



Indian Institute of Management Calcutta

Working Paper Series

WPS No 870 /November 2021

Readability (and beyond) of Financial Disclosures and Earnings Management

Samit Paul*

Finance & Control Group

Indian Institute of Management Calcutta

Email: samit@iimcal.ac.in

*Corresponding Author

Indian Institute of Management Calcutta Joka, D.H. Road Kolkata 700104

URL: <http://facultylive.iimcal.ac.in/workingpapers>

Readability (and beyond) of Financial Disclosures & Earnings Management¹

Samit Paul²

Abstract

The obfuscation is a multidimensional concept. To investigate the same, we explore the association between earnings management and different linguistic features associated with management discussion and analysis (MD&A) section of 10-Ks filed by 10,940 US firms for 67,649 firm-years. Our industry and year fixed-effect regression analysis followed by fixed-effect quantile regression suggest that earnings management significantly influences not only the readability of MD&As but also the lexical diversity, sentiment and uncertainty of textual disclosures. This finding is consistent for both accrual and real activity-based earnings management.

Keywords: Obfuscation, Readability, Lexical Diversity, Sentiment, Uncertainty.

JEL Codes: M1; M4; G3.

¹ Author acknowledges research grant received under category-I project from IIM Calcutta (RP: RFDE/3765/2018-19).

² Finance & Control Group, Indian Institute of Management Calcutta, Email : samit@iimcal.ac.in

Readability (and beyond) of Financial Disclosures & Earnings Management

1. Introduction

Evidence from prior literature suggests that managers mask adverse information by obfuscating financial disclosures. Li (2008) shows that greater complexity and length of financial reports are associated with lower level of firm earnings and its persistence. This is commensurate with the incomplete revelation hypothesis of Bloomfield (2002), which states that the complexity and length of disclosures make the extraction of relevant information costly for investors. In similar line, Bao et al (2019) reports that managers often withhold poor performance when they are incentivized to support the stock price. Hence, it's of no surprise that firms that miss the benchmark of prior year's earnings by a small margin are often motivated to manage the earnings within accounting discretions to achieve or surpass the benchmark. In fact, Lo, Ramos and Rogo (2017) evinces that firms that manage earnings to meet or just beat the prior year's earnings, often obfuscate the information in management discussion and analysis (MD&A) part of annual report by making it difficult to read.

We argue that the obfuscation is a multidimensional concept. Therefore, readability of financial reports may not be the only indicator of obfuscation. Managers may alter other linguistic features of the textual disclosure to obfuscate the information as well. For example, prior researches document that short-term stock price reaction is positively associated with language sentiment or tone of disclosures (Huang et al., 2014). Similarly, Loughran and McDonald (2013) concludes that usage of uncertain, negative and weak modal words enhances volatility and first-day IPO returns. Even, lexical diversity in text can be an indicator of precision of information

(Bozanic & Thevenot, 2015). Thus, it's imperative to look beyond readability and explore other textual attributes as potential outcomes of earnings management.

We investigate the association of earnings management with different linguistic features present in MD&A section of 10-Ks filed with the Security Exchange Commission (SEC) of the US for the period of 1993-2020. More specifically, we explore how firms that manage accrual based and real activity-based earnings to achieve or just surpass prior period earnings alter the readability, lexical diversity, sentiment and uncertainty of textual disclosures. Our findings suggest that managers obfuscate information not only by changing readability of disclosures but also by altering lexical diversity, sentiment and uncertainty of MD&A texts. This finding is robust across different quantiles as well. Our study acts as a complement to Lo et al (2017) by delving deeper into this less explored strand of literature and contribute to the existing knowledge-base.

The rest of the study is structured as follows: section 2 develops relevant hypotheses; section 3 describes research design; section 4 discusses empirical findings; and finally, section 5 concludes the study.

