Linguistic Complexity in Firm Disclosures: Obfuscation or Information

Bryan

June 7, 2024

Outline

Research Quesion

Two previous views on the use of complex language

- 1. obfuscate to increase the information processing costs and delay the market reaction to the news Li, 2008
- 2. It is necessary to convey information about the firm's business transactions and operating strategy Bloomfield, 2008

How do two latent components of complexity language, the first is Obfuscation, the second is information, related to information asymmetry?
Conference calls are informative to market participants and lead to reductions in information asymmetry

Outline

Motivation

Regulators have often expressed concern that the language in firms' disclosures has become increasingly complex.

And previous literature shows concern about the old methods (fog index), which would provide valuable information now.

This one focus on the use of complex language in conference calls, rather than in mandatory SEC filings

Individual investors might overreact, so the management has a stronger incentive to obfuscate. So the call would be a perfect proxy for this analysis.

Outline

equation-1

LC_Manager: the language complexity of manager, determined by two composition.

- $Info^*$:the information content of a manager's language in the absence of obfuscation
- $Obfu^*$: intentional obfuscation, or an intentional reduction in informative language

$$LC\_Manager = \phi_0 + \phi_1 Info^* + \phi_2 Obfu^* + \varepsilon, (1)$$

$$where \phi_1 \geq 0 and \phi_2 \geq 0.$$

We seek to recover empirical estimates of the latent variables, $Info^*$ and $Obfu^*$

equation 2

Analysts have incentives to acquire and convey value-relevant information during conference calls.Accordingly, we assume that the analyst's linguistic complexity on the call, which we represent as LC Analyst, reflects only the intrinsic amount of informative technical disclosure on the call. Formally,

$$LC\_Analyst = \delta_0 + \delta_1 Info^* + \nu, (2)$$

Under the identifying assumption that analysts' linguistic complexity does not reflect intentional obfuscation, we can use the linguistic complexity of analysts on the call as a benchmark level of complexity that one would expect in the absence of obfuscation.

equation 3

In particular, appendix A shows that we can recover estimates of $Info^*$ and $Obfu^*$ from a regression of $LC\_Manager$ on $LC\_Analyst$,

$$LC\_Manager = \beta_0 + \beta_1 LC\_Analyst + \eta, (3)$$

where the fitted value is an estimate of the latent information component $Info$ and the residual is an estimate of the latent obfuscation component $Obfu$.

equation 4

to validate that our estimates of the two latent components measure the respective theoretical constructs, we rely on economic theory that suggests that obfuscatory (informative) disclosure is associated with greater (lower) information asymmetry between shareholders and managers, $InfoAsym$ (e.g., Bloomfield [2002]). Formally we assume,

$$InfoAsym = \gamma_0 + \gamma_1 Info^* + \gamma_2 Obfu^* + \eta, (4)$$

where $\gamma_2 \geq 0$ and $\gamma_1 < 0$. Note that equation (4) expresses information asymmetry as a function of the unobserved latent components of linguistics.

equation 5

$$InfoAsym = \gamma_0 + \gamma_1 Info^* + \gamma_2 Obfu^* + \eta, (5)$$

if our decomposition is empirically descriptive, we expect to decompose a single variable—whose relation to information asymmetry is ambiguous—into two separate components, one that is positively related to information asymme try and one that is negatively related to information asymmetry.
And the epirmrical exanmie show that obfuscation in analyst words is mild and no effect on result .

To equation 3

LC Manager : use Fog(.) = 0.4 ×(average number of words per sentence+percent of complex words).

Fog(Present) represents the Fog index of managers' language during the presentation, Fog(Response) is the Fog of managers' responses to questions.

Fog (analyst), using the analysts' questions and statements during the QA portion of the call

To equation 5

We measure information asymmetry using the Amihud [2002] illiquidity construct (see, e.g., Lang and Maffett [2011]). Following Amihud [2002], Illiquidity, is defined as

$$Illiquidity_t = \frac{|R_t|}{DVolume_t},$$

where Rt is the daily return and DVolumet is the daily dollar volume (in millions).

