ONLINE APPENDIX

This appendix reports (i) results for additional tests that are briefly described in the paper, (ii) the textual parsing details, and (iii) the *BERT-based* methodological implementation.

The appendix can be downloaded at https://federicosiano.com/oa/

Contents:

Online Appendix A

TABLE OA-1 Contextualized News and Future Income Numbers at Different Horizons

TABLE OA-2 Contextualized News and the Concurrent Releases of Analyst Forecasts

TABLE OA-3 Contextualized News and Substitution Effects with Prior Informational Releases

TABLE OA-4 Sensitivity Analyses – Additional Financial Statement and Textual Controls

TABLE OA-5 Sensitivity Analyses – Using Returns Volatility as an Alternative Capital Market Outcome

Online Appendix B

Textual Parsing

Online Appendix C

BERT-based Neural Language Model Implementation Details

TABLE OA-1

Contextualized News and Future Income Numbers at Different Horizons

Panel A: Short Horizons

	Dependent Variable: *Earn_Q+1*						Dependent Variable: *Earn_Q+2*					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Adjusted R^2	**33.7%**	**6.9%**	**31.7%**	**41.8%**	**39.0%**	**44.3%**	**29.8%**	**6.1%**	**24.3%**	**38.8%**	**36.9%**	**39.3%**
Within R^2	**-**	**-**	**-**	**8.3%**	**2.9%**	**11.3%**	**-**	**-**	**-**	**4.9%**	**1.9%**	**5.7%**
Predicted_Earn	Yes	No	No	Yes	No	No	Yes	No	No	Yes	No	No
Text Attributes	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes	No
Earnings	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
Fixed Effects	No	No	No	Firm	Firm	Firm	No	No	No	Firm	Firm	Firm
Observations	101,437	101,437	101,437	100,365	100,365	100,365	99,764	99,764	99,764	98,737	98,737	98,737

Panel B: Medium Horizons

	Dependent Variable: *Earn_Q+3*						Dependent Variable: *Earn_Q+4*					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Adjusted R^2	**29.2%**	**5.7%**	**23.4%**	**38.7%**	**36.6%**	**39.1%**	**28.9%**	**5.1%**	**21.7%**	**39.5%**	**37.2%**	**39.4%**
Within R^2	**-**	**-**	**-**	**4.7%**	**1.5%**	**5.3%**	**-**	**-**	**-**	**4.9%**	**1.1%**	**4.3%**
Predicted_Earn	Yes	No	No	Yes	No	No	Yes	No	No	Yes	No	No
Text Attributes	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes	No
Earnings	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
Fixed Effects	No	No	No	Firm	Firm	Firm	No	No	No	Firm	Firm	Firm
Observations	98,268	98,268	98,268	97,247	97,247	97,247	96,721	96,721	96,721	95,717	95,717	95,717

Notes: This table presents the explanatory power of *Predicted_Earn* (i.e., contextualized news proxy) for the prediction of future reported earnings at different horizons (i.e., from 1 to 4 quarters following the earnings announcement). Panel A focuses on relatively shorter horizons (i.e., 1 to 2 quarters following the disclosure), while Panel B focuses on medium forecasting horizons (i.e., 2 to 4 quarters following the disclosure). Both panels report the Adjusted R^2 and Within R^2 from OLS tests for a sample of firm-quarter observations with available disclosure and financial data (the difference between the reported number of observations and the primary sample of 104,254 observations is due to data points with missing future quarterly earnings). The dependent variable is future earnings at different dates in the future. The experimental variable is *Predicted_Earn* which represents the predicted dependent variable, modeled out-of-sample through contextualized earnings announcement texts. A "RoBERTa" neural language model is employed for modeling purposes. The model is trained each year using a 5-year rolling window (see Section 3.4 of the paper). *Earnings* include both the contemporaneous quarterly accounting earnings and the contemporaneous quarter-to-quarter change in earnings as defined within Appendix A of the paper. *Text Attributes* include *Tone*, *Fog*, *Length*, *Numbers*, and *Future*. All variables are defined within Appendix A of the paper. Continuous variables are winsorized at the 1st and 99th percentiles.

