Measuring non-GAAP Earnings Quality Using Qualitative Disclosure: A Machine-learning Approach

Jack Wang

May 9, 2022

Abstract

I use machine learning methods to examine whether the qualitative information contained in 8-K earnings press releases is useful to assess firms' non-GAAP earnings quality. I train my model with 8-K earnings press release textual information and firm specific earnings response coefficient (ERC) as the measure of non-GAAP earnings quality . The trained model can then be used to generate a non-GAAP earnings quality score (the Score) for each disclosing firm based on the qualitative information. In my out-of-sample test, I find that the higher value of the Score is positively associated with higher firm specific ERC.

1 Introduction

One of the most important implication of accounting earnings is for evaluating firm's performance. However, the accounting earnings calculated based on the General Accepted Accounting Principles (GAAP) is argued to be less informative for evaluation purpose because it contains transitory items that are not relevant with firm's core performance as oppose to the non-GAAP measure of earnings, from which managers exclude the impacts of transitory items on accounting earnings. In practice, managers will not only exclude transitory items but also exclude recurring items from GAAP earnings, because there is not a clear rule to restrict what items can be excluded. Managers have been found to use non-GAAP earnings to both inform and mislead investors ([BL03]; [CW14]; [LM04]; [Bar12]; [BC09]). While tons of research has been done to examine the determinants and consequences of non-GAAP earnings and exclusions, the extant research mainly focuses on quantitative aspects of non-GAAP earnings. Qualitative disclosure, which comprises the majority of financial disclosure, can also provide useful and relevant information to stakeholders. Prior research has documented that textual features conveyed in firm qualitative disclosures (e.g. length, tone, readability, word usage) can provide incremental information about managers' discretion and firm performance and are associated with future stock market return ([Li10]; [DS12]; [LM16]). Therefore, in this paper, I am going to examine whether the qualitative information contained in non-GAAP earnings press releases can be used to predict the quality of non-GAAP earnings.

ERC is market's reaction to firm's earnings. In accounting research, ERC can be used to assess the quality of non-GAAP earnings ([CK89]; [BS02]). However, it is not a perfect measure of earnings quality, because it only captures market's reaction to the quantitative earnings numbers and firm's information environment also affects the effectiveness of using ERC to proxy earnings quality. In a recent research, [CY21] introduced a measure of non-GAAP earnings quality based on qualitative characteristics (narrative information and presentation choices) of firm earnings press release. They hand-collect earnings press release with non-GAAP disclosures for S&P 500 firms and assign score to each observation based on researchers judgements on whether each disclosure meet certain disclosure requirements. However,relying on researchers judgements may introduce unwanted bias and measurement inconsistency for their disclosure quality scores and hand-collecting data limits their sample size. Moreover, the criteria they use to generate the scores may not capture the whole useful qualitative information in press releases.

The advanced computing technology enables me to tackle these issues with more efficient and reliable methods. The purpose of this paper is to use machine learning technique to extract qualitative information from non-GAAP earnings press releases to evaluate firm's non-GAAP earnings quality. I scrape all the earnings press releases with non-GAAP disclosures from SEC EDGAR using Python. Then I use "tidytext" package in R to do the preprocessing (i.e. removing stop words and numbers, stemming) of the textual information, which returns a corpus in document-term matrix (DTM) format. This DTM contains all the qualitative information in earnings press releases. To train the model to capture the information in textual disclosure, I use supervised Latent Dirichlet Allocation (sLDA) method with the DTM as input and the firm specific ERC as output. In the out-of-sample test, the non-GAAP earnings quality scores generated by my trained model based on the provided earnings press release have a significant association with the firm specific ERC, which indicates that the Scores capture non-GAAP earnings quality as measured by ERC.

This research may contribute to the literature from several aspects. First, it provide additional evidence that qualitative information of non-GAAP disclosures contains information that can be used to evaluate non-GAAP earnings quality. While prior research mainly focuses on the quantitative attributes of non-GAAP disclosures, the information usefulness of qualitative non-GAAP disclosure is left under studied. Second, compared to the research that uses manually processed data, my research, using machine learning methods to process all the textual information, is less likely to be affected by researchers' subjectivity. Lastly, using machine learning methods also enables the analysis to incorporate more comprehensive information in the qualitative disclosure.

