Understanding Sentiment Through Financial Context

With Franco Wong

SOAR Accounting Symposium

December 2022

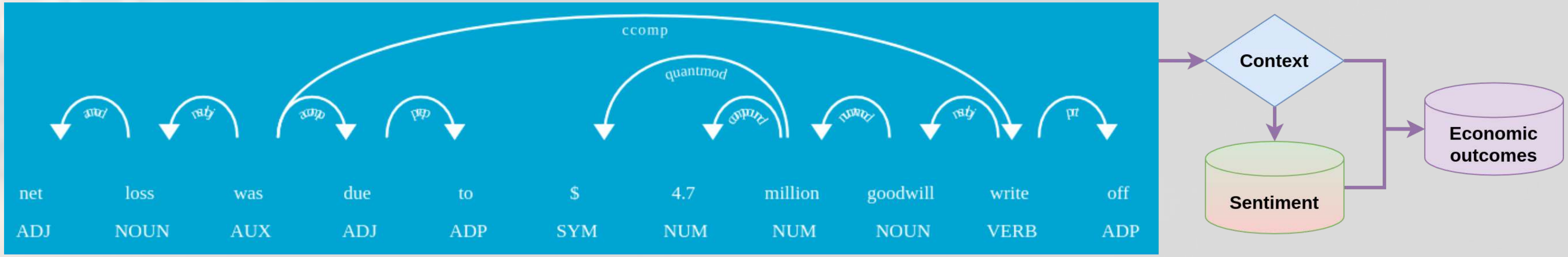
Dr. Richard M. Crowley
rcrowley@smu.edu.sg
Slides: https://rmc.link/SOAR · @prof_rmc

Research question and background

What do we ask?

1. How does sentiment depend on context?
   - A logical approach is to examine the text that sentiment comes from
2. Do prior results using financial sentiment hold across contexts?
3. Are prior results for different outcomes derived from the same underlying contexts?

> Why? To understand what financial sentiment captures and if it is empirically consistent



Main findings

1. Only a *few key contexts* drive each financial sentiment result
    - Aggregation to document-level sentiment adds a lot of noise
2. Sentiment, at the context level, often *contradicts prior results*
    - Aggregation removes nuance from our understanding
3. *Different contexts drive prediction* for different outcomes
    - Sentiment captures different empirical constructs across regressions
4. The above results hold across two other financial sentiment dictionaries
    - Our results are not unique to the LM dictionary
5. The above results hold using a neural network-based sentiment measure
    - Bag-of-words isn't the problem – financial sentiment, as a construct, likely is

> Punchline: Sentiment should be measured on fine-grained contexts, not full documents

- In other words, a precise matching between the text used and the economic question examined is needed

Related literature

1. Bag-of-words (dictionary) methods
   - *Word count* based
   - A few terms, such as 7 ethics terms (Loughran, McDonald and Yun 2009)
   - Longer lists like positive and negative sentiment (Loughran and McDonald 2011; Henry 2008)

> Dictionary methods ignore context entirely

2. Topic modeling
   - Still bag of words, but captures *document-level content*
   - Cannot be used for fine-grained context
   - Used on 10-Ks in Dyer, Lang and Stice-Lawrence (2017) and Brown, Crowley, and Elliott (2020)

> LDA ignores context within document and focuses on measuring the total content of a document

Related literature

3. Naïve Bayes
   - Adds supervised learning to bag-of-words
   - Used for measures of sentiment in Antweiler and Frank (2004) and Li (2010b)
4. Neural network approaches
   - Uses sentence-level context for classification
   - Used in Azimi and Agrawal (2021)
   - BERT-based approaches used in Siano and Wysocki (2021) and Huang, Wang and Yang (2022)

> Both naïve Bayes and neural networks can use context for *training* the model, but they don't provide a direct measure of context to researchers

Methodology: Measuring context

The idea

- Our goal is to replicate a natural approach that one would take to identify contexts by hand:
  1. Take a reference clause
  2. Look to see what the clause is about (the “context”)
  3. Assign the clauses into logical groupings of contexts
  4. After: Interpret sentiment of a clause within context

> In order to better understand context and its link to sentiment, we will examine a broad set of contexts spanning all MD&A content

Implementation

- Step 1a: Clause extraction and reconstruction (OpenIE)
- Step 1b: Filtering overlapping clauses
- Step 2: Extracting a numeric representation of the context (USE)
- Step 3: Clustering into contexts (MB K-means + Gap statistic)

Examples of contexts

Accounting

- *Policies*: Assumptions, Revenue Recognition, Tax, Cautionary Statements
- *Standards*: Standards, New standards
- *General or B/S*: Cash flow, Deferred tax
- *Income statement discussion*: Accounting losses, Depreciation and amortization

Business operations

- *Debt, Equity, and Investment*: Financing, Loans
- *Expectations and future*: Management expectations, Risk factor disclosures
- *Macroeconomics*: Interest rates, Market risk
- *Operations*: Growth, Customers, Products
- *Structure*: Subsidiaries, Partnerships

Changes

- Changes in: sales, expenses, operating measures
- Declines in: value or performance
- Increase in: expenses, income or revenue

Ungrouped

- *Grammatical patterns*
- *Timeframes*
- *Unrelated statements*
- *Unrelated statements with specific words*

What clauses are in the contexts?

