

Media abnormal tone, earnings announcements, and the stock market

INTERNET APPENDIX

D. Ardia, K. Bluteau, K. Boudt

G-2021-50

September 2021

La collection *Les Cahiers du GERAD* est constituée des travaux de recherche menés par nos membres. La plupart de ces documents de travail a été soumis à des revues avec comité de révision. Lorsqu'un document est accepté et publié, le pdf original est retiré si c'est nécessaire et un lien vers l'article publié est ajouté.

Citation suggérée : D. Ardia, K. Bluteau, K. Boudt (Septembre 2021). Media abnormal tone, earnings announcements, and the stock market, Rapport technique, Les Cahiers du GERAD G- 2021-50, GERAD, HEC Montréal, Canada.

Avant de citer ce rapport technique, veuillez visiter notre site Web (<https://www.gerad.ca/fr/papers/G-2021-50>) afin de mettre à jour vos données de référence, s'il a été publié dans une revue scientifique.

The series *Les Cahiers du GERAD* consists of working papers carried out by our members. Most of these pre-prints have been submitted to peer-reviewed journals. When accepted and published, if necessary, the original pdf is removed and a link to the published article is added.

Suggested citation: D. Ardia, K. Bluteau, K. Boudt (September 2021). Media abnormal tone, earnings announcements, and the stock market, Technical report, Les Cahiers du GERAD G-2021-50, GERAD, HEC Montréal, Canada.

Before citing this technical report, please visit our website (<https://www.gerad.ca/en/papers/G-2021-50>) to update your reference data, if it has been published in a scientific journal.

La publication de ces rapports de recherche est rendue possible grâce au soutien de HEC Montréal, Polytechnique Montréal, Université McGill, Université du Québec à Montréal, ainsi que du Fonds de recherche du Québec – Nature et technologies.

Dépôt légal – Bibliothèque et Archives nationales du Québec, 2021
– Bibliothèque et Archives Canada, 2021

The publication of these research reports is made possible thanks to the support of HEC Montréal, Polytechnique Montréal, McGill University, Université du Québec à Montréal, as well as the Fonds de recherche du Québec – Nature et technologies.

Legal deposit – Bibliothèque et Archives nationales du Québec, 2021
– Library and Archives Canada, 2021

Media abnormal tone, earnings announcements, and the stock market

INTERNET APPENDIX

David Ardia ^a

Keven Bluteau ^b

Kris Boudt ^{c, d, e}

^a GERAD & Department of Decision Sciences, HEC Montréal, Montréal (Qc), Canada, H3T 2A7

^b Department of Finance, Université de Sherbrooke, Sherbrooke (Qc), Canada, J1K 2R1

^c Department of Economics, Ghent University, 9000 Ghent, Belgium

^d Solvay Business School, Vrije Universiteit Brussel, 1050 Brussels, Belgium

^e School of Business and Economics, Vrije Universiteit Amsterdam, 1081 HV Amsterdam, The Netherlands

david.ardia@hec.ca

keven.bluteau@usherbrooke.ca

kris.boudt@ugent.be

September 2021
Les Cahiers du GERAD
G–2021–50

Copyright © 2021 GERAD, Ardia, Bluteau, Boudt

Les textes publiés dans la série des rapports de recherche *Les Cahiers du GERAD* n'engagent que la responsabilité de leurs auteurs. Les auteurs conservent leur droit d'auteur et leurs droits moraux sur leurs publications et les utilisateurs s'engagent à reconnaître et respecter les exigences légales associées à ces droits. Ainsi, les utilisateurs:

- Peuvent télécharger et imprimer une copie de toute publication du portail public aux fins d'étude ou de recherche privée;
- Ne peuvent pas distribuer le matériel ou l'utiliser pour une activité à but lucratif ou pour un gain commercial;
- Peuvent distribuer gratuitement l'URL identifiant la publication.

Si vous pensez que ce document enfreint le droit d'auteur, contactez-nous en fournissant des détails. Nous supprimerons immédiatement l'accès au travail et enquêterons sur votre demande.

The authors are exclusively responsible for the content of their research papers published in the series *Les Cahiers du GERAD*. Copyright and moral rights for the publications are retained by the authors and the users must commit themselves to recognize and abide the legal requirements associated with these rights. Thus, users:

- May download and print one copy of any publication from the public portal for the purpose of private study or research;
- May not further distribute the material or use it for any profit-making activity or commercial gain;
- May freely distribute the URL identifying the publication.

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Abstract : In Section A, we describe the data-driven methodology for the calibration of the lexicon. In Section B, we summarize the various software used in our analyses. In Section C, we validate that our corpus contains varied and trustworthy sources. In Section D, we evaluate the use of the GWP model trained on news articles to compute the tone for the earnings press releases and earnings calls. In Section E, we report the most influential words for the tone in news articles, earnings press releases, and earnings calls.