TABLE OA-2
Contextualized News and The Concurrent Release of Analyst Forecasts

Panel A: Pricing Revisions

	Dependent Variable: ***3-day CAR***					
	(1)	(2)	(3)	(4)	(5)	(6)
Text_News_(CAR)	1.747***	1.908***	1.747***	1.679***	1.995***	1.731***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
AF			-0.001**			0.002**
			(0.02)			(0.01)
Text_News_(CAR) x AF			**0.161*** **			**0.218*** **
			(0.00)			**(0.00)**
Sample	**No Forecast**	**Forecast**	**Full**	**No Forecast**	**Forecast**	**Full**
Controls	No	No	No	Yes	Yes	Yes
Adjusted R^2	**12.1%**	**15.1%**	**13.8%**	**13.2%**	**15.9%**	**14.7%**
Fixed Effects	No	No	No	No	No	No
Observations	43,979	60,275	104,254	43,979	60,275	104,254

Panel B: Trading Decisions

	Dependent Variable: ***3-day AVOL***					
	(1)	(2)	(3)	(4)	(5)	(6)
Text_News_(AVOL)	0.985***	0.997***	0.985***	0.976***	0.973***	0.975***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
AF			0.092***			0.073***
			(0.00)			(0.00)
Text_News_(AVOL) x AF			**0.011**			**0.016***
			(0.61)			**(0.08)**
Sample	**No Forecast**	**Forecast**	**Full**	**No Forecast**	**Forecast**	**Full**
Controls	No	No	No	Yes	Yes	Yes
Adjusted R^2	**10.9%**	**20.0%**	**21.0%**	**11.2%**	**20.1%**	**20.6%**
Fixed Effects	No	No	No	No	No	No
Observations	43,979	60,275	104,254	43,979	60,275	104,254

<u>Notes:</u> This table presents differences in the directional magnitude and explanatory power of the contextualized news proxies for pricing revisions and trading decisions (i.e., *Text_News_(CAR)* and *Text_News_(AVOL)*) around earnings announcement dates. Differences are assessed with respect to *AF* (i.e., a dummy variable equal to "1" if at least one analyst forecast is released within the earnings announcement event period – i.e., from day -1 to day 1 – and "0" otherwise). The sample comprises 104,254 firm-quarter observations. Panel A reports results related to returns (i.e., pricing revisions). Panel B shows evidence for trading volumes (i.e., trading decisions).

In Panel A, the dependent variable is the cumulative abnormal return for the 3-day period (i.e., from day -1 to 1) surrounding an earnings announcement event. Cumulative abnormal returns are computed as actual security returns minus the return on the overall market (including dividends). In Panel B, the dependent variable is the abnormal trading volume for the 3-day period (i.e., from day -1 to 1) surrounding an earnings announcement event. Abnormal trading volumes are computed as the difference between average share-outstanding-scaled volumes in the 3-day announcement period (i.e., from day -1 to 1) minus average share-outstanding-scaled volumes in the non-announcement period (i.e., from day -130 to day -10 and from day 10 to day 130); the difference is then divided by the standard deviation of share-outstanding-scaled volumes in the non-announcement period.

In Panel A, the experimental variable is *Text_News_(CAR)* which represents cumulative abnormal returns modeled out-of-sample through contextualized earnings announcement texts. In Panel B, the experimental variable is *Text_News_(AVOL)* which represents abnormal trading volumes modeled out-of-sample through contextualized earnings announcement texts. A "RoBERTa" neural language model with 3 epochs and a 1e-5 learning rate is employed for modeling purposes. The neural language model is trained using a random sampling protocol (i.e., 50-50% train and modeling split) stratified by quarter-year (see Section 3.4 of the paper).

Controls include (i) textual attributes, (ii) financial statement items, and (iii) other controls as defined within Appendix A of the paper. Continuous variables are winsorized at the 1^{st} and 99^{th} percentiles. Cluster-robust-to-heteroskedasticity *p*-values are reported in parentheses; standard errors are clustered by firm. ***, **, * indicate significance at <0.01, <0.05, <0.10.