Accounting assumptions

1. "Option pricing models require input of highly subjective assumptions particularly for expected stock price volatility"
2. "Weighted average assumptions determine net periodic pension benefit expense"

Growth

1. "Growth was partially offset by closure"
2. "Diamond 's capital expenditure budget is Diamond 's highest at approximately $ 250 million with much related to internal growth activities comprised of expansions of facilities"

Deferred tax

1. "Adtalem recognizes future tax benefits associated with tax loss as deferred tax assets"
2. "Company fully impaired deferred tax asset resulting in 5 % effective tax benefit rate"

Market risk

1. "We are exposed to market risk related to interest rate risk on investment of cash in securities with original maturities"
2. "Currency gains related to market risk"

Step 1a: Extracting clauses

Automating with Stanford Open IE

- Open IE is an open information extraction algorithm
- Generates triples of context of the form *(subject; relation verb; object)*
- Multi-step algorithm:
  1. Creates the dependency parse tree
  2. Resolves any co-references (“it,” “her,” etc.)
  3. Determines clause boundaries (multinomial logistic model)
  4. Determines triples within each clause (linguistic patterns)

This nets 179,703,756 extractions which can be formed back into clauses

Step 1b: Cutting this down a bit

- Some clauses are superfluous as we saw earlier
- Approach: Keep the shortest clauses such that…
  1. We cover as much of the sentence as possible without having nested clauses
  2. We don't drop words from LM
  3. We don't drop accounting content

> This cuts out 73% $\Rightarrow$ still have 48,576,229 clauses

Accounting content

- Harvey's hypertextual finance glossary
- "The largest financial glossary on the Internet"
- Some words unique to this dictionary:
  - demonetization, boilerplate, deductible

- NYSSCPA's Accounting Terminology Guide
- "Over 1,000 Accounting and Finance Terms"
- Some words unique to this dictionary:
  - GASB, MD&A, periodicity

Some shared words: collateral, specialist, hedge, debit, inventory

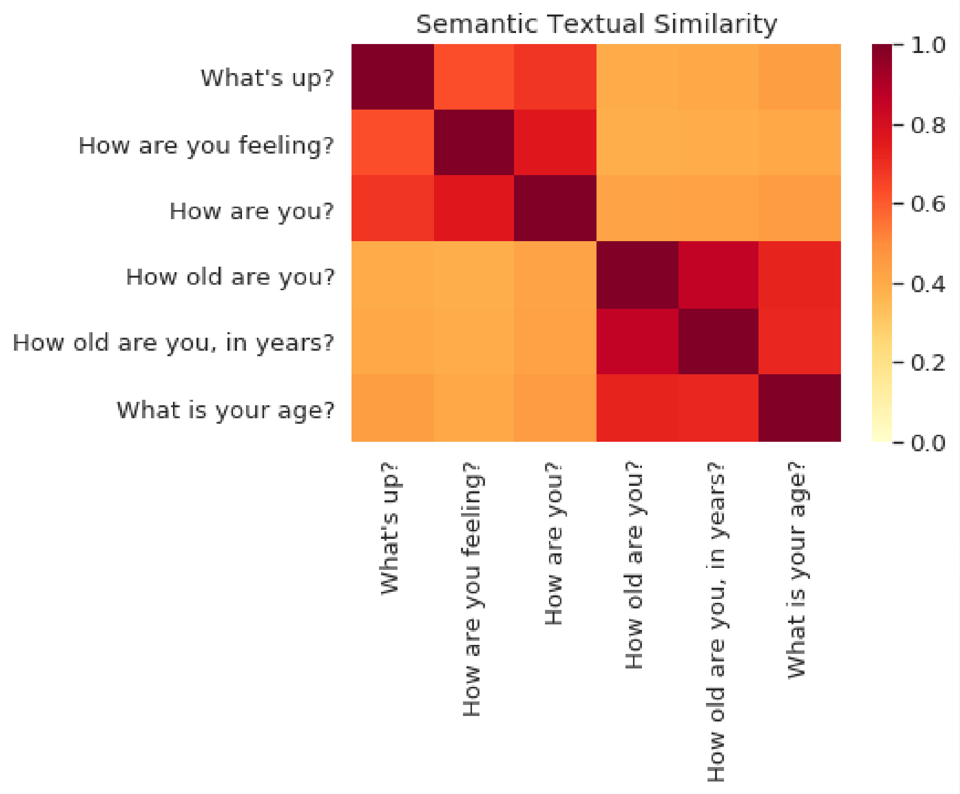
Step 2: Getting a numeric representation

1. Map all clauses to a 512-dimension vector space that represents underlying meaning
   - *Universal Sentence Encoder* (USE; Cer et al. 2018)
   - We mask out certain tokens that USE tends to focus on too much
     - Dates, times, dollar amounts, percentages, quantities, and ordinals

How does USE work?

- Input: Clauses' Words and word order
- Processing: Transformer-based neural network
  - Uses "attention"
- Output: A 512-dim vector per clause

> USE abstracts away from word choice!