Keywords: Abnormal returns, abnormal tone, earnings announcements, event study, news media, sentometrics

A Data-driven lexicon calibration

To estimate the score of sentiment words, we assume that the variable of interest with regards to a firm k at time t has an observable quantitative proxy.¹ In our case, this variable is the stock return of the firm, $r_{k,t}$. Then, we estimate the score of the sentiment words using the following linear regression:

$$r_{k,t} = \gamma + \sum_{j=1}^J \lambda_j f(j, k, t) + \epsilon_{k,t}, \quad (\text{A1})$$

where γ is a constant, λ_j is the regression coefficient for the word j , and $\epsilon_{k,t}$ is an error term.² The parameters γ and λ_j , $j = 1, \dots, J$, are estimated by ordinary least squares. The parameters can also be estimated using machine learning methods such as the Lasso (see, Pröllochs, Feuerriegel, and Neumann 2015).

Following Jegadeesh and Wu (2013), we note that $\lambda_j = \zeta_j \lambda$, where λ is the regression coefficient of the overall tone measure. Thus, we derive the generalized word power (GWP) score for sentiment word j as:

$$\hat{\zeta}_j = \frac{\hat{\lambda}_j - \hat{\mu}_\lambda}{\hat{\sigma}_\lambda}, \quad (\text{A2})$$

where $\hat{\lambda}_j$ is the estimated coefficient for word j , and $\hat{\mu}_\lambda$ and $\hat{\sigma}_\lambda$ are the sample mean and standard deviation of $\hat{\lambda} = (\hat{\lambda}_1, \dots, \hat{\lambda}_J)'$, respectively. Note that the procedure does not only eliminate the regression coefficient λ from the scores but also normalizes the scores.

We follow Manela and Moreira (2017) and compute the importance of a sentiment word v as:

$$imp_v = \frac{\hat{\lambda}_v^2 \sigma_v^2}{\sum_{j=1}^J \hat{\lambda}_j^2 \sigma_j^2}, \quad (\text{A3})$$

where $\hat{\lambda}_v$ is the regression coefficient of sentiment word v and σ_v^2 is the sample variance of $f(v, k, t)$ across all values of k and t .

This calibration is valid for various definitions of $f(j, k, t)$. For instance, one can account for the context in which the sentiment word is situated within a text (e.g. valence shifting words such as negator), word rarity (e.g. term-frequency inverse-document-frequency), or time-distributed frequencies to take into account lagged effect of words on the variable of interest. The standard word power approach of Jegadeesh and Wu (2013) is a special case where $D_{k,t} = 1$ for all values of k and t .

B Computation details

Results of our study are obtained using R 3.6.1 (R Core Team, 2019). For general data manipulation, we use the package **data.table** (Dowle and Srinivasan, 2019). For textual data processing, we use the packages **quanteda** (Benoit et al., 2018), **sentometrics** (Ardia et al., 2021), **stringr** (Wickham, 2019), and **tm** (Feinerer et al., 2008). For near-duplicate detection using minhash and locality-sensitive hashing, we use **textreuse** (Mullen, 2016). We use **textcat** (Hornik et al., 2013) for news language detection to filter out non-English articles miss-classified by LexisNexis. We use **GWP** (Bluteau, 2019) for the calibration of the GWP, which itself is built on **Matrix** (Bates and Maechler, 2019) and

¹In the case where no observable proxy of the variable of interest is available, one can manually build a lexicon with the corresponding sentiment words score for their specific research question as done in Henry (2008), Loughran and McDonald (2011), and Renault (2017).

²An alternative is to model the heterogeneity in the word scores across, for example, different types of firms. The largest gains are to be expected in the case of low correlations between word scores across the heterogeneous groups and for applications that focus on the firm-specific tone. Since our analysis is about aggregated tone and our initial analysis shows a high average correlation between the word scores when we segment them by groups, we decided to favor the robustness of the single score per word approach.

glmnet (Friedman et al., 2010) for efficient sparse linear regression in the computation of the scores. We estimate the structural topic model with the package **stm** (Roberts et al., 2019). We estimate the panel regressions with **plm** (Croissant and Millo, 2008) and the standard error are computed according to the unifying approach presented in Millo (2017). Graphics are generated using **ggplot2** (Wickham, 2016) and **latex2exp** (Meschiari, 2015).

C Corpus analysis

We look at the publisher of the majority of the news articles as it validates if our corpus contains varied and trustworthy sources. Table I reports the top 50 sources along with their source types, the number of publications, and a brief description of the sources. In the newswire category, the top source is “PR newswire” (258,396 articles) followed by “Business wire” (211,117 articles), in the newspaper category, the top source is “The Globe and Mail” (20,978 articles), followed by the “Investor’s Business Daily” (19,282 articles). Finally, in the web publication category, the top source is “TendersInfo – News” (42,390 articles) followed by “RTTNews” (24,295 articles).