2. Hypotheses Development

Although Li (2008) reports an overall negative relation between earnings performance and readability using *Fog* and *Length* of MD&A, it may vary in the context of earnings management. If the management tries to deceive the investors or analysts with biased choice of increased earnings, it will make the cost of extracting information very high (Lo et al., 2017). At the same time, untruthful representation is a difficult task as there involves a chance of contradicting own statement some point later (Hancock et al., 2007). Therefore, we expect that firms that manage

earnings to achieve or just surpass previous year's earnings will disclose shorter but complex MD&As. Hence, we posit the following hypothesis:

H1: Firm-years with (i) zero or very small increase in earnings, and (ii) income-increasing and high income-increasing discretionary accruals or real activity-based earnings management tend to disclose complex and shorter MD&As

If a firm reports similar or close to the previous year's earnings, it has nothing much to explain. Further, if it wants to appear trustworthy without disclosing substantive details it would exhibit lower level of lexical diversity (Humpherys et al., 2011). Therefore, we postulate the following hypothesis:

H2: Firm-years with (i) zero or very small increase in earnings, and (ii) income-increasing and high income-increasing discretionary accruals or real activity-based earnings management tend to disclose MD&As with less lexical diversity

Firms engaging in any fraudulent act of hiding losses or just meeting expectations may willingly illustrate a fake image of success. This can be achieved by reducing usage of negative words and including more of positive words. Although tracking a count of negative words is intuitive and straightforward, any inference based on usage of positive words may be incorrect as negative phrases are often wrapped in positive words (Loughran & McDonald, 2011). Thus, we build our next hypothesis as:

H3: Firm-years with (i) zero or very small increase in earnings, and (ii) income-increasing and high income-increasing discretionary accruals or real activity-based earnings management tend to disclose MD&As with less negative and more/less positive words

Usage of strong modal words in MD&As refers to the tendency of firms to display their overconfidence in investors' mind (Loughran & McDonald, 2013). On the other hand, inclusion of weak modal or uncertain words indicate about firms' uncertain future. We conjecture that firms that manage earnings to achieve or surpass prior year's earnings, would use more of strong modal words and less of weak modal or uncertain words. Hence, we frame our last hypothesis as:

H4: Firm-years with (i) zero or very small increase in earnings, and (ii) income-increasing and high income-increasing discretionary accruals or real activity-based earnings management tend to disclose MD&As with higher strong modal and less weak modal and uncertain words

3. Research Design

3.1 Data

We source all annual MD&A reports for the period of 1993 to 2020 from Electronic Data Gathering, Analysis and Retrieval (EDGAR) database maintained by the SEC. We exclude all firms belonging to utilities and financial services from our sample due to their unique financial and operating structures (Lo et al., 2017). To test our hypotheses, we extract all relevant data from

COMPUSTAT using Central Index Key (CIK) as documented in EDGAR filings. We further drop few firm-year observations based on availability of data required to construct necessary variables. Finally, our sample comprises 67,649 firm-year observations pertaining to 10,940 unique firms. We use word lists of Loughran and McDonald (2011) to gauge the tone of disclosures.

3.2 Methodology

To test our hypotheses, we run four regression model specifications while controlling for industry and year fixed effects. These are:

$$Readability = \beta_0 + \beta_1 EM + \sum \beta_j Control_j + \varepsilon \quad (1)$$

$$Diversification = \beta_0 + \beta_1 EM + \sum \beta_j Control_j + \varepsilon \quad (2)$$

$$Sentiment = \beta_0 + \beta_1 EM + \sum \beta_j Control_j + \varepsilon \quad (3)$$

$$Uncertainty = \beta_0 + \beta_1 EM + \sum \beta_j Control_j + \varepsilon \quad (4)$$

In these models, the dependent variable *Readability* is measured by *FOG* and *MD&A Length* (Li, 2008). *FOG* has been criticized as “poorly specified” in financial applications (Loughran & McDonald, 2014). Hence, we also use *MD&A Length* as a measure of readability. *Diversification* is proxied by *Lexical Diversity* (Humpherys et al., 2011). *Sentiment* is represented by percentage of *Positive* and *Negative* words (Loughran & McDonald, 2013). Finally, *Uncertainty* is measured by percentage of *Strong Modal*, *Weak Modal* and *Uncertain* words (Loughran & McDonald, 2013). The definitions of these linguistic measures with examples are listed in Appendix A.1.

