesis predicts that the pre-news volatility increase will be inversely proportional to the size of the firm. On the other hand, the pre-news volatility increase caused by information leakage should be independent of the firm size. Hence, the results in Table 3 provide further support for the tardy news hypothesis.

7 Does it pay to be informed?

To gauge the economic value of acquiring information before news, we examine the dynamic effect of news on returns using the following IFE model:

$$r_{it} = \alpha_i + \lambda_i f_t + \sum_{k=1}^{K} \sum_{j=-J}^{J} \delta_{k,j} n_{k,i,t-j} + \epsilon_{i,t}, \tag{18}$$

where r_{it} is the continuously compounded close-to-close return for asset i on day t in excess of the risk free rate, α_i is the individual-specific access return, f_t is the latent pricing factor, λ_i is the factor loading, and $\delta_{k,j}$ measures the abnormal excess return corresponding to news type k that occurred j days before/later. The news are divided into the 52 groups listed in Table 2.

In the absence of information leakage or tardy news, we expect positive (negative) returns on the day of a fundamental news item with positive (negative) sentiment, but not before or after. From Figure 8, *accounting* news does behave as expected, further confirming the markets' efficiency in processing this type of news. For *equity* and *operation* news, we observe small deviations from zero in the days preceding news, especially for negative news.

The *media* news reports large price movements, hence it is not surprising to observe significant returns associated with the day before the overnight news. The trading day news is only associated with contemporaneously return impacts.

For analysis news, we observe consistently positive (negative) returns before positive (negative) news in figure 8, and significant return reversal after the negative news. Our results suggest that there is a critical advantage to being informed, either through private information

(the information leakage channel) or from alternative data sources (the tardy news channel). Recent decades have witnessed a fast expansion of the alternative data industry where billions are invested annually in cutting-edge datasets such as satellite imagery, social sentiment analysis, mobile app usage, and internet quality and activity (see an overview in Sun et al. (2024)). The growing interest on these innovative data sources reflects their potential to provide actionable insights and improve investor decision-making. Recognizing this shift, platforms like Bloomberg have integrated alternative data directly into their terminals, allowing professional investors to access these insights alongside traditional financial metrics (Bloomberg, 2024).

8 Conclusion

In this paper, we propose a tardy news hypothesis that explains volatility rises prior to news due to the delay in news reporting time and investor inattention. We incorporate the tardy news hypothesis and information leakage in a framework that relates private and public information to volatility. Under this framework, we show that standard regressions commonly used for estimating the effect of public news on volatility can be biased and inconsistent. We propose combining regressions with upward bias and downward bias for a more robust analysis.

Using a dataset that comprises the realized variance and the Ravenpack news analytics for S&P500 constituents, we find that public news significantly impacts volatility after accounting for possible biases. By separating news by the types of events it covers, we find that *secondary* news is more likely to experience pre-news volatility rise than *primary* news, which can be