Table I: Top 50 sources in the corpus. This table reports the source, the source type, the number of publications, and the description of the top 50 sources of news articles in our corpus. Classification of the type of source is given in the LexisNexis database. The sources’ descriptions are a summary or the full description reported on the LexisNexis website.

Source	Type	#	Description
PR Newswire	newswire	258,396	PR Newswire plays a key role in the dissemination of time-critical financial information. The newswire delivers full-text, unedited news releases as written by the originators.
Business Wire	newswire	211,117	Business Wire transmits to the media the full text of news releases issued by corporations and other organizations. The news sources can be from the banking industry, entertainment, aviation or many other industries.
Targeted News Service	newswire	66,604	Targeted News Service is a national news and editorial services company that produces news and information for America’s newspapers, databases, and services businesses and consumers directly. Coverage areas include news on federal government and congressional activities, regulation, nation’s foundations, corporations, educational institutions, and native Americans.
M2 PressWIRE	newswire	64,870	M2 PressWIRE is the world’s third-largest electronic press release distribution service and UK/Europe’s largest. M2 PressWIRE’s breadth and depth of coverage draws acclaim and high degrees of external use throughout the world.
ENP Newswire	newswire	51,710	ENP Newswire is a global press release distribution service that allows companies to distribute their news both via the world’s largest news agencies and directly to individual journalists.
Marketwire	newswire	51,698	Marketwire is the leading Internet-based distributor of direct company news. Marketwire distributes corporate news, including press releases, financial announcements, and other time-critical business communications.
The Associated Press State & Local Wire	newswire	50,310	The Associated Press State & Local Wire source includes news from all 50 states, drawing news stories from 143 U.S. bureaus and from Associated Press member newspapers and broadcasters. The wire provides coverage on a variety of regional topics such as information on state capitols, legislation and politics, local regional and state sports, cross-state issues, news analysis, and entertainment.
The Associated Press	newswire	46,143	The Associated Press is the the oldest and largest news service in the world. News collected by the Associated Press republished by more than 1,300 newspapers and broadcasters.
TendersInfo – News	web publication	42,390	TendersInfo – News is a comprehensive intra-daily source of business and industry news from around the globe, keeping its readers updated on joint ventures, memorandum of understanding, project launches, contract awards, budget allocations, financial status, stock updates, etc.

Continued on next page

Table I – Continued from previous page

Source	Type	#	Description
US Official News	newswire	38,484	US Official News is a comprehensive source of major happenings, developments and full-text public announcements made through press releases, statements and other documents issued by various federal and state governments. Coverage includes the latest reports on stock exchange filings, patents, financial reports, economic surveys, inflation index, banking and economic performance reviews, and sector-wise export/import trade as well as parliamentary news including congressional legislation updates, bills, laws, business regulations and local government news.
US Fed News	newswire	37,383	US Fed News is a daily publication comprising a comprehensive compilation of publicly distributed government information, department press releases, federal/appellate/district court rulings and data related to federal/state business and grant opportunities.
Market News Publishing	newswire	32,726	Market News Publishing supplies news, commentaries, analysis and related information on public companies and exchanges in Canada and the U.S.
RTTNews	web publication	24,295	RTTNews provides comprehensive corporate news coverage of companies ranging from Blue Chips to Penny Stocks. In addition to reporting financial figures, RTT also provides comments from prominent world leaders, U.S. & International political and general news that can affect the markets, Interest Rate changes, important speeches/comments from US Federal Reserve, ECB, BoE and other central banks from around the world.
The Globe and Mail	newspaper	20,978	The Globe and Mail, and its daily Report on Business section are the definitive sources for Canadian news and business reporting.
Investor's Business Daily	newspaper	19,282	Investor's Business Daily is a national business and financial newspaper that covers all the major business and economic news of the day.
Wireless News	newswire	19,192	Wireless News delivers a daily roundup designed to keep readers on the leading edge of the wireless industry. Special focus includes emerging technologies, convergence, international news, regulations and policy, mergers and acquisitions, legal issues, emerging technologies, trends, and the youth sector.
States News Service	newswire	18,966	States News Service reports on events in Washington that affect programs or projects of major interest in individual cities and states. States News Service tracks federal laws that directly impact a state's key industries. States News Service works principally for newspapers, although its Washington data collection and reporting are now available to corporations, government agencies, and databanks.
The New York Times	newspaper	18,666	The New York Times bears the reputation of being the United States' unofficial newspaper of record. Comprehensive coverage of national, foreign, business and local news comes from The Times' extensive foreign news network and bureaus around the United States.
Agence France Presse	newswire	17,107	Agence France Presse is the world's oldest news agency. AFP covers Europe, Africa, Latin American, the Middle East, Asia, and the Pacific Rim.
Aggregated Regulatory News Service	newswire	16,429	The Aggregated Regulatory News Service (ARNS) provides full regulatory news relating to companies listed with the stock Exchange. ARNS provides all the information companies are required to make public according to FSA regulations. This includes results, new issues, shareholdings, board changes and any other information, which may affect the company's share price.