2 Literature Review

2.1 Qualitative disclosure and earnings quality

Earnings quality represent the extent to which current earnings can be informative about firm's future performance. Non-GAAP earnings are found more informative than GAAP earnings because managers exclude transitory items from GAAP earnings in calculating non-GAAP earnings to provide a more accurate metrics of core performance ([BL03]; [CW14]). However, in the process of arriving at non-GAAP earnings numbers, managers have incentives to exclude less transitory items to report higher earnings as well. Managers may manipulate the narrative information and presentation of qualitative information disclosed with non-GAAP earnings in press releases to inform or mislead investors. [CY21] find that when their non-GAAP disclosure score calculated using qualitative information in press releases indicates lower disclosure transparency, the corresponding non-GAAP earnings are less informative of firm's future performance. The qualitative characteristics of non-GAAP disclosures convey useful information that helps investors distinguish between aggressive and informative non-GAAP reporting.

Early research on GAAP earnings quality also find that textual characteristics convey information about earnings quality. [Li10] documents that when the management discussions in 10-K filings are complex and difficult to understand, managers might be obfuscating information and the reported financial data are therefore likely to have lower quality. [HZ14] develop a model to estimate the abnormal positive tone in firm's disclosure. Their results indicate that managers manipulate the tone in qualitative disclosure to strategically manipulate investors' perception rather than providing private information.

2.2 Using machine learning technique to capture information in qualitative disclosures

Most prior research on firm financial disclosures using textual analysis rely on ex ante identified dictionaries to identify words and phrases. However, the ability of using these dictionaries to capture relevant words and phrases is limited due to the scope of these dictionaries and the suitableness for specific research questions. Compared to dictionary method, supervised machine learning method automates the identification of language that explain a variable of interest ([FL16]). I use a topic based modeling approach (LDA) first introduced by [BJ03] to capture the qualitative information in press releases. Formally, a topic is a probability distribution over terms in a vocabulary. A topic represents an underlying semantic theme. The LDA model is unsupervised, which only includes the terms in the documents. The goal is to infer topics that maximize the likelihood (or the posterior probability) of the collection. Unsupervised LDA has previously been used to construct features for classification. The hope was that LDA topics would turn out to be useful for categorization, since they act to reduce data dimension. It provides useful descriptive statistics for a collection, which facilitates tasks like browsing, searching, and assessing document similarity. However, when the goal is prediction, fitting unsupervised topics may not be a good choice ([BJ07]). In my case, it is used to help determine the optimal number of topics for sLDA.

The sLDA is supervised method, where each document is paired with a output. The goal is to infer latent topics predictive of the output, which in my case is the firm specific ERC. The sLDA technique is widely used in the linguistic related research in multiple areas to identify the thematic structure of collections of textual disclosures (the topics). Given an unlabeled document, the trained model can infer its topic structure using a fitted model, then form its prediction, which is similar as how we do OLS estimation and prediction.

Prior research has shown that LDA/sLDA method is useful in accounting research on extracting qualitative information from disclosures. [BE20] apply LDA algorithm to 10-K narratives and find that their algorithm produces a valid set of semantically meaningful topics that predict fianncial misreporting. In another research, [DL21] use multiple supervised machine learning methods to capture credit risk relevant qualitative information directly from 10-K and earnings conference calls and find that their measure improves the ability to predict credit events relative to previously developed credit risk measures.

2.3 Non-GAAP earnings quality and ERC

In early accounting literature, when non-GAAP earnings have not gained prevalence, ERC is used to measure the information content of GAAP earnings. Colins and Kothari (1989) find that ERC is function of riskless interest rate, riskness, growth and persistence of earnings. [BS02] is the first to use ERCs to compare quality of non-GAAP earnings and GAAP earnings. They find that ERC for firm's non-GAAP earnings are significantly higher than their GAAP earnings, which is consistent with the argument that ERC represent earnings quality and non-GAAP earnings on average have higher quality than GAAP earnings, because non-GAAP earnings exclude transitory items and provide information about firm's core performance which is more persistent than GAAP earnings. [KM08] argue that non-GAAP earnings are of higher quality when they are associated with firm's future performance.

3 Data

The final training and testing data will contain firm specific ERC as the dependent variable and the counts of words in earnings press releases as the independent variables.