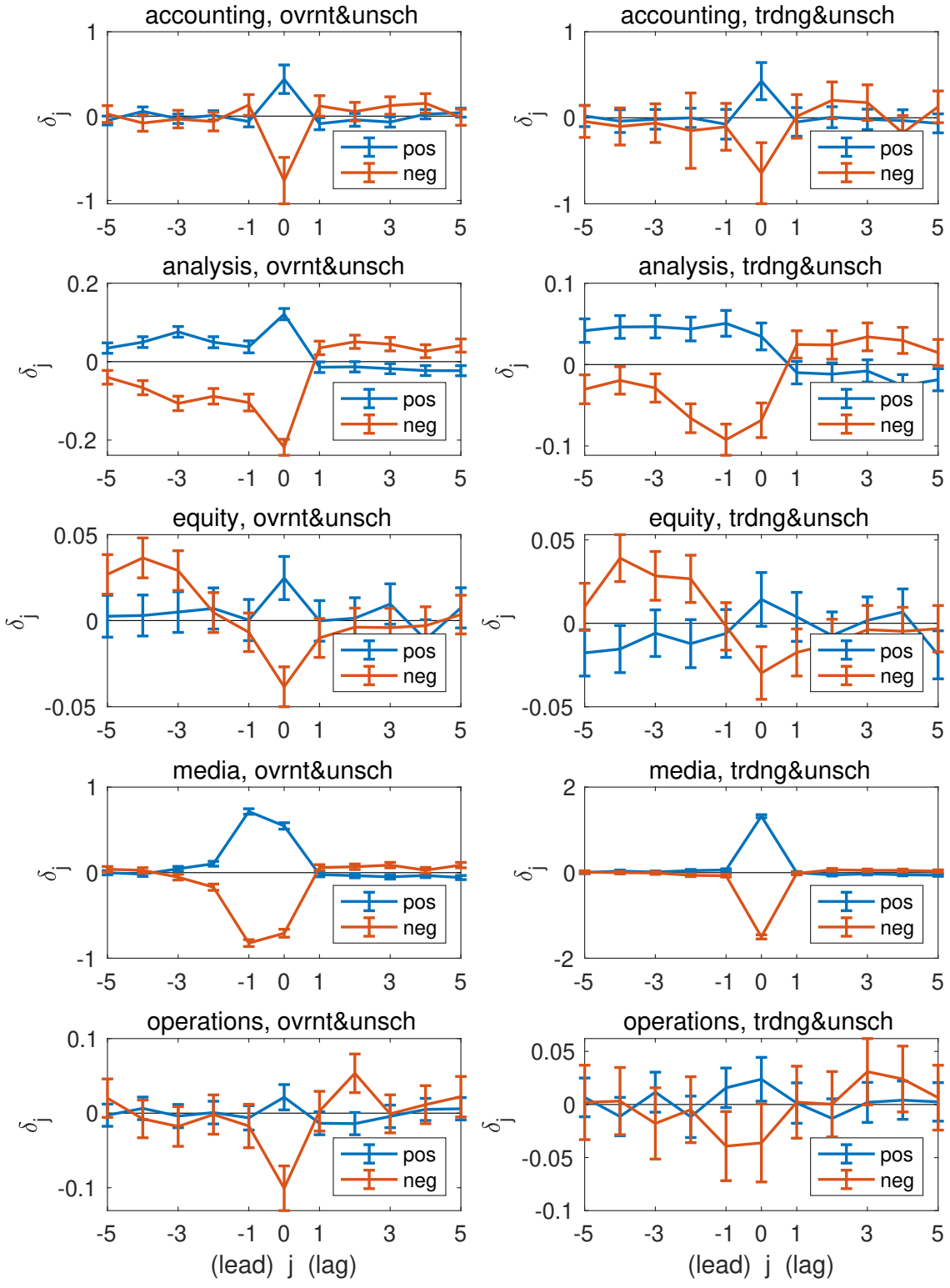

Figure 8: Return dynamics around unscheduled news. The coefficients are estimated using the IFE regression with 52 news groups listed in Table 2. The bars around the point estimates indicate the 95% confidence interval.

viewed as evidence favoring the tardy news hypothesis over information leakage. Moreover, small stocks that are more likely to suffer from a lack of attention from limited-attention traders also experience a higher magnitude of pre-news volatility rise than large stocks, which is consistent with the tardy news hypothesis.

Our empirical results suggest that alternative data is becoming an important outlet for the dissemination of public information in recent years. Hence, both theoretical and empirical research need to incorporate these emerging channels to improve our understanding of market efficiency.