Continued on next page

Table I – Continued from previous page

Source	Type	#	Description
The FinancialWire	newswire	15,725	The FinancialWire news coverage extends to the 100 Most Active NASDAQ and NYSE companies, some 1000 companies in the Investrend platforms, the only daily wrap-up of the largest repository of standards-based professional independent research, forums, webcasts, official filings and events not generally included in other newswires' offerings, as well as socio-political perspectives on today's financial marketplace.
The Associated Press International	newswire	15,108	The Associated Press International Service covers breaking news from around the world. It provides international business, general and sports developments continuously, 24 hours each day, seven days a week.
eWEEK.com	web publication	14,565	eWEEK.com provides the first source for breaking news, vendor analysis and critical examination of recent deployments.
News Bites – people in business	newswire	14,427	News Bites – people in business collects announcements for stock exchanges globally concerning company executives (e.g. CEOs, CFOs, Directors & Company Secretaries). These include appointments, resignations, dismissals, addresses to shareholders, chairman's statements, directors' buying selling of shares, financial results, profit warnings, appointment anniversary and others.
The Mercury News	newspaper	14,402	The Mercury News is a general circulation daily newspaper providing local, national, and international news coverage.
Global Insight	web publication	13,512	Global Insight is a leading global provider of business-critical information and is relied on by thousands of executives in hundreds of multinational corporations, financial institutions, and governments throughout the world. Their political, economic and security risk analysis of 185 markets.
Associated Press – Financial News	newswire	13,094	Associated Press – Financial News provides detailed coverage of the top 1,000 US companies plus major international corporations. Coverage includes world market news, quarterly earnings announcements, executive changes, regulatory actions, mergers and acquisitions, and new product developments.
MailOnline	web publication	12,112	MailOnline offers the latest news, sports, entertainment, science and health stories from around the world from the Daily Mail and Mail on Sunday newspapers;
Benzinga.com	web publication	11,570	Benzinga.com is a news and analysis service that focuses on global markets providing original, accurate and timely global financial content from industry experts and experienced analysts while also covering the news of the day.
The Washington Post	newspaper	11,247	The Washington Post is one of the few U.S. newspapers with a serious interest in foreign news, deploying correspondents from its 16 foreign bureaus to produce in-depth articles from the world's hot spots.
Progressive Media – Company News	newspaper	11,082	Progressive Media Company News is a collection of all the latest news, comments and industry information. It covers the following industry: Pharmaceutical, Technology, Banking, Insurance, Food, Drinks, Automotive, Logistics, Medical Devices, Clean Technology, Energy, Retail, and Packaging.
Network World	newswire	10,284	Network World's self-described mandate is to be an advocate of the user: the communications manager. The weekly is written for the person, whatever the title, coordinating the increasingly complex world of voice and data communications for an organization.
Entertainment Close-Up	newswire	9,658	Entertainment Close-Up covers the deal-makers, companies, products, start-ups, technologies, and trends that are transforming the world of arts and entertainment. Special focus includes global partnerships, emerging technology, legal issues, the convergence of old and new media, telematics, standards, digital rights issues, and youth culture.