TABLE OA-3
Contextualized News and Substitution Effects with Prior Informational Releases

Panel A: Pricing Revisions

	Dependent Variable: ***3-day CAR***					
	(1)	(2)	(3)	(4)	(5)	(6)
Text_News_(CAR)	0.481***	2.721***	0.481***	0.462***	2.706***	0.468***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Low_Prior_Info			0.002***			0.002***
			(0.00)			(0.00)
Text_News_(CAR) x Low_Prior_Info			**2.240*** **			**2.248*** **
			(0.00)			**(0.00)**
Sample	**High Prior Info**	**Low Prior Info**	**Full**	**High Prior Info**	**Low Prior Info**	**Full**
Controls	No	No	No	Yes	Yes	Yes
Adjusted R^2	**4.7%**	**19.9%**	**18.7%**	**5.2%**	**21.4%**	**19.6%**
Fixed Effects	No	No	No	No	No	No
Observations	52,157	52,097	104,254	52,157	52,097	104,254

Panel B: Trading Decisions

	Dependent Variable: ***3-day AVOL***					
	(1)	(2)	(3)	(4)	(5)	(6)
Text_News_(AVOL)	0.968***	1.066***	0.968***	0.930***	1.026***	0.918***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Low_Prior_Info			-0.032***			-0.065***
			(0.00)			(0.00)
Text_News_(AVOL) x Low_Prior_Info			**0.098*** **			**0.124*** **
			(0.00)			**(0.00)**
Sample	**High Prior Info**	**Low Prior Info**	**Full**	**High Prior Info**	**Low Prior Info**	**Full**
Controls	No	No	No	Yes	Yes	Yes
Adjusted R^2	**15.6%**	**24.2%**	**20.4%**	**16.0%**	**24.4%**	**20.7%**
Fixed Effects	No	No	No	No	No	No
Observations	52,157	52,097	104,254	52,157	52,097	104,254

<u>Notes:</u> This table presents differences in the directional magnitude and explanatory power of the contextualized news proxies for pricing revisions and trading decisions (i.e., *Text_News_(CAR)* and *Text_News_(AVOL)*) around earnings announcement dates. Differences are assessed with respect to *Low Prior Info* (i.e., a dummy variable equal to "1" whenever the ratio of absolute returns in the event period to the standard deviation of returns in the non-event quarterly period is above the median and "0" otherwise). The sample comprises 104,254 firm-quarter observations. Panel A reports results related to returns (i.e., pricing revisions). Panel B shows evidence for trading volumes (i.e., trading decisions).

In Panel A, the dependent variable is the cumulative abnormal return for the 3-day period (i.e., from day -1 to 1) surrounding an earnings announcement event. Cumulative abnormal returns are computed as actual security returns minus the return on the overall market (including dividends). In Panel B, the dependent variable is the abnormal trading volume for the 3-day period (i.e., from day -1 to 1) surrounding an earnings announcement event. Abnormal trading volumes are computed as the difference between average share-outstanding-scaled volumes in the 3-day announcement period (i.e., from day -1 to 1) minus average share-outstanding-scaled volumes in the non-announcement period (i.e., from day -130 to day -10 and from day 10 to day 130); the difference is then divided by the standard deviation of share-outstanding-scaled volumes in the non-announcement period.

In Panel A, the experimental variable is *Text_News_(CAR)* which represents cumulative abnormal returns modeled out-of-sample through contextualized earnings announcement texts In Panel B, the experimental variable is *Text_News_(AVOL)* which represents abnormal trading volumes modeled out-of-sample through contextualized earnings announcement texts. A "RoBERTa" neural language model with 3 epochs and a 1e-5 learning rate is employed for modeling purposes. The neural language model is trained using a random sampling protocol (i.e., 50-50% train and modeling split) stratified by quarter-year (see Section 3.4 of the paper).

Controls include (i) textual attributes, (ii) financial statement items, and (iii) other controls as defined within Appendix A of the paper. Continuous variables are winsorized at the 1st and 99th percentiles. Cluster-robust-to-heteroskedasticity *p*-values are reported in parentheses; standard errors are clustered by firm. ***, **, * indicate significance at <0.01, <0.05, <0.10.