The main explanatory variable of the study is *EM* which is the proxy for the earnings management. We compute EM in multiple ways. Our first EM measure is a dummy variable with

value 1 where change of earnings per share (ΔEPS) is marginally “small positive” (within [$\$0$, $\$0.01$ to $\$0.03$]), and 0 otherwise. The second and third measures are based on accrual-based earnings management where the discretionary accruals (DA) are computed using modified-Jones model (Dechow et al., 1995), and Raman and Shahrur (2008) model.³ More specifically, our second measure is *PosEM(DA)* dummy which is 1 for the firm-years when there is a positive DA, and 0 otherwise. The third measure, *HighPosEM(DA)*, carries value 1 for the firm-years where the positive DA is higher than the median value, and 0 otherwise. To increase the power of the test, we interact these two measures with ΔEPS (Lo et al., 2017). For measuring EM based on real activities (RA), we express discretionary expenses as linear form of lagged sales (Roychowdhury, 2006).⁴ Similar to our earlier two measures, we compute fourth and fifth EM proxies as *PosEM(RA)* and *HighPosEM(RA)* dummies and interact them with ΔEPS .

In our regression equations, we include 12 control variables that were also used in earlier studies (e.g., Li, 2008; Lo et al., 2017). The descriptions of these variables and their basic descriptive statistics are provided in Appendix A.3.

4. Empirical Findings

Table 1 presents the descriptive statistics of linguistic features. The mean and median values of these variables are closely comparable which suggest that our sample observations are free from outliers. The standard deviation of *FOG* is highest although it's more or less in line with the standard deviation of *FOG* reported by Lo *et al* (2017). Table 2 lists the correlations among these

³ For brevity, we report results of only Raman and Shahrur (2008) model. However, results of modified Jones are largely similar. We provide a description of both models in Appendix A.2

⁴ The detailed description of the model is also provided in Appendix A.2

variables. Except the correlation between *FOG* and *MD&A Length*, all other correlation coefficients are significant. Only correlation between *Strong Modal* and *MD&A Length* is negative. This indicates that larger MD&As use relatively a smaller proportion of strong modal words. All other correlations are positive suggesting that when length of MD&A increases, the diversity as well as usage of positive, negative, weak modal and uncertain words increase.

To explore how these linguistic features vary year wise we plot their respective scores separately for firms that manage earnings and for those which don't manage, and present them in Figure 1. The figure suggests that across years firms that manage earnings disclose MD&As with consistently higher *FOG*, shorter *MD&A Length*, and lower *Lexical Diversity* compared to the firms that don't manage earnings. Interestingly, MD&As of these suspect firms started using higher proportion of *Strong Modal*, *Weak Modal*, and *Uncertain* words after 2007.

Table 3 reports the impact of earnings management on the above-mentioned linguistic features of MD&A disclosures. For brevity of space, we list only the coefficients of explanatory variables and not that of control variables. The number of observations for accrual-based earnings management is little less due to sacrifice of few additional data points in the process of constructing the DA measure. The results of the table suggest that the association between *FOG* and "small positive" ΔEPS is positive and highly significant. This association is consistent when we interact *PosEM(DA)*, *HighPosEM(DA)*, *PosEM(RA)* and *HighPosEM(RA)* with ΔEPS and use these combinations to explain *FOG*. This indicates that firms that achieve or just surpass the prior year EPS and perform accrual or real activity-based earnings management make MD&As more

complex. This finding is in similar line with Lo *et al* (2017). Interestingly, when we use *MD&A Length* as a proxy of readability, the relationship becomes negative and significant. This suggests that firms managing earnings disclose shorter MD&As. This differs from the finding of Lo *et al* (2017), which documents a negative but insignificant association between size of MD&A and earnings management. Thus, we find support for our first hypothesis *H1* which suggests that suspect firms disclose less information in MD&As but in complex ways. We also find a negative and significant relationship between earnings management and *Lexical Diversity* scores. This shows that firms that manage earnings dilute the diversity of language by using less unique words in MD&As. This finding is in line with Humpherys *et al* (2011) and corroborates our hypothesis *H2*. Next, we find a significant and negative relationship between earnings management and usage of *Negative* and *Positive* words. However, the impact on usage of negative words is larger than that of positive words. This asymmetric response is well documented in literature (e.g., Epstein & Schneider, 2008). Thus, we find support for our hypothesis *H3* as well. While exploring uncertainty, we find that impact of earnings management on the usage of *Strong Modal* words is positive and significant. But there is no significant relationship between usage of *Weak Modal* words and EM. Interestingly, the association between EM and usage of *Uncertain* words is positive although inconsistent. Hence, our hypothesis *H4* is only partially supported. Overall, it's quite evident that earnings management significantly influences not only the readability of MD&As but also the lexical diversity, sentiment and uncertainty of textual disclosures. These findings are consistent for both accrual and real activity-based earnings management.