Appendix A Proofs for propositions

Proof of Proposition 1. From Equations (1), (2), and (3), the idiosyncratic variance in our information processing framework takes the following form:

$$V_{it} = \alpha_i + \sum_{j=-\underline{J}}^{\overline{J}} \gamma_j n_{i,t-j} + \sum_{j=0}^{\infty} \beta_j u_{i,t-j}. \tag{19}$$

Comparing Equation (19) with Regression (4), it is easy to see that $e_{it}^{(1)} = \sum_{j=0}^{\infty} \beta_j u_{i,t-j}$. Without information leakage, $e_{it}^{(1)}$ is uncorrelated with any of the regressors, hence $\hat{\theta}_j^{(1)} \xrightarrow{p} \gamma_j$ for any j.

For Regression (5), if $\beta_j = \beta^j$, we can add and subtract $\beta V_{i,t-1}$ to the right-hand side of Equation (19) to obtain

$$V_{it} = (1-\beta)\,\alpha_i + \beta V_{i,t-1} + \gamma_{-\underline{J}} n_{t+\underline{J}} + \sum_{j=-\underline{J}+1}^{\overline{J}} (\gamma_j - \beta\gamma_{j-1})\, n_{i,t-j} - \beta\gamma_{\overline{J}} n_{i,t-\overline{J}-1} + u_{i,t} \tag{20}$$

Comparing Equation (20) with Regression (5), we have $e_{it}^{(2)} = u_{it}$ if $J \geq \max(\underline{J}, \overline{J}+1)$. Since u_{it} is uncorrelated with any of the regressors under the assumption of no information leakage, standard OLS yields $\hat{\phi} \xrightarrow{p} \beta$ and

$$\hat{\theta}_j^{(2)} \xrightarrow{p} \begin{cases} 0 & j < -\underline{J} \\ \gamma_j, & j = -\underline{J} \\ \gamma_j - \beta\gamma_{j-1}, & j = -\underline{J}+1, \ldots, \overline{J} \\ -\beta\gamma_{\overline{J}}, & j = \overline{J}+1 \\ 0. & j > \overline{J}+1 \end{cases} \tag{21}$$

Noting that $\gamma_j = 0$ for $j < -\underline{J}$ or $j > \overline{J}+1$, we can simplify (21) to $\hat{\theta}_j^{(2)} \xrightarrow{p} \gamma_j - \beta\gamma_{j-1}$ for $j = -J, \ldots, J$. For transformation (6), we have $\hat{\gamma}_{-J} = \hat{\theta}_{-J}^{(2)} \xrightarrow{p} \gamma_{-J}$ since $J \geq \underline{J}$. Using $\hat{\phi} \xrightarrow{p} \beta$, we then obtain $\hat{\gamma}_j = \hat{\theta}_j^{(2)} + \hat{\phi}\hat{\gamma}_{j-1} \xrightarrow{p} \gamma_j$, for $j = -J+1, \ldots, J$.

If V_{it}^{priv} follows an AR(p) process, we can rearrange Equation (19) by adding and subtracting $\sum_{k=1}^{p} \psi_k V_{i,t-k}$ to obtain

$$V_{it} = \left(1 - \sum_{k=1}^{p} \psi_k\right) \alpha_i + \sum_{k=1}^{p} \psi_k V_{i,t-k} + \sum_{j=-\underline{J}}^{\overline{J}+p} \left(\gamma_j - \sum_{k=1}^{p} \psi_k \gamma_{j-k}\right) n_{i,t-j} + u_{i,t}. \tag{22}$$

When $J \geq \max(\underline{J}, \overline{J}+p)$, $e_{it}^{(2)}$ in Regression (7) equals u_{it}. Hence, $\hat{\phi}_k \xrightarrow{p} \psi_k$ for $k = 1, \ldots, p$, and $\hat{\theta}_j^{(2)} \xrightarrow{p} \gamma_j - \sum_{k=1}^{p} \psi_k \gamma_{j-k}$ for $j = -J, \ldots, J$. Then, transformations similar to (6) can be applied to obtain the estimator $\hat{\gamma}_j$ that converges to γ_j for $j = -J, \ldots, J$. □