Continued on next page

Table I – Continued from previous page

Source	Type	#	Description
Product News	web publication	9,454	Product News is an information service designed to keep industrial professionals informed on hundreds of new products announced each week.
USA Today	newspaper	9,172	USA TODAY is the one of the US most-read newspaper with more than 6.3 million readers. USA TODAY provides coverage of issues and events from across the US and the world.
Plus Company updates	newswire	8,781	Plus Company updates is a comprehensive source of company information, financial statements, and corporate announcements, etc. It covers Pakistan, India, Bangladesh, Brazil, Singapore, Malaysia, UAE, Philippines, Taiwan, and other countries of the world.
The Australian	newspaper	8,634	Includes The Australian newspaper, The Weekend Australian newspaper, and its inserted Australian Magazine. The Australian is a national morning broadsheet newspaper which is published six days a week.
Associated Press Online	newswire	8,035	Associated Press Online is a news service tailored specifically for use in databases or similar online environments. The service is comprised of the top national, international, Washington, financial and sports news on a given day. Stories cover various topics including Politics, Business, Wall Street, Sports, Entertainment and Weather, and are transmitted 24 hours a day, seven days a week.
Professional Services Close-Up	newswire	8,001	Professional Services Close-Up covers the firms, products, services, start-ups, technologies, and trends that make up this growing market. Particular focus includes new products, legal services and issues, technology support, brand loyalty, customer satisfaction, and trends as well as emerging technologies and opportunities.
The Deal Pipeline	web publication	7,982	The Deal Pipeline is a premier news service covering the deal economy. Deal news and analysis combined with coverage of the people and the personalities behind the deals. Extensive coverage of Venture Capital, Mergers and Acquisitions, IPO's, Private Equity and Bankruptcy.
US States News	newswire	7,654	US State News is a daily publication with a broad collection of publicly distributed information and press releases by state government officials, including Governor, Secretary of the State, Lt. Governor and Attorney General.
Class Action Reporter	newswire	7,623	Class Action Reporter covers all significant class action litigation throughout the United States.
The Daily PAK banker	newspaper	7,455	The Daily PAK banker is published in Pakistan's three major cities, i.e., Karachi, Lahore, and Islamabad. The Daily PAK banker covers banking and financial sectors of Pakistan exclusively and also the major developments of the world banking and financial scenario.
St. Louis Post-Dispatch	newspaper	7,309	St. Louis Post-Dispatch is the major regional newspaper in St. Louis, Missouri, serving St. Louis City and County, St. Charles County, the Metro East and surrounding counties. It is the only daily newspaper in the city.
InfoWorld	web publication	7,293	InfoWorld is an information technology media business. Founded in 1978, it began as a monthly magazine. In 2007, it transitioned to a Web-only publication.
Australian Financial Review	newspaper	7,232	The Australian Financial Review is an Australian business and finance newspaper published by Nine six days a week. The Financial Review is informally referred to as the Fin or the Fin Review. Its satellite publications include Financial Review Smart Investor, Financial Review Asset and BRW.
CNNmoney.com	web publication	7,223	CNN Digital is the world leader in online news and information and seeks to inform, engage and empower the world. Staffed 24 hours, seven days a week by a dedicated team in CNN bureaus around the world, CNN's digital platforms deliver news from almost 4,000 journalists in every corner of the globe.

Continued on next page

Table I – Continued from previous page

Source	Type	#	Description
Contra Costa Times	newspaper	6,780	Contra Costa Times provides an emphasis on the rapidly expanding high tech field in the Contra Costa County area as new companies migrate away from Silicon Valley and San Francisco. The paper also covers several oil refineries.
AFX International ProFeed	newswire	6,628	AFX International ProFeed is a comprehensive financial newswire that monitors all the major markets around the globe and provides comprehensive news coverage of U.S., European and Asian companies, stock and currency markets, specific commodities, economic data and events that impact local, national and global economies.
Financial Post	newspaper	6,560	The Financial Post was established in 1907, and has been Canada’s only all-business daily since February 2, 1998.

D GWP model trained news articles to build earnings press releases and earnings calls control variables

As explained in Section 3.4, we use the GWP model trained on news articles to build two (out of four) control variables for the tone of earnings press releases and earnings calls. The motivation of this choice is two-fold.

First, training a GWP model on earnings press releases and earnings calls only suffers from poor reliability due to the low number of observations compared to the number of parameters to estimate. Inspection of the estimated lexicons reported in Table II confirms this in the case of earnings press releases. For instance, we have counter-intuitive scoring: “weaken” has a positive score while “leadership” and “agreement” have a negative score.

Second, when comparing the out-of-sample performance on the task of predicting firm returns, we obtain an R^2 of 0.255% with a GWP model trained on news data compared to 0.074% with a GWP model trained on earnings press releases; thus an improvement factor of 3.45. For earnings calls, the improvement factor is 3.39. The out-of-sample analysis is performed for the period 2003–2016. Before 2003, the number of earnings press releases and earnings calls is insufficient to train the two GWP models.

In addition, as a robustness check, we repeated the analysis of Section 5.2 of the paper with the additional controls obtained with GWP models trained on earnings press releases and earnings calls. We still found a positive relationship between the abnormal tone and abnormal returns; compare Table 5 of the paper and Table III of this Internet Appendix. Since we cannot train the two GWP models for 2000–2002 using an expanding window because data are insufficient, we rely on a full-sample calibration for these periods. Because of this look-ahead bias, we do not present this robustness analysis in the main paper.

E Important words in news articles, earnings press releases, and earnings calls

In addition to comparing the textual content, it is also interesting to compare the words driving the tone in news articles, earnings press releases, and earnings calls around earnings announcements. One way to achieve this is by looking at the most important positive and negative words obtained with the trained GWP over the last estimation window applied to the three types of textual documents. Importance is defined as the fraction of tone variance each word drives (see Manela and Moreira, 2017); see Section A for details regarding the computation.