To examine whether our results are influenced by outliers, we perform a robustness check by employing fixed-effect quantile regression technique in a panel data framework (Tripathi & Dixit, 2021). Specifically, we estimate the association between earnings management and all eight linguistic features under different quantiles [0.05, 0.95] and plot the coefficients of estimates in Figure 2. The figure suggests that the *MD&A Length* and *Lexical Diversity* experience strong and statistically significant negative relationship with earnings management in both upper and lower quantiles whereas *FOG* and *Strong Modal* bears a similarly strong positive association. Thus, it further supports our earlier argument that managers obfuscate information through multiple attributes of textual disclosures.

5. Conclusion

We argue in favor of multidimensionality of obfuscation. To investigate the same, we first identify firms that manage earnings to achieve or just surpass the prior year's earnings within the period of 1993 to 2020. Next, we compute different linguistic features such as readability, lexical diversity, sentiment and uncertainty associated with MD&A section of 10-Ks filed in these firm-years. Following Lo et al (2017), we apply industry and year fixed regression technique to establish the linkage between earnings management and those linguistic features. Further, we perform a robustness check with fixed-effect quantile regression under panel data framework. Overall, our findings suggest that earnings management significantly influences not only the readability of MD&As but also the lexical diversity, sentiment and uncertainty of textual disclosures. These findings are consistent for both accrual and real activity-based earnings management.

Future research can focus on different other benchmarks, e.g. analyst's forecast, and its impact on linguistic features of MD&As. Further, investigating these linguistic features in the context of earnings management in other countries across the globe will be of much interest.

References:

- Bao, D., Kim, Y., Mian, G. M., & Su, L. (2019). Do managers disclose or withhold bad news? Evidence from short interest. *Accounting Review*, *94*(3), 1–26.
- Bloomfield, R. J. (2002). The incomplete revelation hypothesis and financial reporting. *Accounting Horizons*, *16* (3), 233–244
- Bozanic, Z., & Thevenot, M. (2015). Qualitative disclosure and changes in Sell-Side financial analysts' information environment. *Contemporary Accounting Research*, *32*(4), 1595–1616.
- Dechow, P. M., Sloan, R. G., & Sweeney, A. P. (1995). Detecting earnings management. *Accounting Review*, 193–225.
- Epstein, L. G., & Schneider, M. (2008). Ambiguity, information quality, and asset pricing. *Journal of Finance*, *63*(1), 197–228.
- Hancock, J. T., Curry, L. E., Goorha, S., & Woodworth, M. (2007). On lying and being lied to: A linguistic analysis of deception in computer-mediated communication. *Discourse Processes*, *45*(1), 1–23.
- Huang, X., Teoh, S. H., & Zhang, Y. (2014). Tone management. *Accounting Review*, *89*(3), 1083–1113.
- Humpherys, S. L., Moffitt, K. C., Burns, M. B., Burgoon, J. K., & Felix, W. F. (2011). Identification of fraudulent financial statements using linguistic credibility analysis. *Decision Support Systems*, *50*(3), 585–594.

- Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics*, 45(2–3), 221–247.
- Lo, K., Ramos, F., & Rogo, R. (2017). Earnings management and annual report readability. *Journal of Accounting and Economics*, 63(1), 1–25.
<https://doi.org/10.1016/j.jacceco.2016.09.002>
- Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance*, 66(1), 35–65.
- Loughran, T., & McDonald, B. (2013). IPO first-day returns, offer price revisions, volatility, and form S-1 language. *Journal of Financial Economics*, 109(2), 307–326.
- Loughran, T., & McDonald, B. (2014). Measuring readability in financial disclosures. *Journal of Finance*, 69(4), 1643–1671.
- Raman, K., & Shahrur, H. (2008). Relationship-specific investments and earnings management: Evidence on corporate suppliers and customers. *Accounting Review*, 83(4), 1041–1081.
- Roychowdhury, S. (2006). Earnings management through real activities manipulation. *Journal of Accounting and Economics*, 42(3), 335–370.
- Tripathi, A., & Dixit, A. (2021). Liquidity commonality in extreme quantiles: Indian evidence. *Finance Research Letters*, 38, 101448.