Step 3: Clustering to contexts

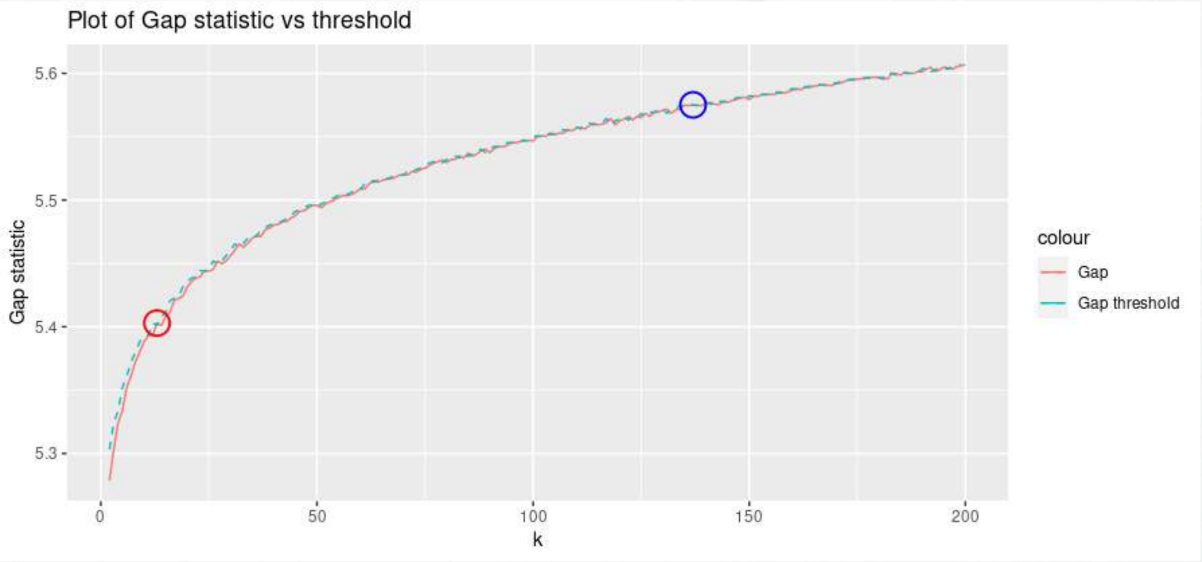
- We cluster within the 512-dim vector space with *Mini-Batch K-means* (Sculley 2010)
  - Mini-Batch K-means is an *online* version of K-means
    - Output is the same as K-means, but the process is more memory-friendly

Optimizing with Gap statistic

- Gap statistic (Tibshirani et al. 2001) is a simulation approach to supervising clustering
- Goal: Select the lowest k by comparing the informativeness of clustering on real data vs. synthetic data
  - Compare informativeness at k vs. at $k+1$, look for a gap <1 S.D.
  - Caveat: Optimal k may be too small in more varied text; thus we compare k to $k+1$ and $k+2$

13 is the lowest k vs $k+1$ (red circle), 137 is the lowest k vs $k+1$ and $k+2$ (blue circle)

137 contexts in the data



Validating our context methodology

1. Intrusion task
   - Take 3 clauses from 1 context and an “intruder” from another
   - E.g.:
     1. average market rate is in effect
     2. price swings are due to commodity costs
     3. net sales impact is in same store sales
     4. Volatility is in commodity prices
   - 4 RAs average 86% on the task; 500 questions each
     - **This is a very high score on the task!**
2. Overlap of original extractions with accounting dictionaries:
   - 95.2% contain at least 1 word in the Campbell Harvey’s dictionary
   - 84.8% contain at least 1 word in the NYSSCPA dictionary
3. Regress MD&A sentiment on clusters conditional on sentiment
   - 82.3% (68.6%) of variation captured for negative (positive) sentiment

Empirical approach

Data

- All *10-K and 10-K405 MD&A* sections to build the text model
  - 107,596 MD&As
  - 48,576,229 extractions
- Only MD&As subject to many requirements for empirical tests
  - 35,362 MD&As
  - 22,669,186 extractions
- Loughran McDonald sentiment from their *10X File summaries file*
- MD&A LM sentiment based on the 10-K parser from *Brown, Crowley and Elliott (2020)* (BCE)
  - The BCE parser has Pearson correlations $> 80\%$ for full text sentiment measures with LM
- Accounting data from *Compustat*
- Stock data from *CRSP*
- Material weaknesses from *Audit Analytics*

Empirics sketch

Three regression structures used throughout

1. To examine how sentiment relates to context
   - $Sentiment_{f,t} = \alpha + \sum_{i=1}^{137} \beta_i Context_{i,f,t} + \gamma \cdot Controls_{f,t} + \delta \cdot Industry\ FE + \varepsilon$
     - Run using a LASSO regression
2. To replicate results from Loughran McDonald (2011)
   - $DV_{f,t} = \alpha + \beta_0 Sentiment_{f,t} + \gamma \cdot Controls_{f,t} + \delta \cdot Industry\ FE + \varepsilon$
     - Run using a linear regression
3. To partition the replication on context
   - $DV_{f,t} = \alpha + \sum_{i=1}^{137} \beta_i Sentiment_{Context,i,f,t} + \gamma \cdot Controls_{f,t} + \delta \cdot Industry\ FE + \varepsilon$
     - Run using a LASSO regression

Practical issues with 137 IVs

- 137 text-derived measures means multicollinearity could flip coefficient signs and drop adjusted R^2

Solution 1: LASSO

- Replace OLS problem of $\min_{\beta,\gamma,\delta} \frac{1}{N} \left|\varepsilon\right|_2^2$ with:

$$\min_{\beta,\gamma,\delta} \frac{1}{N} \left|\varepsilon\right|_2^2 + \lambda \sum_{b\in\{\beta,\gamma,\delta\}} \left|b\right|_1$$

- Optimize λ with 10-fold cross-validation
- LASSO is is also called L^1 regularization
  - Standard technique for dealing with high VIFs
- Derive p-values using Post-LASSO estimator

> Much less worry about multicollinearity

- But some worry about dropping causal links

Solution 2: Double LASSO

- Determine causal links with $1 + \#IV$ LASSO regressions
- Estimate result using post-LASSO
- Ensures statistically significant links aren't dropped

> Avoids biasing against finding significant coefficients

- If we find that only a few contexts matter, it's because they really don't predict the outcome

Empirical results

Sentiment regressed on context

92 (79) contexts drive negative (positive) sentiment

- Some uniformly drive both positive and negative sentiment
  - "Cautionary statements" and "reduction in accounts"
- Some only drive negative sentiment
  - "Accounting losses" and "risk factor disclosures"
- Some only drive positive sentiment
  - "Increases in performance" and "tax"
- Some drive a lack of sentiment
  - "Depreciation and amortization" and "credit facilities"

As more coefficients' signs match to our intuition for negative sentiment, we argue that negative sentiment is more tied to context.