TABLE OA-4
Sensitivity Analyses – Additional Financial Statement and Textual Controls

Panel A: Pricing Revisions

	Dependent Variable: *3-day CAR*									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Text_News_(CAR)	**1.835*****				**1.871*****	**1.960*****				**2.011*****
	(0.00)				**(0.00)**	**(0.00)**				**(0.00)**
Adjusted R^2	**13.8%**	**1.3%**	**3.7%**	**4.3%**	**15.6%**	**15.7%**	**3.5%**	**5.5%**	**6.3%**	**17.6%**
Within R^2	**-**	**-**	**-**	**-**	**-**	**13.7%**	**1.2%**	**3.4%**	**4.2%**	**15.8%**
Text Attributes	No	Yes	No	Yes	Yes	No	Yes	No	Yes	Yes
Other Text Attributes	No	Yes	No	Yes	Yes	No	Yes	No	Yes	Yes
Financial Statement Items	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Surprise	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Fixed Effects	No	No	No	No	No	Firm	Firm	Firm	Firm	Firm
Observations	72,733	72,733	72,733	72,733	72,733	71,617	71,617	71,617	71,617	71,617

Panel B: Trading Decisions

	Dependent Variable: *3-day AVOL*									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Text_News_(AVOL)	**1.028*****				**0.990*****	**1.135*****				**1.135*****
	(0.00)				**(0.00)**	**(0.00)**				**(0.00)**
Adjusted R^2	**20.4%**	**3.8%**	**2.2%**	**4.7%**	**21.0%**	**23.0%**	**17.4%**	**17.0%**	**17.2%**	**23.4%**
Within R^2	**-**	**-**	**-**	**-**	**-**	**7.1%**	**0.1%**	**0.1%**	**0.1%**	**7.8%**
Text Attributes	No	Yes	No	Yes	Yes	No	Yes	No	Yes	Yes
Other Text Attributes	No	Yes	No	Yes	Yes	No	Yes	No	Yes	Yes
Financial Statement Items	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Surprise	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Controls	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Fixed Effects	No	No	No	No	No	Firm	Firm	Firm	Firm	Firm
Observations	72,733	72,733	72,733	72,733	72,733	71,617	71,617	71,617	71,617	71,617

<u>Notes:</u> This table reports sensitivity analyses related to the effectiveness of contextualized news to explain actions in equity markets in the presence of further financial statements and textual controls. The sample comprises 72,733 observations (the difference with the 104,254 main sample observations is due to data points with missing earnings surprise; the difference between 72,733 and 71,617 is due to singletons that are dropped for fixed effects estimations). Panel A reports results related to returns (i.e., pricing revisions). Panel B shows evidence for trading volumes (i.e., trading decisions).

In Panel A, the dependent variable is the cumulative abnormal return for the 3-day period (i.e., from day -1 to 1) surrounding an earnings announcement event. Cumulative abnormal returns are computed as actual security returns minus the return on the overall market (including dividends). In Panel B, the dependent variable is the abnormal trading volume for the 3-day period (i.e., from day -1 to 1) surrounding an earnings announcement event. Abnormal trading volumes are computed as the difference between average share-outstanding-scaled volumes in the 3-day announcement period (i.e., from day -1 to 1) minus average share-outstanding-scaled volumes in the non-announcement period (i.e., from day -130 to day -10 and from day 10 to day 130); the difference is then divided by the standard deviation of share-outstanding-scaled volumes in the non-announcement period.

In Panel A, the experimental variable is *Text_News_(CAR)* which represents cumulative abnormal returns modeled out-of-sample through contextualized earnings announcement texts. In Panel B, the experimental variable is *Text_News_(AVOL)* which represents abnormal trading volumes modeled out-of-sample through contextualized earnings announcement texts. A "RoBERTa" neural language model with 3 epochs and a 1e-5 learning rate is employed for modeling purposes. The neural language model is trained using a random sampling protocol (i.e., 50-50% train and modeling split) stratified by quarter-year (see Section 3.4 of the paper).