Table 1: Descriptive Statistics

	Observations (firm-year)	Mean	Median	Standard Deviation
<i>Linguistic Features</i>				
FOG	67,649	15.584	15.534	2.118
MD&A Length	67,649	8.389	8.475	0.814
Lexical Diversity	67,649	10.051	10.139	1.354
Positive	67,649	0.887	0.832	0.381
Negative	67,649	1.834	1.715	0.851
Strong Modal	67,649	0.375	0.329	0.230
Weak Modal	67,649	0.517	0.420	0.394
Uncertain	67,649	1.695	1.592	0.724

Table 2: Correlation between linguistic features of MD&A disclosures

	FOG	MD&A Length	Lexical Diversit y	Negativ e	Positiv e	Weak Modal	Strong Modal	Uncertai n
FOG	1							
MD&A Length	0.004	1						
Lexical Diversity	0.164** *	0.611** *	1					
Negative	0.112** *	0.226** *	0.283***	1				
Positive	0.035** *	0.075** *	0.261***	0.133***	1			
Weak Modal	0.265** *	0.143** *	0.351***	0.368***	0.199** *	1		
Strong Modal	0.259** *	0.067** *	0.224***	0.114***	0.210** *	0.379** *	1	
Uncertain	0.248** *	0.076** *	0.293***	0.269***	0.131** *	0.691** *	0.325** *	1

*, **, *** indicate significance at the 10%, 5%, and 1% levels.

Table 3 (Continued.): The effect of earnings management on linguistic features of MD&A disclosures

Dependent Variable	Estimated coefficient of the “Earnings Management” indicator variable (<i>EM</i>); <i>EM</i> = 1 when					
	<i>PosEM(RA) = 1 and ΔEPS ∈</i>			<i>HighPosEM(RA) = 1 and ΔEPS ∈</i>		
	[\$0, \$0.01]	[\$0, \$0.02]	[\$0, \$0.03]	[\$0, \$0.01]	[\$0, \$0.02]	[\$0, \$0.03]
FOG	0.223*** (4.610)	0.166*** (4.275)	0.174*** (3.437)	0.174*** (3.437)	0.154*** (3.476)	0.120*** (2.955)
MD&A Length	-0.148*** (-9.169)	-0.139*** (-9.871)	-0.129*** (-9.975)	-0.158*** (-9.406)	-0.143*** (-9.684)	-0.132*** (-9.801)
Lexical Diversity	-0.151*** (-4.533)	-0.136*** (-4.668)	-0.162*** (-6.057)	-0.176*** (-5.065)	-0.152*** (-4.984)	-0.175*** (-6.282)
Positive Words	-0.028*** (-3.049)	-0.028*** (-3.532)	-0.031*** (-4.269)	-0.022** (-2.319)	-0.023*** (-2.776)	-0.028*** (-3.713)
Negative Words	-0.068*** (-3.466)	-0.082*** (-4.762)	-0.092*** (-5.831)	-0.068*** (-3.304)	-0.085*** (-4.712)	-0.096*** (-5.857)
Strong Modal	0.023*** (4.304)	0.014*** (3.067)	0.009** (2.201)	0.023*** (4.136)	0.014*** (2.854)	0.008* (1.853)
Weak Modal	0.009 (0.886)	0.002 (0.255)	-0.001 (-0.110)	0.012 (1.150)	0.004 (0.485)	0.000 (0.006)
Uncertain Words	0.038** (2.204)	0.028* (1.853)	0.021 (1.522)	0.038** (2.059)	0.027* (1.686)	0.020 (1.356)
Observations with <i>EM</i> = 1	624	862	1,065	417	553	670
Total Observations	67,649	67,649	67,649	67,649	67,649	67,649

Notes: All models are estimated with the full set of controls, and industry & year fixed effects. *t*-statistics are reported in parenthesis below the regression coefficients. *, **, *** indicate significance at the 10%, 5%, and 1% levels.