Outline

Stats

TABLE 1
Descriptive Statistics

Panel A: Firm characteristics

Variable	Mean	Std. Dev.	P25	Median	P75
Acquisitions	0.096	0.025	0.000	0.000	0.000
BM	0.509	0.382	0.257	0.431	0.670
CapIntensity	0.247	0.242	0.059	0.159	0.367
Capex	0.012	0.016	0.003	0.007	0.015
Coverage	9.240	6.318	4.000	7.000	13.000
Dispersion	0.002	0.004	0.000	0.001	0.002
Financing	0.056	0.096	0.000	0.003	0.021
Goodwill	0.028	0.165	0.000	0.000	0.000
Inkmod	0.023	0.013	0.014	0.019	0.028
Leverage	0.211	0.200	0.028	0.176	0.328
Loss	0.167	0.373	0.000	0.000	0.000
MgmtForecast	0.514	0.500	0.000	1.000	1.000
ROTI	0.011	0.023	0.000	0.000	0.013
Restructuring	0.228	0.419	0.000	0.000	0.000
Returns	2.314	23.284	−11.012	1.826	14.299
Size	7.168	1.599	6.033	6.998	8.154
SmallBeat	0.150	0.357	0.000	0.000	0.000
SpinItems	−0.003	0.015	−0.001	0.000	0.000
Surprise	0.000	0.010	−0.001	0.001	0.002
σ CFO	0.060	0.069	0.022	0.039	0.069

Panel B: Measures of linguistic complexity and information asymmetry

Variable	Mean	Std. Dev.	P25	Median	P75
Fog(Analyst)	8.851	2.513	8.384	9.231	10.098
Fog(Present)	15.861	1.991	14.849	15.899	16.932
Fog(Response)	13.956	1.644	10.804	11.840	12.957
Illiquidity	0.076	2.297	0.000	0.002	0.008

This table presents descriptive statistics for firms in our sample. Our sample is constructed from the intersection of Thomson Reuters StreetEvents, I/B/E/S, and CRSP/Compustat. The sample spans the time period January 2002 to December 2011 and covers a total of 40,172 firm-quarter observations. Panel A reports the distribution of various firm characteristics used in our analysis; panel B reports the distribution of measures of linguistic complexity and information asymmetry. All variables are winsorized at the 1st and 99th percentiles. See appendix C for variable definitions.

Figure: stats

test result for equation 3

TABLE 2

Estimating the Latent Components of Managers' Linguistic Complexity

	Dependent Variable:						*p*-Value: Test of Difference in Coefficients (3) − (6)
	Fog(Present)			*Fog(Response)*			
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Fog(Analyst)	0.09***	.	0.09***	0.22***	.	0.21***	[<0.001]
	(20.19)	.	(20.57)	(30.24)	.	(31.49)	
Size	.	−0.56***	−0.58***	.	0.34***	0.28***	[<0.001]
	.	(−6.38)	(−6.78)	.	(4.73)	(4.10)	
Leverage	.	0.23***	0.24***	.	0.08	0.09	[0.049]
	.	(2.99)	(3.10)	.	(1.29)	(1.54)	
BM	.	−0.08	−0.07	.	−0.21***	−0.18***	[0.10]
	.	(−1.09)	(−0.98)	.	(−3.39)	(−3.20)	
Returns	.	−0.13***	−0.15***	.	0.02	−0.05**	[<0.001]
	.	(−4.54)	(−5.93)	.	(0.64)	(−2.11)	
Acquisitions	.	−0.08	−0.09	.	−0.03	−0.05	[0.488]
	.	(−1.26)	(−1.40)	.	(−0.57)	(−1.00)	
CapIntensity	.	−0.47***	−0.39***	.	−0.78***	−0.59***	[0.042]
	.	(−4.51)	(−3.78)	.	(−9.15)	(−7.26)	
Capex	.	−0.04	−0.07	.	0.07	0.00	[0.329]
	.	(−0.53)	(−0.92)	.	(1.07)	(0.08)	
R&D	.	0.58***	0.57***	.	0.44***	0.42***	[0.128]
	.	(5.54)	(5.57)	.	(4.99)	(5.05)	
Financing	.	0.02	0.02	.	0.02	0.02	[0.975]
	.	(0.46)	(0.45)	.	(0.54)	(0.51)	
σ CFO	.	0.29***	0.30***	.	0.22***	0.24***	[0.369]
	.	(3.74)	(3.90)	.	(3.40)	(3.91)	
Goodwill	.	0.12**	0.11**	.	0.08*	0.07	[0.44]
	.	(2.25)	(2.22)	.	(1.66)	(1.62)	
Restructuring	.	0.00	−0.02	.	0.12***	0.08**	[0.030]
	.	(0.04)	(−0.42)	.	(3.29)	(2.14)	
N-obs	60,172	60,172	60,172	60,172	60,172	60,172	
Adj R^2	2.15	5.02	7.08	11.65	3.73	14.34	