Text Attributes include *Tone*, *Fog*, *Length*, *Numbers*, and *Future*. *Other Textual Attributes* include *Ambiguity*, *Litigation*, *Macro_Topic*, *Investment_Topic*, and *Debt_Topic*. *Financial Statement Items* include *Earn*, *Earn_Chg*, *Div*, *Div_Chg*, *Lev*, *Lev_Chg*, *Restr*, and *Spec_Items*. *Controls* include *Size*, *Following*, and *AF*. *Surprise* is the earnings surprise. All variables are defined within Appendix A of the paper. Continuous variables are winsorized at the 1st and 99th percentiles. Cluster-robust-to-heteroskedasticity *p*-values are reported in parentheses; standard errors are clustered by firm (quarter-year) for models without (with) fixed effects. The Adjusted R-squared – "Adjusted R^2" – is used to compare models without fixed effects. The within R-squared – "Within R^2" – is used to compare models including firm fixed-effects. ***, **, * indicate significance at <0.01, <0.05, <0.10.

TABLE OA-5
Sensitivity Analyses – Using Returns Volatility as an Alternative Capital Market Outcome

	Dependent Variable: *RetVol*									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Text_News_(RetVol)	**1.011***** **(0.00)**				**1.021***** **(0.00)**	**1.013***** **(0.00)**				**0.932***** **(0.00)**
Adjusted R^2	**21.0%**	**3.5%**	**5.4%**	**7.2%**	**21.7%**	**26.2%**	**21.8%**	**23.3%**	**23.4%**	**27.1%**
Within R^2	**-**	**-**	**-**	**-**	**-**	**6.0%**	**0.3%**	**1.0%**	**1.2%**	**6.0%**
Text Attributes	No	Yes	No	Yes	Yes	No	Yes	No	Yes	Yes
Other Text Attributes	No	Yes	No	Yes	Yes	No	Yes	No	Yes	Yes
Financial Statement Items	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Surprise Controls	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Fixed Effects	No	No	No	No	No	Firm	Firm	Firm	Firm	Firm
Observations	72,733	72,733	72,733	72,733	72,733	71,617	71,617	71,617	71,617	71,617

Notes: This table reports sensitivity analyses related to the effectiveness of contextualized news to explain returns volatility around earnings announcement events. The sample comprises 72,733 observations (the difference with the 104,254 main sample observations is due to data points with missing earnings surprise; the difference between 72,733 and 71,617 is due to singletons that are dropped for fixed effects estimation).

The dependent variable is *RetVol* or the standard deviation of abnormal returns for the 3-day period (i.e., from day -1 to 1) surrounding an earnings announcement event. Abnormal returns are computed as actual security returns minus the return on the overall market (including dividends).

The experimental variable is *Text_News_(RetVol)* which represents the standard deviation of abnormal returns modeled out-of-sample through contextualized earnings announcement texts. A "RoBERTa" neural language model with 3 epochs and a 1e-5 learning rate is employed for modeling purposes. The neural language model is trained using a random sampling protocol (i.e., 50-50% train and modeling split) stratified by quarter-year (see Section 3.4 of the paper).

Text Attributes include *Tone*, *Fog*, *Length*, *Numbers*, and *Future*. *Other Textual Attributes* include: *Ambiguity*, *Litigation*, *Macro_Topic*, *Investment_Topic*, and *Debt_Topic*. *Financial Statement Items* include *Earn*, Earn_Chg, *Div*, *Div_Chg*, *Lev*, *Lev_Chg*, *Restr*, and *Spec_Items*. *Controls* include: *Size*, *Following*, and *AF*. *Surprise* is the earnings surprise. All variables are defined within Appendix A of the paper. Continuous variables are winsorized at the 1st and 99th percentiles. Cluster-robust-to-heteroskedasticity *p*-values are reported in parentheses; standard errors are clustered by firm (quarter-year) for models without (with) fixed effects. The Adjusted R-squared – "Adjusted R^2" – is used to compare models without fixed effects. The within R-squared – "Within R^2" – is used to compare models including firm fixed-effects. ***, **, * indicate significance at <0.01, <0.05, <0.10.

Online Appendix B – Textual Parsing

1. Downloading and Parsing of Earnings Announcement Texts

I download 8-K filings from SEC EDGAR using Python and the *sec-edgar-downloader* package (https://pypi.org/project/sec-edgar-downloader/). I download all 8-K filings for all available CIK identifiers within Compustat Fundamentals Quarterly within +/- 6 calendar days of the quarterly earnings announcement date.