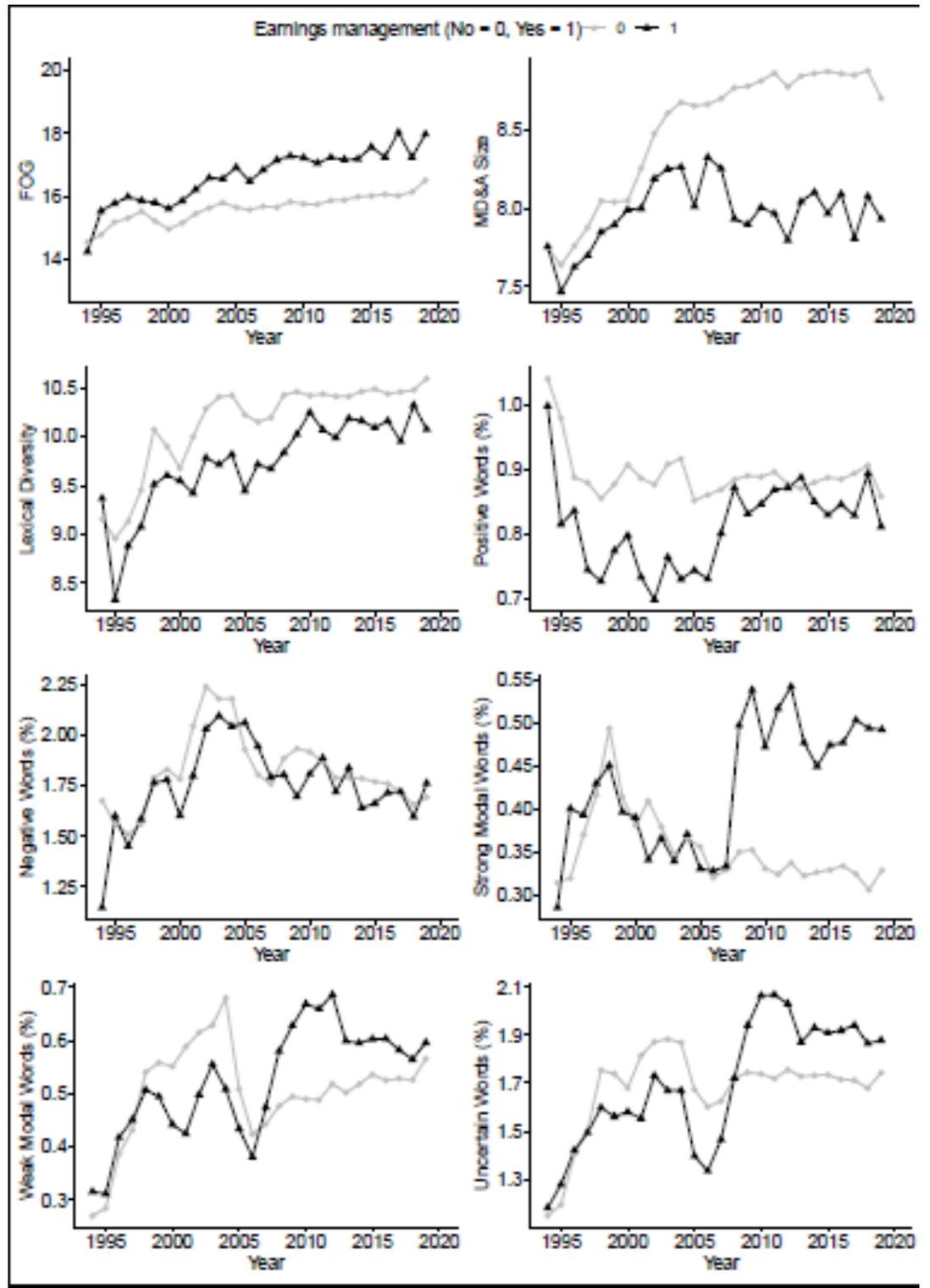


Figure 1: Plot of Linguistic Scores

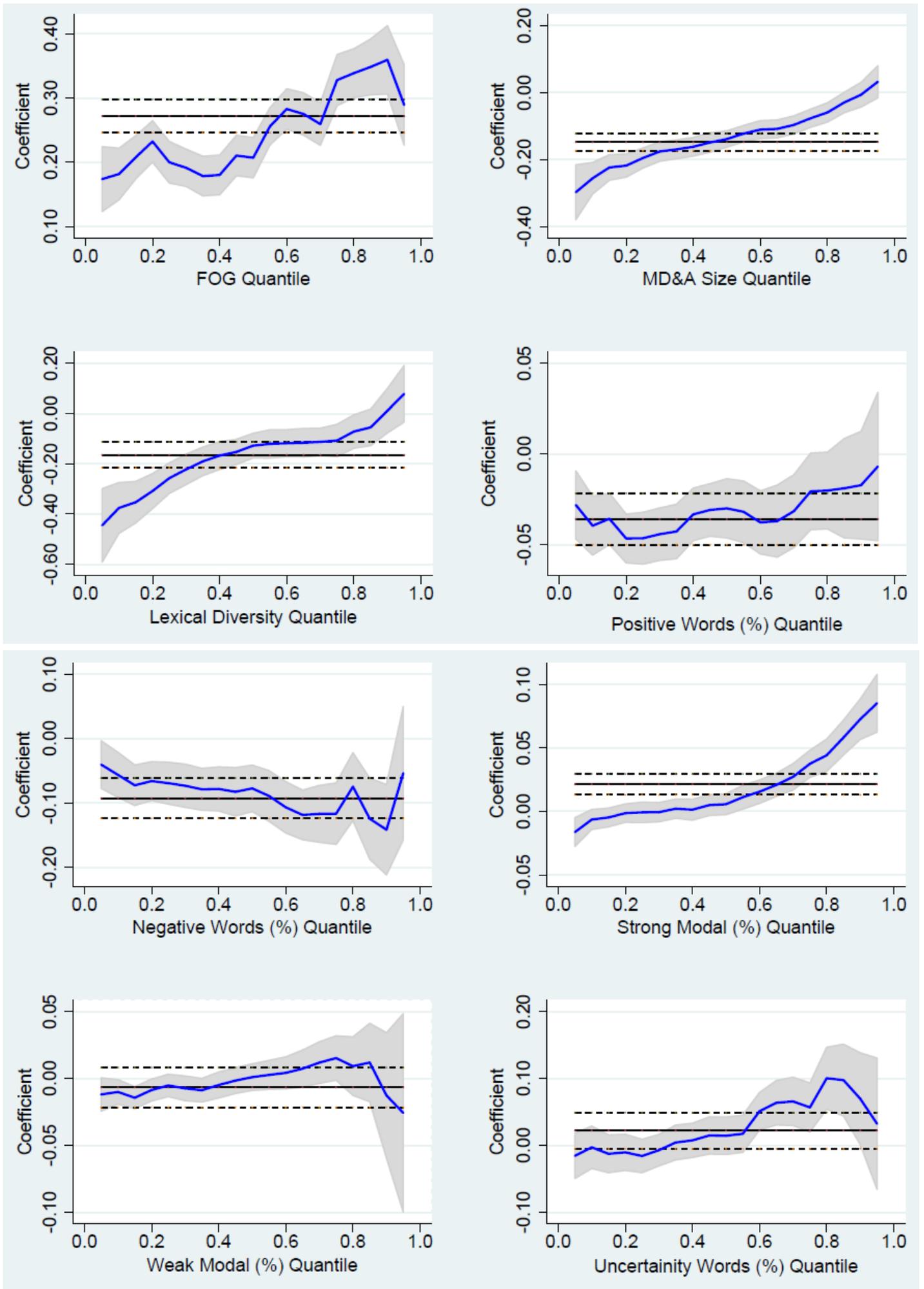


Figure 2: Plot of Quantile Regressions

Appendix

A.1 Definitions and examples of linguistic features

Linguistic features	Definitions	Examples
FOG	FOG = 0.4 * (words per sentence + percent of complex words)	
MD&A Length	Logarithm of total words	
Lexical Diversity	Number of unique words / ($\sqrt{2} * \sqrt{\text{Total words}}$)	
Positive	Percentage of positive words out of total words	'success', 'strong', 'good', 'benefit' etc
Negative	Percentage of negative words out of total words	'failure', 'decline', 'loss', 'difficult' etc
Strong Modal	Percentage of strong modal words out of total words	'must', 'always', 'never', 'will' etc
Weak Modal	Percentage of weak modal words out of total words	'could', 'possibly', 'may', 'nearly' etc
Uncertain	Percentage of uncertain words out of total words	'approximate', 'believe', 'uncertain', 'pending' etc

A.2 Models for measuring earnings management

A.2.1 Accrual based earnings management

In accrual-based earnings management, discretionary accruals (DA) are computed as residuals of following two regression specifications:

Modified Jones model (Dechow et al., 1995):