Filing period excess return

Prediction: Positive relation between sentiment and return

- Expected signs: Negative for negative sentiment, positive for positive sentiment
- Replication: Expected result for negative sentiment, null result for positive sentiment
- Contexts: Mixed findings, both sentiments drive results in both directions
- Double LASSO: Results are consistent

Filing period abnormal volume

Prediction: More sentiment (either), higher volume

- Expected signs: Positive for both
- Replication: Opposite result for negative sentiment, null result for positive sentiment
- Contexts: Mixed findings, but mostly in line with predictions
- Double LASSO: Results are consistent

Post-filing return volatility

Prediction: More sentiment (either), higher volatility

- Expected signs: Positive for both
- Replication: Expected result for negative sentiment, null result for positive sentiment
- Contexts: Mixed findings, both sentiments drive results in both directions
- Double LASSO: Results are consistent

Future material weakness

> Prediction: Inverse relation between sentiment and Material weaknesses

- Expected signs: Positive for negative sentiment, negative for positive sentiment
- Replication: Null result for negative sentiment, expected result for positive sentiment
- Contexts: Mixed findings, both sentiments drive results in both directions
- Double LASSO: Results are consistent

Falsification test

> Randomly assign each clause to one of 137 groups using a uniform distribution

- Table 4 replication: Difference between simulated and real data
  - Context drives 31.5% of the variation in negative sentiment
  - Context drives 10.8% of the variation in positive sentiment
- Tables 5 through 8 replication
  - All falsification tests have fewer significant coefficients on the context measures than our main results
  - All falsification tests have lower adjusted R^2 than our main results

> Our main results are unlikely to be driven by disaggregation in general

- Context is likely meaningful for sentiment

Construct validity of sentiment

Is sentiment a consistent construct? **It doesn't appear to be.**

- Negative sentiment:
  - No context always loads
  - "Discussion of accounting procedures" and "decreases in expenses or performance" load 3/4 of the time
  - 13 contexts significant only twice
  - 35 contexts significant only once

- Positive sentiment:
  - No context always loads
  - "decrease' + unrelated statements" loads 3/4 of the time
  - 4 contexts significant only twice
  - 43 contexts significant only once

This appears to violate how we approach sentiment empirically

- Aggregation is likely a problem

Other sentiment measures

Results are the same with the Henry (2008) and Harvard General Inquirer dictionaries

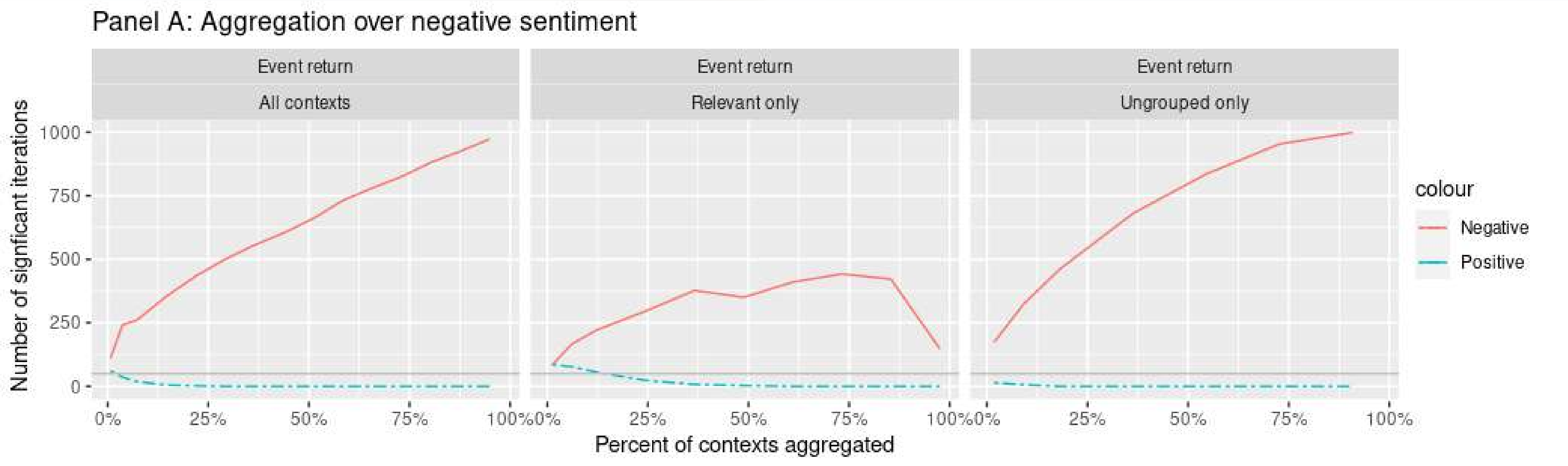
- The problem we document is not due to the LM dictionary's construction

Results are the same using FinBERT

- This means that bag-of-words isn't the source of the problem
- It also means that the problem source likely isn't classification accuracy