I parse the HTML version of the downloaded 8-K filings using Python and the *Beautiful Soup* library (https://beautiful-soup-4.readthedocs.io/en/latest/). I extract the filing date using regular expressions searching for *"FILED AS OF DATE"* and then identify the earnings press release based on a (i) starting marker (e.g., "EX-99.1", "REPORTS", "EARNINGS") and an (ii) ending marker (e.g., "GRAPHIC"). I test the parsing algorithm on 30 disclosures and find that it accurately identifies the earnings announcement text in all instances. I discard documents for which starting and/or ending markers cannot be reliably identified.

I also download earnings press releases from Dow Jones Factiva. I collect textual documents from the Business Wire U.S. repository. I identify end-of-document markers (e.g., "DOCUMENT BWR0") and separate each earnings announcement. I also scrape each disclosure using regular expressions to extract the company's Ticker symbol (also verified by a firm's name) and the earnings announcement date.

I finally retain only non-duplicate SEC EDGAR and Business Wire U.S. textual disclosures and exclude documents with less than 50 words or less than 5 sentences – both identified using Python and the *NLTK* library (https://www.nltk.org/).

2. Identification of Tables and Generic Statements

I split each textual disclosure into sentences using Python and the *NLTK* library (https://www.nltk.org/). I test the sentence tokenization algorithm for 30 disclosures and find a 93% parsing accuracy. I then classify a sentence as "table" whenever it contains more than 20 non-breaking spaces (i.e., "\xa0"), dash symbols (i.e., "-"), or plus symbols (i.e., "+"). I exclude sentences classified as "tables" from the main analyses (see Section 3.1 of the paper for rationales and robustness checks).

I identify and exclude generic cautionary statements (see Section 3.1 of the paper for rationales and robustness checks) using regular expressions that match multiple tokens (e.g., "CERTAIN STATEMENTS", "CAUTIONARY STATEMENTS", "WORDS SUCH AS 'BELIEVE', 'MAY', 'WILL', 'EXPECT'"). One example of cautionary statements for Ocean Bio-Chem Inc. is reported thereafter.

Ocean Bio-Chem Inc. (CIK: 350737) – Second Quarter 2017
(Announcement Date: August 14th, 2017)

"Certain statements contained in this Press Release including without limitation, Company performance in the second half of 2017, the Company's entry into the pet market and commencement of production in the expanded portion of the Company's plant, constitute forward-looking statements. For this purpose, any statements contained in this report that are not statements of historical fact may be deemed forward-looking statements. Without limiting the generality of the foregoing, words such as "believe," "may," "will," "expect," "anticipate," "intend," "could" including the negative or other variations thereof or comparable terminology are intended to identify forward-looking statements."

Since long documents tend to display relatively more tables with non-standard formatting and relatively longer cautionary or generic discussions (that could go undetected using the prior algorithms), I exclude the last (and least informative) 512 tokens from disclosures with a number of tokens of at least 1 standard deviation above the average. Note that 512 tokens represent the maximum sequence length that *BERT-based* models can process – see Section 3.5 of the paper and Online Appendix C for details about applying "rolling windows" to still be able to consider the entire earnings announcement text.

3. *Identification of Relevant Numerical Tokens*

I identify and count a number in the following cases: (i) a numerical substring is preceded by a dollar sign ("$"); (ii) a numerical substring is followed by the words million/billion/trillion; (iii) a numerical substring is followed by a percentage sign ("%") or by the words "percent"/"pct". I also identify numbers in parentheses (negative sign) and/or for which the previous markers (i) are preceded by one or two white spaces; (ii) are not preceded by any white spaces; (iii) are capitalized, fully or in part (applies to words).

4. *Word Lists*

I use the Henry (2008) word list to compute *Tone*. I also use the Henry (2008) dictionary to perform "probing tasks" involving *Tone* words (see Section 5.6 of the paper).