3.1 Data for calculating ERC

ERC is the coefficient on the unexpected non-GAAP earnings when we regress accumulated abnormal return (CAR) of the accounting period on the unexpected non-GAAP earnings. To calculate CAR, I use daily stock return and market return data from CRSP dataset. For unexpected earnings calculation, the real non-GAAP earnings data is from [BW18] and the expected non-GAAP earnings are from I/B/E/S analysts forecast consensus. Since the sample period of updated [BW18] dataset starts from 2003 and I require data of 20 quarters before reporting periods to calculate ERC, my sample starts from fiscal year 2008. After restricting observations to have enough data to calculate quarterly firm specific ERC, I got firm specific ERCs for 27791 firm-quarter observations. Firms usually provide more detailed earnings press releases for their annual earnings with which the forth quarter earnings are reported. Therefore, I only use firm specific ERCs of the forth quarter (the same rule applies to earnings press release data) for model training, which makes my sample to only includes 6637 firm-year ERCs.

3.2 Earnings Press Releases Data

The textual information that is used as the input to the model is from the earnings press releases that include non-GAAP disclosure. Public firms are required to file earnings press releases every quarter. In earnings press releases firms discuss their operations and performance in that period. If firms disclose non-GAAP earnings for that period, they need to show the reconciliation between their GAAP earnings and non-GAAP earnings (the direct discussions) and firms usually provide some other discussions in press releases to rationalize their non-GAAP numbers (the other discussions). The direct- and other discussions of non-GAAP earnings contain information to interpret managers' intentions to be informative or misleading. I use python to scrape the original earnings press releases for the forth quarter, as discussed above, from SEC EDGAR website. The whole textual information of each earnings press release is then put into one cell of the earnings press releases column of the corresponding firm-year observation in the data frame. In total, I got over 45,000 firm-year observations. After matching earnings press releases data with ERC data, I eventually got 6483 firm-year observations for analysis. However, the file size of my original dataset is so large for my computer to process. I randomly select 2000 firm-year observations from the original dataset for LDA and sLDA modeling and provide descriptive findings based on the whole original sample in section 5.

The ”PressRelease” column is the vector contains all the earnings press release textual information. I preprocess the selected sample using ”tidytext” package in R before they can be used to train model. All the words are converted to lower case and are stemmed to the common root. All the stop words and numbers are removed.

4 Methods

The training process of sLDA requires input and output. In my case, the input is the DTM generated based on earnings press releases and the output is the firm specific ERC, which represent the non-GAAP earnings quality.

4.1 To Calculate ERC

$$CAR_i = \alpha_i + \beta_i UE + \epsilon_i \quad (1)$$

$$Return_{i,t} = \gamma_i + \delta_t Return_{mkt} + \nu_i \quad (2)$$

Following prior research ([CK89]; [TW96]) I use the above model (equation (1)) to calculate the firm specific ERC (β_i).

CAR is the accumulated abnormal return, which is the sum of the abnormal return of each day through two days after last announcement date to one day after current announcement date (the reporting period). Daily abnormal return is the residual of the market model (equation (2)). I require firms to have quarterly earnings announcement date before the next fiscal quarter end to eliminate observations that have abnormal reporting date and I assign the last announcement date as three month before current reporting date for observations that miss last reporting date. Daily stock- and market returns of the past 300 days before two days after last announcement date are used to estimate the coefficients (i.e. γ, δ) for each firm-quarter observation and then the coefficients are applied to the reporting period to calculate the daily abnormal returns.

UE is the unexpected non-GAAP earnings of quarter q, which is the difference between the real non-GAAP earnings and expected non-GAAP earnings calculated on a per share basis scaled by firm's closing price at the end of quarter t. I use the quarterly managerial non-GAAP earnings per share (EPS) from [BW18] as the real non-GAAP earnings and median of the most recent analysts' consensus non-GAAP EPS estimate from I/B/E/S as the expected non-GAAP earnings.

The model (equation (1)) is estimated using firm specific time series quarterly data of 20 quarters before the reporting period of t for each firm-quarter observation on a rolling basis. I further require each observation has at least ten non-missing CAR and UE for estimating ERC.