Table II: Most positive and negative words obtained with GWP model trained on EPR. This table reports the 20 most important positive and negative words for earnings press releases among the 3,585 words in our lexicon. Importance is defined as the fraction of tone variance each word drives (Manela and Moreira, 2017). We use GWP scores estimated on earnings press releases for the largest estimation sample possible, namely from 2000 to 2016 (9,976 observations). See Section A for details. Note that stop words (e.g., “and”, “the”, “to”, etc.) have been removed.

Positive		Negative	
Root	Imp.	Root	Imp.
improv	5.49	leadership	2.07
encourag	3.18	challeng	1.70
bui	1.73	intellectu	1.64
achiev	1.46	slow	1.59
learn	1.45	compani	1.36
stronger	1.44	volatil	1.34
strong	1.07	short	1.26
great	1.05	lower	1.15
pleas	0.94	agreement	1.08
pro	0.85	special	1.03
decreas	0.85	weak	1.03
weaken	0.80	aggress	1.02
except	0.78	asset	0.99
innov	0.74	deterior	0.92
abl	0.73	declin	0.85
valu	0.72	pai	0.83
lead	0.70	competit	0.77
impress	0.67	writedown	0.77
despit	0.65	health	0.70
accur	0.65	save	0.70

In Table IV, we report the ten most important positive and negative root words for news articles (Panel A), earnings press releases (Panel B), and earnings calls (Panel C). We see that the most important positive (negative) words reasonably qualify as positive (negative) from a finance viewpoint for the three types of documents. Interestingly, we note that the level of importance is not spread identically within the three information sources. For news articles, the top positive word “beat” drives 18.87% of the tone’s variability compared to the top negative word “disappoint” with 6.32%. In earnings press releases, both top positive and negative words are key drivers of the tone’s variability, with importance at respectively 17.09% and 12.33%. The presence of top driving words in news articles and earnings press releases contrasts with earnings calls, where there is no major contributor of tone’s variability. Finally, while total importance is relatively balanced between positive and negative words for news articles and earnings press releases, we see that negative words drive much more the tone’s variability than positive words in earnings calls.

To summarize the context surrounding the appearance of the important words, we also report in Table IV the most frequent two-word combinations for which a root word is observed. Looking at the common top words in the three document types, we see that the context is similar for most of them. This is the case for “gain,” for instance, that is related to shares or sales. On the contrary, “beat” appears as “beating Wall Street” or “beating analysts” for news media compared to “beating earnings” or “beating estimates” for earnings press releases. For earnings calls, “beat” appears with idioms such as “beating a dead horse.” This example illustrates the limitation of the dictionary approach using unigrams. Our estimates indicate that, on aggregate, the word “beat” is a positive word. However, in a financial context, the sequence of words “beating a dead horse” typically has a negative connotation. Due to the large number of possible combinations of words and the low number of occurrences of these combinations, it is challenging to estimate the tone of word combinations using the word power approach in a reliable way.

Table III: CAR regression results with additional GWP controls. This table reports the panel regression results on the relationship between $RCAT(-1,1)$ and the CAR computed over various time spans. Two additional control variables have been added compared to Table 5 in the paper: $ERPGWP(EPR)$, the GWP calibrated on the earnings press releases, and $ECGWP(EC)$, the GWP calibrated on the earnings call transcripts. For the years 2000 to 2003, the calibrations are performed in-sample as the number of earnings press releases and the earnings calls are not sufficient to train the GWP model.

	$CAR(-1,1)$	$CAR(2,5)$	$CAR(2,20)$
$RCAT(-1,1)$	0.810*** (0.109)	-0.022 (0.056)	-0.347*** (0.120)
SUE	5.924*** (0.762)	0.132 (0.347)	-2.493*** (0.838)
$I[SUE < 0]$	-0.019*** (0.003)	0.001 (0.001)	0.004 (0.003)
$SUE \times I[SUE < 0]$	-6.511*** (0.946)	0.281 (0.703)	3.111** (1.306)
$EPRLM$	0.148 (0.108)	0.023 (0.065)	0.307* (0.159)
$EPRGWP$	-0.189 (0.257)	0.134 (0.136)	-0.157 (0.414)
$EPRGWP(EPR)$	0.224 (0.257)	-0.143 (0.129)	-0.315 (0.262)
$ECLM$	0.706*** (0.208)	0.109 (0.130)	-0.477 (0.339)
$ECGWP$	1.359** (0.563)	-0.761** (0.296)	-0.297 (0.699)
$ECGWP(EC)$	2.043*** (0.559)	-1.096*** (0.343)	-2.135** (0.895)
$CAR(-5,-2)$	0.087 (0.060)		
$CAR(-1,1)$		0.060*** (0.015)	0.195*** (0.027)
ROA	-0.012 (0.088)	0.143*** (0.051)	0.298** (0.130)
$\log(B/M)$	-0.014* (0.008)	0.016*** (0.005)	0.053*** (0.012)
$\log(M)$	0.001 (0.004)	-0.007*** (0.003)	-0.035*** (0.006)
Firm fixed effects	Yes	Yes	Yes
Year-quarter fixed effects	Yes	Yes	Yes
Observations	4,662	4,662	4,662
$R^2 (\times 100)$	12.7	3.1	7.0