Proof of proposition 2. For Regression (4), given that $\mathsf{cov}(n_{i,t-j}, n_{i,t-j^*}) = 0$ for $j \neq j^*$,

standard OLS yields

$$\theta_j^{(1)} \xrightarrow{p} \frac{\mathsf{cov}(V_{it}, n_{i,t-j})}{\mathsf{var}(n_{i,t-j})} = \gamma_j + \sigma_{n,i}^{-2}\mathsf{cov}\left(e_{it}^{(1)}, n_{i,t}\right),$$

where $\sigma_{n,i}^2 = \mathsf{var}(n_{i,t}) = \mathsf{var}(n_{i,t-j})$ due to stationarity. Since $e_{it}^{(1)} = \sum_{j=0}^{\infty}\beta_j u_{i,t-j}$ and assuming $\beta_j = 0$ for $j < 0$,

$$\mathsf{cov}\left(e_{it}^{(1)}, n_{i,t}\right) = \mathsf{cov}\left(\sum_{j=0}^{\infty}\beta_j u_{i,t-j}, n_{i,t}\right) = \sum_{k=1}^{J^*}\beta_{j+\underline{J}+k}\rho_k, \tag{23}$$

where $\sigma_{n,i}^2 = \mathsf{var}(n_{i,t})$ and $\rho_k = \mathsf{cov}(u_{i,t-\underline{J}-k}, n_{i,t})$ for $k = 1, \ldots, J^*$. □

proof of Proposition 3. Comparing Equation (9) with (20), we have

$$e_{it}^{(3)} = \gamma_{-\underline{J}} n_{t+\underline{J}} + \sum_{j=-\underline{J}+1}^{-1}\left(\gamma_j - \beta\gamma_{j-1}\right) n_{i,t-j} + u_{i,t},$$

for $\underline{J} \geq 1$. Information leakage allows $u_{i,t}$ to be correlated with $n_{i,t+j}$ when $j = \underline{J}+1, \ldots, \underline{J}+J^*$, but we still have $\mathsf{cov}(u_{i,t}, n_{i,t+j}) = 0$ when $j \leq \underline{J}$. Since we also assume $\mathsf{cov}(n_{i,t-j}, n_{i,t-j^*}) = 0$ for $j \neq j^*$, $e_{it}^{(3)}$ is not correlated with any of the regressors. Standard OLS yields $\hat{\phi} \xrightarrow{p} \beta$ and

$$\hat{\theta}_j^{(3)} \xrightarrow{p} \begin{cases} \gamma_j - \beta\gamma_{j-1}, & j = 0, \ldots, \overline{J} \\ -\beta\gamma_{\overline{J}}, & j = \overline{J}+1 \\ 0. & j = \overline{J}+2, \ldots J \end{cases} \tag{24}$$

Given that $\gamma_j = 0$ for $j \geq \overline{J}+1$, we can simplify (24) to $\hat{\theta}_j^{(3)} \xrightarrow{p} \gamma_j - \beta\gamma_{j-1}$ for $j \geq 0$.

For transformation (10), we start with $\hat{\gamma}_0 = \hat{\theta}_0^{(3)} \xrightarrow{p} \gamma_0 - \beta\gamma_{-1}$, then using $\hat{\phi} \xrightarrow{p} \beta$,

$$\begin{aligned}
\hat{\gamma}_j &= \hat{\theta}_j^{(3)} + \hat{\phi}\hat{\gamma}_{j-1} \\
&= \sum_{k=0}^{j} \hat{\phi}^k \hat{\theta}_{j-k}^{(3)} \\
&\xrightarrow{p} \gamma_j - \beta^{j+1}\gamma_{-1},
\end{aligned}$$