This table presents results from estimating the linguistic complexity of managers during the respective portion of the conference call, *Fog(Present)* and *Fog(Response)*, as a function of linguistic complexity of analysts, *Fog(Analyst)*, and variables related to business complexity. We use the following variables to measure business complexity: firm size (*Size*), firm leverage (*Leverage*), book-to-market ratio (*BM*), historical stock performance (*Returns*), acquisitions (*Acquisitions*), capital intensity (*CapIntensity*), capital expenditures (*Capex*), research and development (*R&D*), debt and equity issuance (*Financing*), cash flow volatility (*σ CFO*), goodwill impairments (*Goodwill*), and restructuring charges (*Restructuring*). See appendix C for variable definitions. For ease of interpretation, each of these variables is ranked into deciles and scaled to range from 0 to 1. *t*-statistics appear in parentheses (two-tailed *p*-values appear in brackets) and are based on standard errors clustered by firm and date.

***, **, and * denote statistical significance at the 0.01, 0.05, and 0.10 levels (two-tail), respectively.

Figure: stats

Decomposition on Obfuscation and information.

TABLE 3

Distribution of Estimated Latent Components of Linguistic Complexity

Panel A: Descriptive statistics

Variable	Mean	Std. Dev.	P25	Median	P75
Obfu(Present)	0.000	1.533	−0.978	0.028	1.026
Info(Present)	15.861	0.423	15.599	15.855	16.132
Obfu(Response)	0.000	1.521	−1.061	−0.122	0.913
Info(Response)	11.956	0.623	11.724	12.038	12.332

Panel B: Correlation matrix of estimated latent components of linguistic complexity

Variable	(1)	(2)	(3)	(4)	(5)	(6)
Presentation Portion						
(1) *Fog(Present)*	1.00	**0.96**	0.25	0.37	0.32	0.19
(2) *Obfu(Present)*	**0.96**	1.00	0.00	0.31	0.34	0.02
(3) *Info(Present)*	0.27	0.00	1.00	0.25	0.02	**0.62**
Response Portion						
(4) *Fog(Response)*	0.38	0.32	0.26	1.00	**0.91**	0.42
(5) *Obfu(Response)*	0.34	0.35	0.00	**0.93**	1.00	0.09
(6) Info(Response)	0.19	0.00	**0.70**	0.38	0.00	1.00

This table presents descriptive statistics and correlations among the latent components of linguistic complexity. Panel A reports the distribution of our empirical estimates of the latent components of managers' linguistic complexity. Panel B reports the correlations among the latent components of linguistic complexity. *Fog(.)* is the Fog index of the respective portion of the conference call. *Obfu(.)* is the latent obfuscation component during the respective section of the call and *Info(.)* is the latent information component during the respective section of the call. Spearman (Pearson) correlations appear above (below) the diagonal and bold denotes correlations in excess of 0.5.