References

- David Ardia, Keven Bluteau, Samuel Borms, and Kris Boudt. The R package `sentometrics` to compute, aggregate and predict with textual sentiment. *Journal of Statistical Software*, 99(2):1–40, 2021. doi: 10.18637/jss.v099.i02.
- Douglas Bates and Martin Maechler. *Matrix: Sparse and dense matrix classes and methods*, 2019. URL <https://CRAN.R-project.org/package=Matrix>.
- Kenneth Benoit, Kohei Watanabe, Haiyan Wang, Paul Nulty, Adam Obeng, Stefan Müller, and Akitaka Matsuo. `quanteda`: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*, 3(30):774, 2018. doi: 10.21105/joss.00774.
- Keven Bluteau. *Generalized Word Power Approach*, 2019. URL <https://github.com/keblu/GWP>.
- Yves Croissant and Giovanni Millo. Panel data econometrics in R: The `plm` package. *Journal of Statistical Software*, 27(2):1–43, 2008. doi: 10.18637/jss.v027.i02.
- Matt Dowle and Arun Srinivasan. `data.table`: Extension of “`data.frame`”, 2019. URL <https://CRAN.R-project.org/package=data.table>.
- Ingo Feinerer, Kurt Hornik, and David Meyer. Text mining infrastructure in R. *Journal of Statistical Software*, 25(5):1–54, 2008. doi: 10.18637/jss.v025.i05.
- Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1–22, 2010. doi: 10.18637/jss.v033.i01.

Table IV: Most important positive and negative words. This table reports the ten most important positive and negative words for news articles (Panel A), earnings press releases (Panel B), and earnings calls (Panel C) among the 3,585 words in our lexicon. “All” refers to the total importance of positive and negative words. Importance is defined as the fraction of tone variance each word drives (Manela and Moreira, 2017). We also report (up to two) two-word combinations for which the root word is observed. We limit to combinations that appear more than 1% of the time. In parenthesis, we report the fraction of times the combination is observed relative to all combinations. We use GWP scores estimated using data from 1999 to 2016 (full sample analysis) on newswire articles, for which we get 1,286 positive and 1,188 negative words. See Section A for details. Note that stop words (e.g., “and”, “the”, “to”, etc.) have been removed.