for $j > 0$. □

Appendix B Ravenpack event categories and providers

Table 4
News count and descriptions of Ravenpack event categories

Ravenpack Event	News count	Typical Events
equity-actions	235277	Shareholders change positions, IPOs
acquisitions-mergers	181188	Acquisitions, mergers, and shareholders changing positions
earnings	149285	Earnings preview, guidance, and reports
insider-trading	128752	Executives or large sharesholders changing their holdings
analyst-ratings	127554	Financial analysts and institutions give buy/hold/sell ratings
technical-analysis	102769	Financial analysts and institutions give price predictions
price-targets	81815	Financial analysts and institutions set price targets
stock-prices	71709	Actual price changes
products-services	68845	Changes in product/service such as new releases or discontinuotions
revenues	51352	Revenues reports
labor-issues	42992	Changes in personnel, such as CEOs or directors
partnerships	28302	Changes in collaboration with other companies
investor-relations	27642	Press-releases, earnings conference calls
marketing	22889	Participating in conferences
legal	18652	Lawsuits and settlements
assets	18324	Sale/purchase of facilities, patent filings
dividends	15366	Dividend distribution
credit-ratings	10473	Financial institutions rate corporate bonds
regulatory	3097	Investigations from government bodies
credit	3053	Bond offering

Notes: This table reports the number of news items in the 20 most common Ravenpack event categories and summarizes the typical events corresponding to each category.

References

Aït-Sahalia, Y., Li, C. X., and Li, C. (2024). So many jumps, so few news. Technical report, National Bureau of Economic Research.

Table 5
Top 10 news providers in Ravenpack

Provider Name	News Count	News Percentage
TICKER REPORT	335704	24.0%
DOW JONES NEWSWIRES	303057	21.7%
LEDGER GAZETTE	141768	10.2%
SLEEKMONEY	66327	4.7%
INDIVIDUAL.COM	50082	3.6%
MT NEWSWIRES	44098	3.2%
BENZINGA	43767	3.1%
THE FLY	43738	3.1%
PR NEWSWIRE	24156	1.7%
SEEKING ALPHA	22533	1.6%

Notes: This table reports the number and percentage of news items from the top 10 Ravenpack news providers.

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

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, 96(453):42–55.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, 77(4):1229–1279.

Bajgrowicz, P., Scaillet, O., and Treccani, A. (2016). Jumps in high-frequency data: Spurious detections, dynamics, and news. *Management Science*, 62(8):2198–2217.

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

Barndorff-Nielsen, O. E. and Shephard, N. (2002). Econometric analysis of realized volatility

and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 64(2):253–280.

Bloomberg (2024). Bloomberg makes alternative data accessible alongside traditional financial data. Accessed: 2024-12-06.

Boudoukh, J., Feldman, R., Kogan, S., and Richardson, M. (2019). Information, trading, and volatility: Evidence from firm-specific news. *Review of Financial Studies*, 32(3):992–1033.

Brunnermeier, M. K. (2005). Information leakage and market efficiency. *The Review of Financial Studies*, 18(2):417–457.

Choi, H. M. and Gupta-Mukherjee, S. (2022). Analysts' Reliance on Industry-level versus Firm-specific Information: Implications for Information Production. *Journal of Banking & Finance*, 143:106555.

Christensen, K., Oomen, R. C., and Podolskij, M. (2014). Fact or friction: Jumps at ultra high frequency. *Journal of Financial Economics*, 114(3):576–599.

Christophe, S. E., Ferri, M. G., and Hsieh, J. (2010). Informed trading before analyst downgrades: Evidence from short sellers. *Journal of Financial Economics*, 95(1):85–106.

Clark, P. K. (1973). A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices. *Econometrica*, 41(1):135–155.

Connor, G. and Korajczyk, R. A. (1988). Risk and return in an equilibrium apt: Application of a new test methodology. *Journal of financial economics*, 21(2):255–289.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196.

Dubinsky, A., Johannes, M., Kaeck, A., and Seeger, N. J. (2019). Option pricing of earnings announcement risks. *The Review of Financial Studies*, 32(2):646–687.