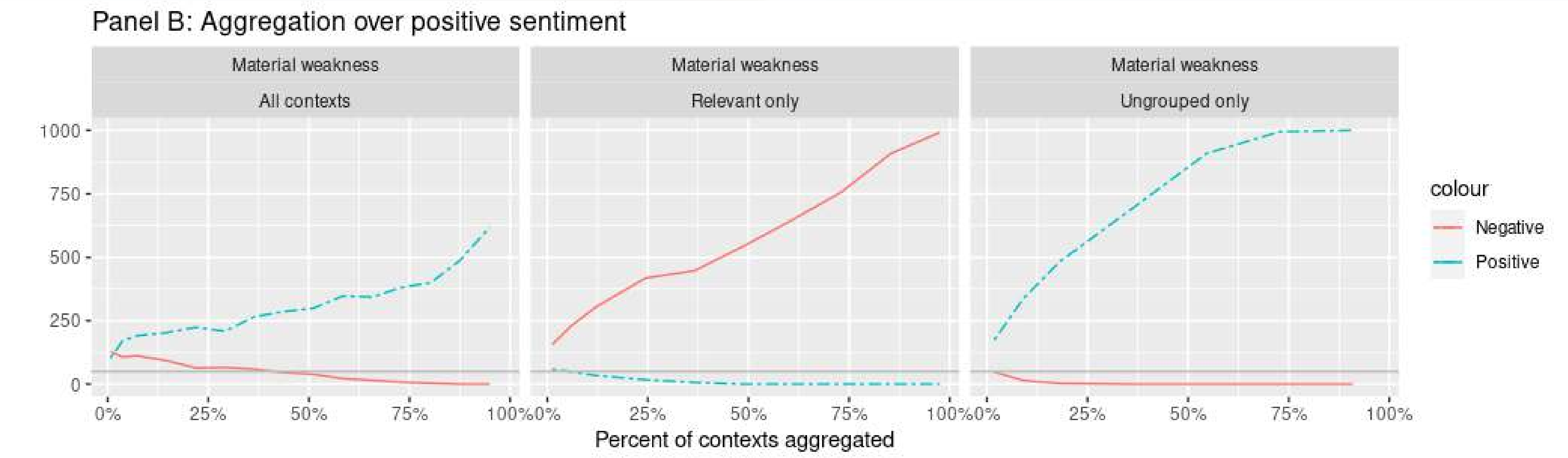
Aggregation is very likely to be the source of the problem

Simulating aggregation: Negative sentiment



Aggregation suppresses the mixed results found at low levels of aggregation

Simulating aggregation: Positive sentiment



> Aggregation can completely flip results depending on what is included in the aggregation

Conclusion

Findings and takeaways

1. Sentiment relies more on some contexts than others
2. *Context matters* for when regressing on sentiment
   - Some contexts behave as expected for sentiment, *many others do not!*
3. The regression DV matters
   - Sentiment results are driven by *different contexts* for *different DVs*

> Takeaway: Sentiment, at the document level, is not a consistent construct.

What should we, as researchers, do then?

For most papers

- Focus on *one context* within documents
  - Match to theory
- See, e.g., Hassan, Hollander, Van Lent, and Tahoun (2019 QJE) on political discussion

For papers needing a broad set of discussion

- Our *context methodology* offers a solution
- Unsupervised, automated, replicable
- Works for any document type

Thanks!

Richard M. Crowley
Singapore Management University
https://rmc.link/
@prof_rmc

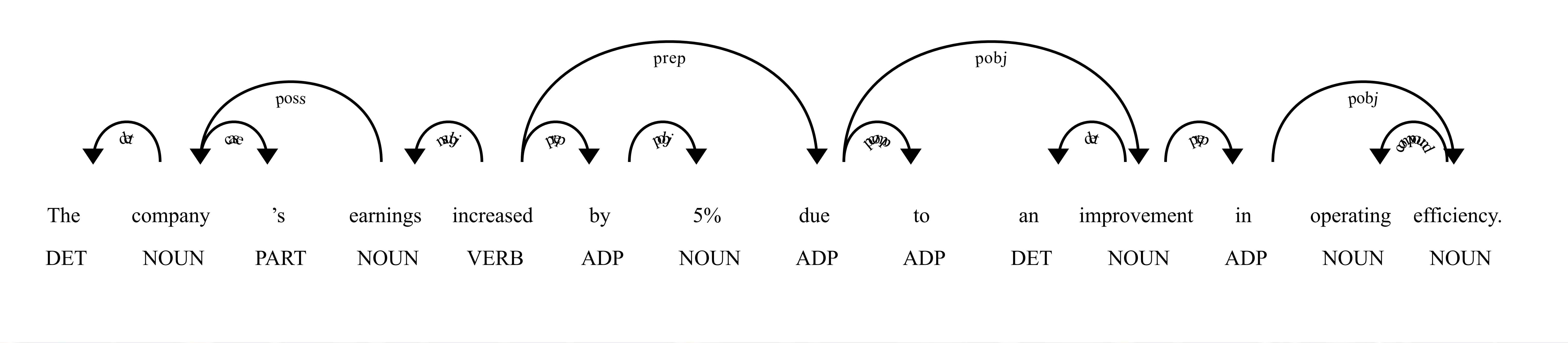
Packages used for these slides

- dplyr
- ggplot
- gridExtra
- kableExtra
- knitr
- revealjs

Other slides

Illustration of extracting clauses

> "The company's earnings increased by 5% due to an improvement in operating efficiency."



- (company; has; earnings)
- (company's earnings; increased by; 5%)
- (company's earnings; increased due; improved operating efficiency)
- (company's earnings; increased due; operating efficiency)

How does the Gap statistic work?