<i>Panel A. News at earnings announcements</i>					
Positive			Negative		
Root	Imp.	Two-word combination	Root	Imp.	Two-word combination
beat	18.87	beat_wall (15), beat_analysts (10)	disappoint	6.32	disappointing_earnings (7), disappointing_results (6)
gain	7.58	gain_sale (3), shares_gained (2)	drop	4.69	percent_drop (9), sales_dropped (4)
boost	1.99	helped_boost (5), boost_sales (3)	miss	4.58	missed_analysts (12), missed_wall (10)
stronger	1.85	stronger_dollar (9), stronger_sales (4)	lower	4.46	lower_sales (3))
surpass	1.51	surpassed_wall (12), results_surpassed (8)	weaker	3.40	weaker_expected (8), weaker_dollar (5)
strong	1.29	strong_sales (4), strong_growth (4)	weak	2.93	weak_economy (4), weak_demand (3)
upbeat	0.93	upbeat_earnings (18), upbeat_outlook (6)	sank	1.68	profit_sank (11), sales_sank (7)
strength	0.61	continued_strength (6), financial_strength (6)	hurt	1.62	hurt_sales (4), results_hurt (4)
rebound	0.36	economic_rebound (4), sales_rebounded (3)	shortfal	0.86	revenue_shortfall (13), sales_shortfall (9)
upturn	0.33	expecting_upturn (16), gradual_upturn (16)	declin	0.86	sales_declined (4), percent_decline (4)
All	49.10		All	50.90	
<i>Panel B. Earnings press releases</i>					
Positive			Negative		
Root	Imp.	Two-word combination	Root	Imp.	Two-word combination
gain	17.09	gain_sale (9)	lower	12.33	due_lower (7), lower_sales (3)
stood	5.11	quarter_stood (42), equivalents_stood (34)	weaker	3.96	offset_weaker (7), due_weaker (6)
strong	3.43	strong_growth (4), strong_performance (4)	disappoint	2.85	results_disappointing (12), disappointed_results (9)
beat	2.55	beat_earnings (10), beats_estimate (6)	declin	2.38	sales_declined (4), percent_decline (2)
stronger	1.97	stronger_US (6), even_stronger (4)	weak	1.82	continued_weakness (6), weak_economic (3)
upturn	0.84	upturn_business (23), inevitable_upturn (16)	shortfal	1.14	sales_shortfall (14), volume_shortfalls (12)
improv	0.80	productivity_improvements (1)	slower	0.95	slower_growth (16), due_slower (6)
surpass	0.61	sales_surpassed (6), surpassing_company (6)	slowdown	0.82	economic_slowdown (25), seasonal_slowdown (5)
strength	0.51	continued_strength (8), financial_strength (6)	neg	0.54	negatively_impacted (15), negative_impact (14)
exception	0.50	exceptionally_strong (36), exceptionally_well (20)	drop	0.53	quarter_dropped (9), percent_drop (5)
All	52.91		All	47.09	
<i>Panel C. Earnings calls</i>					
Positive			Negative		
Root	Imp.	Two-word combination	Root	Imp.	Two-word combination
gain	6.78	share_gains (7), gaining_share (3)	lower	7.93	due_lower (3), offset_lower (2)
strong	4.73	strong_growth (4), strong_performance (3)	disappoint	5.67	little_disappointed (3), disappointed_results (3)
beat	3.50	beat_dead (11), want_beat (3)	weaker	4.86	bit_weaker (6), weaker_expected (5)
stronger	2.74	much_stronger (5), even_stronger (4)	declin	2.73	sales_declined (3), revenue_declined (1)
thank	1.11	thanks_much (11)	weak	2.73	continued_weakness (3), weak_demand (2)
improv	0.98	margin_improvement (2), continue_improve (1)	slower	2.10	slower_growth (8), bit_slower (6)
boost	0.78	big_boost (3), boost_sales (3)	slowdown	1.86	economic_slowdown (6), see_slowdown (7)
upturn	0.63	see_upturn (8), next_upturn (7)	drop	1.84	big_drop (1)
strength	0.56	continued_strength (4), financial_strength (2)	shortfal	1.78	revenue_shortfall (9), sales_shortfall (6)
rebound	0.34	strong_rebound (3), nice_rebound (2)	miss	1.578	i_missed (32), i_missing (9)
All	40.79		All	59.21	

E. Henry. Are investors influenced by how earnings press releases are written? *Journal of Business Communication*, 45(4):363–407, 2008. doi: 10.1177/0021943608319388.

Kurt Hornik, Patrick Mair, Johannes Rauch, Wilhelm Geiger, Christian Buchta, and Ingo Feinerer. The textcat package for n -gram based text categorization in R. *Journal of Statistical Software*, 52(6):1–17, 2013. doi: 10.18637/jss.v052.i06.

Narasimhan Jegadeesh and Di Wu. Word power: A new approach for content analysis. *Journal of Financial Economics*, 110(3):712–729, 2013. doi: 10.1016/j.jfineco.2013.08.018.

Tim Loughran and Bill McDonald. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance*, 66(1):35–65, 2011. doi: 10.1111/j.1540-6261.2010.01625.x.

Asaf Manela and Alan Moreira. News implied volatility and disaster concerns. *Journal of Financial Economics*, 123(1):137–162, 2017. doi: 10.1016/j.jfineco.2016.01.032.

Stefano Meschiari. latex2exp: Use LaTeX expressions in plots, 2015. URL <https://CRAN.R-project.org/package=latex2exp>.

Giovanni Millo. Robust standard error estimators for panel models: A unifying approach. *Journal of Statistical Software*, 82(1):1–27, 2017. doi: 10.18637/jss.v082.i03.

- Lincoln Mullen. **textreuse**: Detect text reuse and document similarity, 2016. URL <https://CRAN.R-project.org/package=textreuse>.
- Nicolas Pröllochs, Stefan Feuerriegel, and Dirk Neumann. Generating domain-specific dictionaries using Bayesian learning. In *Twenty-Third European Conference on Information Systems*, pages 1–14, 2015. doi: 10.18151/7217444.
- R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2019. URL <https://www.R-project.org/>.
- Thomas Renault. Intraday online investor sentiment and return patterns in the US stock market. *Journal of Banking & Finance*, 84:25–40, 2017. doi: 10.1016/j.jbankfin.2017.07.002.
- Margaret E. Roberts, Brandon M. Stewart, and Dustin Tingley. stm: An R package for structural topic models. *Journal of Statistical Software*, 91(2):1–41, 2019. doi: 10.18637/jss.v091.i02.
- Hadley Wickham. ggplot2: Elegant graphics for data analysis. Springer-Verlag New York, 2016. ISBN 978-3-319-24277-4. URL <https://ggplot2.tidyverse.org>.
- Hadley Wickham. stringr: Simple, consistent wrappers for common string operations, 2019. URL <https://CRAN.R-project.org/package=stringr>.