Engle, R. F., Hansen, M. K., Karagozoglu, A. K., and Lunde, A. (2021). News and Idiosyncratic Volatility: The Public Information Processing Hypothesis. *Journal of Financial Econometrics*, 19(1):1–38.

Evans, K. P. (2011). Intraday jumps and US macroeconomic news announcements. *Journal of Banking and Finance*, 35(10):2511–2527.

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

Groß-Klußmann, A. and Hautsch, N. (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.

Hendershott, T., Livdan, D., and Schürhoff, N. (2015). Are institutions informed about news? *Journal of financial economics*, 117(2):249–287.

Hirshleifer, D., Lim, S. S., and Teoh, S. H. (2011). Limited investor attention and stock market misreactions to accounting information. *The Review of Asset Pricing Studies*, 1(1):35–73.

Huang, A. G., Tan, H., and Wermers, R. (2020). Institutional trading around corporate news: Evidence from textual analysis. *The Review of Financial Studies*, 33(10):4627–4675.

Indjejikian, R., Lu, H., and Yang, L. (2014). Rational information leakage. *Management Science*, 60(11):2762–2775.

Irvine, P., Lipson, M., and Puckett, A. (2007). Tipping. *The Review of Financial Studies*, 20(3):741–768.

Jeon, Y., McCurdy, T. H., and Zhao, X. (2022). News as sources of jumps in stock returns: Evidence from 21 million news articles for 9000 companies. *Journal of Financial Economics*, 145(2, Part A):1–17.

Jiang, G. J., Konstantinidi, E., and Skiadopoulos, G. (2012). Volatility spillovers and the effect of news announcements. *Journal of Banking Finance*, 36(8):2260–2273.

Kalev, P. S. and Duong, H. N. (2011). Firm-specific news arrival and the volatility of intraday stock index and futures returns. In Mitra, G. and Mitra, L., editors, *The Handbook of News Analytics in Finance*, chapter 12, pages 271–288. John Wiley & Sons, Ltd, Chichester, West Sussex, UK, first edition.

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

Kerssenfischer, M. and Schmeling, M. (2024). What moves markets? *Journal of Monetary Economics*, page 103560.

Liu, L. Y., Patton, A. J., and Sheppard, K. (2015). Does anything beat 5-minute RV? A

comparison of realized measures across multiple asset classes. *Journal of Econometrics*, 187(1):293–311.

Maćkowiak, B., Matějka, F., and Wiederholt, M. (2023). Rational inattention: A review. *Journal of Economic Literature*, 61(1):226–273.

Park, Y.-H. (2022). Informed trading in foreign exchange futures: Payroll news timing. *Journal of Banking Finance*, 135:106372.

Ross, S. A. (1989). Information and volatility: The no-arbitrage martingale approach to timing and resolution irrelevancy. *The Journal of Finance*, 44(1):1–17.

Sims, C. A. (2003). Implications of rational inattention. *Journal of monetary Economics*, 50(3):665–690.

Sun, Y., Liu, L., Xu, Y., Zeng, X., Shi, Y., Hu, H., Jiang, J., and Abraham, A. (2024). Alternative data in finance and business: emerging applications and theory analysis. *Financial Innovation*, 10(1):127.

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

Tetlock, P. C. (2010). Does public financial news resolve asymmetric information? *The Review of Financial Studies*, 23(9):3520–3557.

Tetlock, P. C. (2011). All the news that's fit to reprint: Do investors react to stale information? *The Review of Financial Studies*, 24(5):1481–1512.

Tetlock, P. C., Saar-Tsechansky, M., and Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *The journal of finance*, 63(3):1437–1467.

Timmermann, A., Christensen, K., and Veliyev, B. (2023). Warp speed price moves: Jumps after earnings announcements. CEPR Discussion Papers 18032, C.E.P.R. Discussion Papers.

Van Nieuwerburgh, S. and Veldkamp, L. (2010). Information acquisition and under-diversification. *The Review of Economic Studies*, 77(2):779–805.