I use the Loughran and McDonald Master Dictionary (https://sraf.nd.edu/textual-analysis/resources/) to compute *Ambiguity* (i.e., "uncertainty" words) and *Litigation* (i.e., "litigious" words).

To perform "probing tasks" involving performance-related words, I tabulate and use the 100 most frequent accounting performance terms) found within earnings disclosures (some of these keywords overlap with the Henry 2008 *Tone* words):

Word	Freq.	Word	Freq.	Word	Freq.	Word	Freq.
million	2.1%	months	0.3%	more	0.1%	record	0.1%
quarter	2.0%	loss	0.3%	services	0.1%	acquisition	0.1%
year	0.9%	growth	0.3%	stock	0.1%	continue	0.1%
net	0.8%	business	0.2%	strong	0.1%	development	0.1%
share	0.7%	revenues	0.2%	lower	0.1%	executive	0.1%
compared	0.7%	operations	0.2%	continued	0.1%	investment	0.1%
income	0.6%	adjusted	0.2%	nine	0.1%	comparable	0.1%
sales	0.5%	interest	0.2%	decrease	0.1%	reports	0.1%
company	0.5%	reported	0.2%	performance	0.1%	value	0.1%
second	0.4%	expenses	0.2%	full	0.1%	profit	0.1%
ended	0.4%	tax	0.2%	ebitda	0.1%	shares	0.1%
results	0.4%	related	0.2%	product	0.1%	customers	0.1%
increased	0.4%	costs	0.2%	segment	0.1%	debt	0.1%
increase	0.4%	up	0.2%	capital	0.1%	portfolio	0.1%
operating	0.4%	margin	0.2%	well	0.1%	markets	0.1%
diluted	0.4%	expense	0.1%	six	0.1%	charges	0.1%
revenue	0.4%	basis	0.1%	decreased	0.1%	guidance	0.1%
period	0.3%	three	0.1%	loans	0.1%	measures	0.1%
fourth	0.3%	billion	0.1%	exhibit	0.1%	service	0.1%
fiscal	0.3%	higher	0.1%	loan	0.1%	flow	0.1%
earnings	0.3%	products	0.1%	offset	0.1%	improved	0.1%
financial	0.3%	market	0.1%	cost	0.1%	production	0.1%
percent	0.3%	gross	0.1%	impact	0.1%	losses	0.1%
cash	0.3%	average	0.1%	information	0.1%	sale	0.1%
gaap	0.3%	release	0.1%	result	0.1%	equity	0.1%

To run the interpretability tests described within Section 3.7 of the paper, I delete from earnings disclosures the set of "stop-words" (i.e., connecting words) listed within the spaCy library (https://spacy.io/).

Within sensitivity analyses (see Section 5.7 of the paper), I also identify and control for relevant *topics* discussed within the earnings disclosures. Specifically, I use the following lists of words to identify *Macro*, *Investment*, and *Debt* narrative topics.

Macro_Topic
competition
competitive
competitor
economic
economy
industrial
industry

Investment_Topic		
acquisition	investment projects	sale proceeds
acquisitions	investments	spinoff
asset sale	joint venture	spin-off
asset sales	joint-venture	spun-off
been acquired	joint-ventures	the company acquired
capex	merger	to acquire
capital expenditure	mergers	was acquired
capital expenditures	partnership invested	we acquired
capital improvement	partnership sold	we have acquired

capital project	pipeline project	were invested
capital projects	pipeline projects	
divested	PP&E	
divestiture	(PP&E)	
divestitures	property and equipment	
expansion	property, plant and equipment	
facilities renewal	property, plant, and equipment	
facility renewal	refurbish	
invested assets	refurbished	
investing activities	refurbishes	
investment project	refurbishment	

Debt_Topic
debt
debts
debenture
debentures
indebtedness
interest
credit
obligation
note
notes
debit
debits
covenant
covenants
loan
loans

5. *Other Textual Attributes*

To compute the Gunning (1952) *Fog* (i.e., "readability") index, I tokenize each disclosure in words and sentences using Python and the *NLTK* library (https://www.nltk.org/). I also classify "complex" words (i.e., words with more than two syllables) using Python and the *Pyphen* library (https://pypi.org/project/pyphen/).