$$TotAcc_t/TA_{t-1} = \alpha_0 + \alpha_1(1/TA_{t-1}) + \alpha_2((\Delta Rev_t - \Delta Rec_t)/TA_{t-1}) + \alpha_3(PPE_t/TA_{t-1}) + \varepsilon_t \quad (A.1)$$

Raman and Shahrur (2008) model:

$$TotAcc_t/TA_{t-1} = \alpha_0 + \alpha_1(1/TA_{t-1}) + \alpha_2((\Delta Rev_t - \Delta Rec_t)/TA_{t-1}) + \alpha_3(PPE_t/TA_{t-1}) + \alpha_4(ROA_t) + \alpha_5(MTB_t) + \varepsilon_t \quad (A.2)$$

TotAcc_t = total operating accruals in year *t*

TA_{t-1} = total assets at year *t* - 1

ΔRev_t = change in revenue from year *t*-1 to *t*

ΔRec_t = change in receivables from year *t*-1 to *t*

PPE_t = gross property, plant and equipment during year *t*

ROA_t = net income before extra-ordinary items at year *t* scaled by lagged total assets

PPE_t = market value of equity scaled by book value of equity in year *t*

A.2.2 Real activity-based earnings management

For real activity-based earnings management, we estimate abnormal discretionary expenses as suggested by Roychowdhury (2006). It is computed from following equation:

$$DISEXP_t/TA_{t-1} = \alpha_0 + \alpha_1(1/TA_{t-1}) + \alpha_2(Rev_{t-1}/TA_{t-1}) + \varepsilon_t \quad (A.3)$$

$DISEXP_t$ = discretionary expenses (R&D, advertising, and SG&A expenses) in year t

TA_{t-1} = total assets at year $t - 1$

Rev_{t-1} = revenue at year $t-1$