- Let…
  - k be the number of clusters,
  - B the number of simulated samples
  - W_k be the K-Means inertia score on actual data
  - $W^*_{k,r}$ be the K-Means inertia score for iteration r with synthetic data
  - $\bar{l}$ be the average of the $W^*_{k,r}$ s

$$Gap(k) = \left(\frac{1}{B}\right) \sum_{r=1}^{B} \log\left(W^*_{k,r}\right) - \log\left(W_k\right) \text{ and}$$

$$s_k = sd_k \sqrt{1 + \frac{1}{B}}, \text{ where } sd_k = \sqrt{\left(\frac{1}{B}\right) \sum_{r=1}^{B} \left\{\log\left(W^*_{k,r} - \bar{l}\right)\right\}^2}$$

- Select the lowest k such that $Gap(k) \geq Gap(k+1) - s_{k+1}$

> I.e., select the lowest k s.t. the log-scaled error removed by clustering on real data at k is no worse than 1 SD below the log-scaled error removed at $k+1$

Double LASSO

- The drawback of handling multicollinearity is removing variables that are potentially causally important
  - This can lead to questions on the validity of inferences derived from LASSO-based coefficients

> Solution: Double LASSO (Belloni, Chernozhukov and Hansen 2014 JEP)

$$(1) \qquad DV_{f,t} = \alpha + \sum_{i=1}^{137} \beta_i Sentiment_{Context,i,f,t} + \gamma \cdot Controls_{f,t} + \delta \cdot Industry\ FE + \varepsilon$$

$$(2) \qquad Sentiment_{Context,i,f,t} = \alpha + \sum_{j \neq i} \beta_j Sentiment_{Context,j,f,t} + [\ldots]$$

$$(3) \qquad DV_{f,t} = \alpha + \sum_{i \in S} \beta_i Sentiment_{Context,i,f,t} + \gamma \cdot Controls_{f,t} + \delta \cdot Industry\ FE + \varepsilon$$

$$S = [1, 137] \backslash \{i \text{ s.t. } 0 < \sum_{\beta_i \text{ from } \{(1),(2)\}} I(\beta_i \neq 0)\}$$

1. Run 138 LASSO regressions to determine significant links between outcome or IVs and IVs
2. Run a post OLS keeping only variables that had significant impact on the LASSO regressions

Other tables

Table 1

Table 1: Sample Construction

	Filings	
	Documents	Documents dropped
Unique 10-K filings	188,030	
Unique 10-K405 filings	20,139	
Total filings	208,169	
10-K with MD&A	93,551	-94,479
10-K405 with MD&A	14,045	-6,094
Total files with MD&As	107,596	

Sample restriction	MD&As	MD&As dropped	Extractions (clauses)	Extractions dropped
MD&A has extractions from OpenIE	105,921	1,675	48,576,229	
Filing matched to the Loughran McDonald data library	103,137	2,784	47,317,492	1,258,737
First filing per year	102,079	1,058	47,023,707	293,785
At least 180 days after last filing	101,877	202	46,942,952	80,755
CIK In CRSP Compustat Merged	56,460	45,417	31,219,059	15,723,893
Data available in Compustat	49,812	6,648	28,110,347	3,108,712
Market cap available in CRSP	49,411	401	27,896,026	214,321
Price on t-1 >= $3	41,693	7,718	23,988,897	3,907,129
Return & volume has >= 60 obs from trading days [-252,-6]	40,489	1,204	23,344,479	644,418
NYSE, AMEX, or NASDAQ listed	40,476	13	23,336,694	7,785
Book to market available and positive	39,466	1,010	22,734,045	602,649
At least 2000 words in the 10-K	39,357	109	22,730,774	3,271
At least 250 words in the MD&A	35,362	3,995	22,669,186	61,588

Table 2, Panels A and B

Table 2: Context Frequencies

Panel A: Most and least frequent contexts by clause count (excluding Ungrouped text)

	Most frequent contexts	Number of clauses	Number of documents		Least frequent contexts	Number of clauses	Number of documents
1	Increases in accounts	337,697	32,100	128	Depreciation and amortization	92,892	19,481
2	Loans issued	279,540	18,325	129	Company expectations	92,490	20,344
3	Mixed business activities	266,742	18,069	130	Risk factor disclosures 2	87,818	23,780
4	Revenue recognition	260,647	28,908	131	Prices	76,571	20,986
5	Sales of goods or assets	253,431	29,193	132	Deferred tax	73,184	16,587
6	Interest rates	248,814	26,361	133	Economic and business conditions	69,642	20,402
7	Funds and financing activities	247,055	17,051	134	New accounting standards	65,962	17,518
8	General business description	241,394	25,927	135	Partners in partnership	61,671	13,074
9	Increases in income or revenue	225,729	27,029	136	Cautionary statements	55,234	23,860
10	Tax	221,030	27,630	137	Partnerships	35,387	5,272

Panel B: Most and least frequent contexts by document count (excluding Ungrouped text)

	Most frequent contexts	Number of clauses	Number of documents		Least frequent contexts	Number of clauses	Number of documents
1	Increases in accounts	337,697	32,100	128	Funds and financing activities	247,055	17,051
2	Increases in performance	220,800	31,038	129	Energy	160,500	17,027
3	Operating performance	172,906	30,441	130	Leases	123,008	16,892
4	Income statement items	167,842	30,417	131	Deferred tax	73,184	16,587
5	Worsening performance outcomes	203,605	30,125	132	Discussion of accounting procedures	182,118	16,257
6	Cash flows	192,820	29,837	133	Negative accounting outcomes	103,508	15,124
7	Financing and investment	180,431	29,267	134	Contracting with other entities	99,712	14,695
8	Sales of goods or assets	253,431	29,193	135	Accounting standards	108,039	13,485
9	Expenses and provisions	176,610	29,180	136	Partners in partnership	61,671	13,074
10	Decreases in expenses or performance	218,754	29,105	137	Partnerships	35,387	5,272

Table 2, Panels C and D

Table 2 (Continued): Context Frequencies

Panel C: Most and least frequent contexts by percent of negative clauses within context