To compute *Future* and identify future-tense verbs, I use Python and the part-of-speech (POS) tagger tool provided by the *NLTK* library (https://www.nltk.org/).

Online Appendix C – BERT-based Neural Language Model Implementation Details

1. *Pre-trained Models, Hyperparameters, and Computational Time*

I fine-tune a pre-trained *RoBERTa* neural language model that can be freely downloaded from *GitHub* (https://github.com/pytorch/fairseq/tree/main/examples/roberta). In particular, I make use of the "RoBERTa Base" model (i.e., the least computationally intensive among the available implementations) and I utilize Python and the *Pytorch* framework (https://pytorch.org/) for fine-tuning.

I fine-tune all (*RoBERTa*) *BERT-based* neural language models using a "learning rate" of 1e-5 and 3 "training epochs".[1] The choice of this set of hyperparameters largely follows the procedure and implementation described in Siano and Wysocki (2021) for earnings announcement texts.

For each of the main prediction tasks related to *contemporaneous* abnormal stock returns (i.e., 3-day *CAR*) and *contemporaneous* abnormal trading volumes (i.e., 3-day *AVOL*), the average computational time for fine-tuning is 8 hours per training epoch (i.e., about 24 hours in total) using an NVIDIA Tesla V100 GPU accessible through Google Colab (https://research.google.com/colaboratory/faq.html). The out-of-sample modeling/prediction time amounts to about 10 hours for each outcome of interest.

For tasks involving the prediction of *future* outcomes (i.e., future-quarter earnings and the standard deviation of future 4-quarter earnings), I fine-tune a *RoBERTa* model each year between 1994 and 2019 using the disclosure observations over the prior 5 years (e.g., 1989-1993 to fine-tune the model in 1994, and so on). The fine-tuning time for each outcome is about 36 hours.[2] The out-of-sample modeling/prediction time amounts to about 10 hours, for each outcome of interest.

2. *"Rolling Windows" as a Solution to Process the Entire Earnings Announcement Text*

BERT-based language models (including *RoBERTa*) can process textual sequences of maximum 512 tokens. To overcome this limit and offer a more complete characterization of

---

[1] The *learning rate* represents the size of the step through which a loss or cost function is minimized. Too high learning rates could cause the global minimum of the cost function to be missed in the optimization process, while too large ones may increase computational time and render the model's training unfeasible. *Training epochs* can be thought of as the number of times that the entire dataset (training) is passed through the neural network's artificial neurons. Epochs are needed to properly minimize the models' loss function through an iterative process of gradient descent. The choice of the number of epochs is critical: too few iterations do not allow the model to properly minimize the loss or cost function, while too many iterations generally lead to in-sample overfitting and low out-of-sample accuracy.

[2] When this iterative approach is applied for fine-tuning purposes, the number of training observations for each yearly model will be significantly lower than the number of training observations used in the 50-50% split protocol (see Section 3.4 of the paper). This implies that the hyperparameters (i.e., learning rate and training epochs) may also need to be adapted. Since the fine-tuning time largely depends upon the number of training epochs, modeling alternative outcomes (to future income numbers) may well require a larger number of training epochs which would lead to substantially higher training time. In other words, 36 hours is not necessarily the fine-tuning time to be expected for modeling different/alternative outcomes.

contextualized news within corporate disclosures, I implement "rolling windows" that slide through the disclosure contents and allow to model the entire earnings press release. I therefore divide earnings announcements longer than 512 tokens into "windows" or subsequences of (maximum) 512 tokens. Each "window" or subsequence overlaps with the prior "window" or subsequence for 20% of the tokens (i.e., I define a "stride" of 80%).[3] Each generated subsequence is used for fine-tuning purposes. During out-of-sample modeling, *RoBERTa* outputs a prediction for each "window" or subsequence. I consider them all by computing the arithmetic mean of the predictions for all "windows", or subsequences, within an earnings disclosure.

[3] The "stride" can be defined as the distance chosen to slide the window when generating textual subsequences. The choice of the stride, as usual, should balance computational time and the model's accuracy. Smaller strides allow the model to learn from a larger number of examples and textual sequences but also require higher fine-tuning time.