All model specifications listed under Eqs A.1 - A.3 are estimated in cross-section of year and industry with at least 15 observations.

A.3 Description and descriptive statistics of control variables

	Description	Observations (firm-year)	Mean	Median	Standard Deviation
Negative Earnings Change	1 if $\Delta EPS < 1$; 0 otherwise	67,649	0.444	0.000	0.002
Earnings	Operating earnings scaled by total assets	67,649	-0.297	0.010	0.005
Loss	1 if Earnings < 0 ; 0 otherwise	67,649	0.418	0.000	0.002
Size	Logarithm of market value of equity	67,649	5.223	5.323	0.009
Special Items	Amount of special items scaled by total assets	67,649	-0.500	0.000	0.248
Leverage	Total debt scaled to total assets	67,649	0.309	0.179	0.002
Earnings volatility	Standard deviation of operating earnings in last five years	67,624	4.105	7.822	0.428
Return volatility	Standard deviation of monthly returns in last year	67,358	0.178	0.137	0.001
Age	Number of years of firms since appearing in Compustat	67,649	2.403	2.397	0.003
M&A	1 if the firm is an acquirer in a given year; 0 otherwise	67,649	0.292	0.000	0.002
Delaware	1 if the firm is incorporated in Delaware; 0 otherwise	67,649	0.004	0.000	0.000
International	1 if the firm is present internationally; 0 otherwise	67,649	0.062	0.000	0.001