	Most frequent contexts	Number of clauses	Percent of clauses		Least frequent contexts	Number of clauses	Percent of clauses
1	**Losses**	187013	94.77%	128	**Income statement items**	3239	1.93%
2	**Declines in value or performance**	108445	93.49%	129	Dates with events	2462	1.86%
3	**Negative accounting outcomes**	78293	75.64%	130	Reporting periods	1696	1.68%
4	**Accounting losses**	79245	68.05%	131	Headers	980	1.45%
5	**Risk factor disclosures**	82938	63.03%	132	"Company" + mixed accounts	2082	1.33%
6	**Worsening performance outcomes**	65922	32.38%	133	Increases with time reference	2152	1.17%
7	Unrelated statements 6	37606	30.06%	134	Cash headings	1172	1.07%
8	**Economic impact on company**	39218	27.01%	135	Dates	268	0.58%
9	**Decreases in different measures**	36761	23.71%	136	"Increase" + unrelated statements	515	0.36%
10	**Economic and business conditions**	15730	22.59%	137	Percents in year	407	0.24%

Panel D: Most and least frequent contexts by percent of positive clauses within context

	Most frequent contexts	Number of clauses	Percent of clauses		Least frequent contexts	Number of clauses	Percent of clauses
1	**Tax**	66248	29.97%	128	"Notes" + unrelated statements	1217	1.27%
2	Unrelated statements 6	20927	16.73%	129	"Decrease" + unrelated statements	1063	1.25%
3	**Accounting standards**	17981	16.64%	130	Headers	671	0.99%
4	**Partners in partnership**	9368	15.19%	131	**Changes in interest and forex rates**	1095	0.95%
5	"Company" + unrelated statements 2	11194	13.50%	132	**Costs or expenses**	1001	0.80%
6	"We" or "our" + change statements	17447	13.24%	133	Cash headings	667	0.61%
7	**Economic and business conditions**	8828	12.68%	134	Dates	179	0.39%
8	**Changes in operating measures**	16781	12.24%	135	**Losses**	586	0.30%
9	**Growth**	19165	11.81%	136	**Declines in value or performance**	272	0.23%
10	**General business description**	27256	11.29%	137	Percents in year	191	0.11%

Table 3

Table 3: Univariate Statistics

Variable	Obs	Mean	SD	5%	Median	95%
Sentiment measures						
Negative, Full 10-K, LM parser	35,362	1.55%	0.45%	0.81%	1.54%	2.29%
Negative, Full 10-K, Our parser	35,362	1.30%	0.48%	0.58%	1.27%	2.13%
Negative, MD&A, Our parser	35,362	1.22%	0.59%	0.42%	1.14%	2.32%
Positive, Full 10-K, LM parser	35,362	0.68%	0.18%	0.44%	0.65%	1.01%
Positive, Full 10-K, Our parser	35,362	0.64%	0.19%	0.38%	0.61%	0.97%
Positive, MD&A, Our parser	35,362	0.65%	0.29%	0.26%	0.61%	1.16%
Extraction measures						
Clauses per MD&A	35,362	641.1	457.9	75.0	548.0	1,511.0
Negative clauses per MD&A	35,362	36.6	34.8	2.0	27.0	105.0
Positive clauses per MD&A	35,362	20.1	16.9	1.0	16.0	52.0
Dependent variables						
Event period excess return	35,362	-0.36%	7.65%	-11.47%	-0.27%	10.26%
Event period abnormal volume	35,361	0.493	3.848	-0.771	-0.059	3.062
Post-event return volatility	35,362	0.160	0.131	0.000	0.143	0.331
Material weakness count, t+1	23,034	0.153	0.782	0	0	1
Control variables						
log(Market value)	35,362	12.72	1.72	10.14	12.60	15.74
log(BTM)	35,362	-7.63	0.926	-9.21	-7.527	-6.35
log(Share turnover)	35,362	1.37	1.09	-0.553	1.45	2.98
Pre-event FF alpha	35,362	0.08%	2.50%	-2.91%	0.04%	3.17%
I(Nasdaq)	35,362	59.50%	4.91%	0	1	1

What contexts are high in both sentiments?

	Handcoded prediction (1)	Negative MD&A Tone (2)	Positive MD&A Tone (3)	Handcoded prediction (4)
Part A: High sentiment contexts				
Context: Accounting				
Cautionary statements	+	0.082 ***	0.018 **	
Context: Business operations				
Economic impact on company		0.029 ***	0.017 ***	
Employee matters		0.037 ***	0.015 ***	
Market condition and competition		0.013 **	0.039 ***	
Operating performance		0.033 ***	0.014 ***	
Context: Changes				
Reduction in accounts		0.023 ***	0.024 ***	
Worsening performance outcomes		0.071 ***	0.016 ***	

What contexts skew towards negative sentiment?

	Handcoded prediction (1)	Negative MD&A Tone (2)	Positive MD&A Tone (3)	Handcoded prediction (4)
Part B: Contexts skewed toward negative				
Context: Accounting				
Accounting losses	+	0.180 ***	0.004	
Expenses and provisions		0.023 ***	-0.002	
Losses		0.136 ***	-0.005 **	
Negative accounting outcomes	+	0.123 ***	-0.003	
Noncurrent assets		0.016 ***	-0.011 ***	
Context: Business operations				
Customers		0.017 ***	-0.011 ***	
Economic and business conditions		0.029 ***	0.011 *	
Loans		0.026 ***	-0.005 **	
Loans issued		0.008 ***	0.000	
Management expectations		0.068 ***	-0.004	
Market risk		0.010 ***	-0.018 ***	
Risk factor disclosures	+	0.252 ***	-0.008 **	
US Regulatory		0.011 ***	.	
US-centric statements		0.014 ***	-0.005 **	
Context: Changes				
Declines in value or performance	+	0.140 ***	-0.004	
Decreases in different measures		0.029 ***	0.005	

What contexts skew towards positive sentiment?