Figure: stats

To eqution 5

TABLE 4

Linguistic Complexity and Information Asymmetry

	Linguistic Complexity		Linguistic Complexity Decomposed		Control for Additional Language Attributes	
	(1)		(2)		(3)	
Variable	coeff.	zstat	coeff.	zstat	coeff.	zstat
Linguistic Complexity						
Fog(Present)	0.01***	(3.41)	.	.	.	.
Fog(Response)	−0.01***	(−3.04)	.	.	.	.
Linguistic Complexity Decomposed into Latent Components						
Obfu(Present)	.	.	0.02***	(4.02)	0.01***	(2.95)
Obfu(Response)	.	.	0.01	(1.58)	0.01***	(4.07)
Info(Both)	.	.	−0.04***	(−8.56)	−0.03***	(−8.18)
Base Model Controls						
Size	−0.61***	(−73.75)	−0.62***	(−74.54)	−0.63***	(−78.98)
RM	0.03***	(6.27)	0.02***	(4.29)	0.01	(1.19)
Returns	−0.10***	(−28.54)	−0.11***	(−29.50)	−0.10***	(−29.55)
Idiorisk	0.19***	(13.92)	0.19***	(14.11)	0.19***	(14.51)
Coverage	−0.19***	(−29.08)	−0.19***	(−28.55)	−0.16***	(−25.82)
Dispersion	−0.04***	(−10.16)	−0.04***	(−10.06)	−0.05***	(−11.10)
MgmtForecast	−0.01***	(−6.19)	−0.01***	(−5.64)	−0.01***	(−5.56)
Surprise	−0.04***	(−16.96)	−0.04***	(−16.29)	−0.03***	(−14.16)
Loss	0.04***	(13.19)	0.05***	(14.06)	0.04***	(13.30)
SpecItems	0.00	(−0.09)	0.00	(−1.06)	−0.00	(−0.47)
SmallStock	0.01***	(5.92)	0.01***	(6.15)	0.01***	(7.06)
Additional Controls for Characteristics of Presentation and Response Language						
Length(Present)	.	.	.	.	−0.06***	(7.45)
Jargon(Present)	.	.	.	.	0.02***	(2.91)
ForwardLook(Present)	.	.	.	.	−0.02***	(−4.64)
PositiveTone(Present)	.	.	.	.	−0.02***	(−4.22)
NegativeTone(Present)	.	.	.	.	0.06***	(14.06)
Length(Response)	.	.	.	.	−0.07***	(−9.07)
Jargon(Response)	.	.	.	.	0.01***	(2.69)
ForwardLook(Response)	.	.	.	.	−0.01***	(−2.33)
PositiveTone(Response)	.	.	.	.	−0.01	(−1.21)
NegativeTone(Response)	.	.	.	.	0.03***	(7.33)
Nobs / Adj R^2	60,172 / 82.03		60,172 / 82.82		60,172 / 83.48	

Figure: stats

Other test

1. more complex businesses have permanently higher information asymmetry, and that our latent components simply proxy for business complexity that is independent of the disclosure response to complexity
2. examine whether our results are robust to including industry fixed effects firm fixed effects , and both firm and manager fixed effects
3. assess the relative importance of analyst linguistic complexity and the business complexity variables to our empirical estimation strategy by estimating the latent components separately using only analyst complexity and only business complexity.

other test

Also this paper examine the complexity with the earning Candice and Consistent with Li, 2008, we find that loss firms have significantly higher Fog during both the presentation and response

ANd they do experiment with the unique decomposition elements they find.

Mante Carlo Test

TABLE 6

Linguistic Complexity and Information Asymmetry: Monte Carlo Placebo Tests

Panel A: Monte carlo placebo test: Random non-conference call dates			
	Table 4 Column 3 β	Random Non-Call Date $E[\beta]$	Diff (1) − (2) p-value
Variable	(1)	(2)	(3)
Obfu(Present)	0.011***	0.004	[<0.001]
Obfu(Response)	0.013***	0.008	[<0.001]
Infi(Both)	−0.054***	−0.018	[<0.001]
Controls Included	Base Model Controls + Additional Controls for Characteristics of Presentation and Response Language		
Panel B: Monte carlo placebo test: Random conference call dates			
	Table 4 Column 3 β	Random Call Date $E[\beta]$	Diff (1) − (2) p-value
Variable	(1)	(2)	(3)
Obfu(Present)	0.011***	0.010	[0.820]
Obfu(Response)	0.013***	−0.002	[<0.001]
Infi(Both)	−0.054***	−0.015	[<0.001]
Controls Included	Base Model Controls + Additional Controls for Characteristics of Presentation and Response Language		
Panel C: Monte carlo placebo test: Random decomposition			
	Table 4 column 3 β	Random Decomposition $E[\beta]$	Diff (1) − (2) p-value
Variable	(1)	(2)	(3)
Obfu(Present)	0.011***	0.000	[<0.001]
Obfu(Response)	0.013***	0.000	[<0.001]
Infi(Both)	−0.054***	0.000	[<0.001]
Controls Included	Base Model Controls + Additional Controls for Characteristics of Presentation and Response Language		

Figure: stats

Outline

Conclusion

- develop an empirical approach to estimate these two latent components within the context of quarterly earnings conference calls.
- Testify their assumption.
- find a predicted negative (positive) relation between the estimated information (obfuscation) component of linguistic complexity and information asymmetry.

Outline

References I

Bloomfield, Robert (2008). “Discussion of “annual report readability, current earnings, and earnings persistence””. *Journal of Accounting and Economics* 45.2-3, pp. 248–252.

Li, Feng (2008). “Annual report readability, current earnings, and earnings persistence”. *Journal of Accounting and Economics* 45.2. Economic Consequences of Alternative Accounting Standards and Regulation, pp. 221–247. ISSN: 0165-4101. DOI: https://doi.org/10.1016/j.jacceco.2008.02.003.