	Handcoded prediction (1)	Negative MD&A Tone (2)	Positive MD&A Tone (3)	Handcoded prediction (4)
Part C: Contexts skewed toward positive				
Context: Accounting				
Accounting standards		0.00529	0.00583 **	
Cash flows		-0.00099	0.01313 ***	
Income statement items		-0.00493	0.02429 ***	
Interest income or expense		.	0.01266 ***	
New accounting standard		-0.01006 *	0.01408 ***	
Tax		-0.00907 ***	0.02547 ***	
Context: Business operations				
Continuation or going concern		-0.008	0.066 ***	
Contracting with other entities		-0.031 ***	0.015 ***	
Expected outcomes		0.005	0.014 ***	
Financing and investment		-0.039 ***	0.012 ***	
General business description		-0.007 ***	0.015 ***	
Growth		-0.016 ***	0.035 ***	
Investments		-0.010 **	0.013 ***	
Investments and horizons		-0.031 ***	0.018 ***	
Leases		-0.007 **	0.004 **	
Partners in partnership		-0.004	0.011 ***	
Products		0.003	0.005 **	
Context: Changes				
Change in sales		-0.024 ***	0.018 ***	
Changes in operating measures		-0.006	0.010 ***	
Increases in performance		-0.018 ***	0.020 ***	+

What contexts are low in both sentiments?

	Handcoded prediction (1)	Negative MD&A Tone (2)	Positive MD&A Tone (3)	Handcoded prediction (4)
Part D: Low sentiment contexts				
Context: Accounting				
Accounting processes		-0.012 **	-0.010 **	
Depreciation and amortization		-0.036 ***	-0.011 ***	
Large expenses		-0.034 ***	-0.021 ***	
Revenue recognition		-0.022 ***	-0.009 ***	
Context: Business operations				
Credit facilities		-0.019 ***	-0.015 ***	
Energy		-0.012 ***	-0.003 **	
Options and ESOs		-0.051 ***	-0.019 ***	
Sales of goods or assets		-0.010 ***	-0.007 ***	
Subsidiaries		-0.029 ***	-0.014 ***	
Context: Changes				
Decreases in expenses or performance		-0.013 ***	-0.024 ***	
Increases in accounts		-0.026 ***	-0.010 ***	

Full Table 5

Table 5: Predicting Filing Period Excess Return

Contexts conditional on:	Negative sentiment		Positive Sentiment		Neutral
Variable	(1)	(2)	(3)	(4)	(5)
Negative, MD&A, Our parser	-0.241 ***				
Positive, MD&A, Our parser			-0.13		
Accounting policies					
Accounting processes		0.226 **		.	.
Discussion of accounting procedures		0.566 ***		0.19	0.197 ***
Fair value measurement		0.591 ***		.	0.032
Tax		.		0.144 **	0.043
Accounting standards					
Accounting standards		0.326 ***		0.146 *	0.006
General and balance sheet discussion					
Negative accounting outcomes		-0.009 **		.	0.18 *
Income statement discussion					
Accounting losses		-0.098 ***		0.257	-0.014
Debt, Equity, and investment					
Debt transactions		0.072 **		0.053	.
Expectations / future					
Company expectations		.		-0.338 *	-0.102 ***
Continuation or going concern		-0.532 ***		.	.
Expected outcomes		.		-0.289 **	.
Risk factor disclosures		-0.215 ***		.	-0.196 **
Operations					
Products		-0.233		-0.542 ***	-0.125 ***
US-centric statements		.		-0.253 **	.
Structure					
Contracting with other entities		.		.	0.086 ***
Changes					
Changes in expenses		.		.	-0.069 **
Decreases in different measures		0.139 **		.	.
Increase in expenses		.		.	0.093 **
Worsening performance outcomes		.		.	0.179 ***
Ungrouped					
Modal weak statements		-0.429 ***		-0.011	-0.147 **
Percents in year		.		-1.937 **	0.032
Dates with unrelated statements		-0.034 *		.	-0.046 **
Time references + "our"		.		0.827 ***	.
Unrelated statements 4		.		.	0.028 **
"Our" + unrelated statements		.		-0.293 **	-0.035
"Statements" + unrelated statements		-0.11 **		.	.
"We" + unrelated statements 2		.		-0.456 ***	-0.144 ***
"We/our" + operations statements		-0.225		.	-0.165 ***
Controls					
log(Market value)	0.002 ***	0.001 ***	0.002 ***	0.002 ***	0.002 ***
log(BTM)	0.001 **	.	0.001 *	0.000	.
log(Share turnover)	-0.005 ***	-0.004 ***	-0.005 ***	-0.004 ***	-0.004 ***
Pre-event FF alpha	0.012	.	0.012	.	0.001
I(Nasdaq)	0.001	0.000	0.001	.	.
FF48 Industry FE	Included	Included	Included	Included	Included
Adjusted R^2	0.009	0.014	0.009	0.012	0.014
# of negative and significant contexts		6		6	8
# of positive and significant contexts		6		2	5
Double LASSO adjustment					
Adjusted R^2		0.013		0.010	0.013
# of negative and significant contexts		6		5	6
# of positive and significant contexts		7		3	7

Columns (1) and (3) report linear regressions, while columns (2), (4), and (5) report LASSO regressions including all 137 contexts. All contexts are restricted to only the sentiment specified in the column. Only contexts that are significant at p<0.05 for at least one regression from columns (2), (4), and (5) are included. All regressions are based on 35,362 observations. P-values are indicated as follows: * indicates p<0.10, ** indicates p<0.05, and *** indicates p<0.01. A period indicates that the variable was dropped in the regression by the LASSO procedure. The bottom section presents the adjusted R-squared and number of significant contexts by sign (at p<0.05) when using a Double LASSO procedure as in Belloni, Chernozhukov and Hansen (2014) as described in Section 3.1.