4.2 Using sLDA to Estimate the Score

To get the DTM that can be used as input for LDA model, I preprocess the original data using tidytext package. All the paragraph breaks in press releases have been removed.Then press release text is tokenized to one-word term and turned into lower case. Then I remove all the stop words from tokenized corpus based on the "snowball" dictionary. After this step, I reorganize the corpus back to firm-quarter observations and then tokenize textual disclosures into two-word terms, which convey more meaningful information. The two-word terms are then stemmed to root and all numbers are removed from the corpus as well. After all the textual information has been preprocessed, it can then be transformed into DTM format for further analysis. The figure below is the first few lines of DTM for my dateset. In the matrix, every line represents a press release in my dataset, column names are all the terms in the corpus and the number in each cell represent the number of times this term appears in a specific press release.

```
<<DocumentTermMatrix (documents: 2000, terms: 4465)>>
Non-/sparse entries: 874024/8055976
Sparsity           : 90%
Maximal term length: 24
Weighting          : term frequency (tf)
Sample             :
      Terms
Docs    ended decemb financial measur fourth quart gaap financi income tax months end net incom non gaap operating incom
  152           171                0           42            0         30        183       25      14             35
  1661            0                3           50            3         12          0       41     120              0
  167            67               39            7           43         66         23       50     106             67
  1686            0               27           26           32         40         48       91     131             29
  1737           86               46            9           50         65         23       50     199             15
  297            94                0           22            0         14        154        0     166              5
  437            59               13            1           11         34         94       57      15              0
  550           108                4           53            2        113        110       66      93             24
  578            53                0           11            0         42          0        5       0              0
  709            35                9           65            7         49          8      170      16              0
```

Figure 1: An Example of DTM

The supervised latent Dirichlet allocation (sLDA) model is used to estimate the firm specific ERC (the non-GAAP disclosure quality measure) based on qualitative input. sLDA chooses latent topics that are associated with a dependent variable by grouping phrases based on the probability of the phrases co-occurring within disclosures ([BJ07]). Therefore, sLDA is more likely to identify the importance of groups of words and phrases when explaining a dependent variable.

One input for training sLDA model is the number of topics in the corpus. To empirically choose the optimal number of topics, I use five-fold cross-validation, iterate number of topics from 2 to 9 for LDA and calculate the perplextiy, which is the measure for model fit in topic models, for each iteration. The number of topics with smallest perplexity will be chosen as the optimal number of topics for corpus used in sLDA modelling.

The "lda" package in R can be used to conduct the sLDA analysis. Specifically, the function "slda.em" is used for the analysis, which requires multiple inputs. The "documents" requires a list input, which in my case will be the preprocessed data. "K" is the number of topics, which is the optimal number of topics obtained from LDA model. "vocab" is character vector including the vocabulary words used in documents, which can be got from the preprocessed corpus."params" is the initial values for the regression parameters, which I need to randomly assign. "annotations" is the outcome variable, which in my case would be the firm specific ERC.

The results of sLDA analysis contain coefficients on each of topic identified by the model. These coefficients indicate to what extent each topic is associated with the outcome variable. However, obtaining topics is not the focus of this research. The predictions of outcome variable based on these topics is the interest of this research. The slda.predict function will give the estimated outcome, which is the ERC predictions based on qualitative information in earnings press release.

5 Findings

5.1 Descriptive Findings

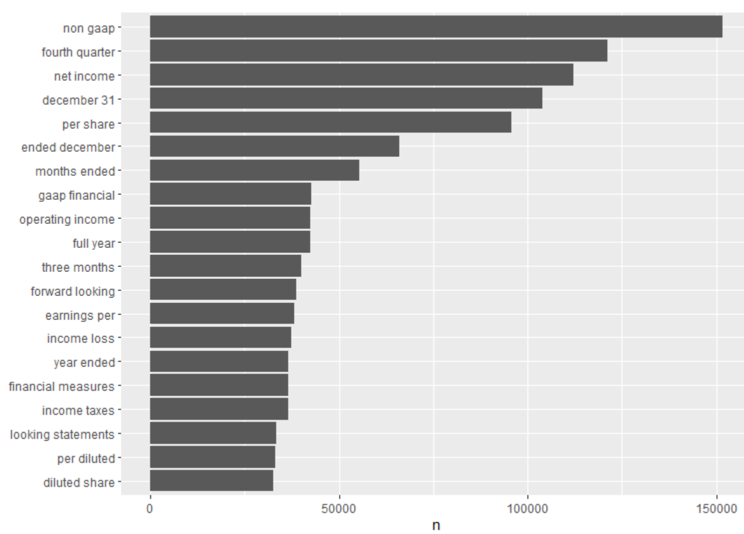

Figure 2: Count of Most Frequent Terms

The above chart shows the 20 most frequent terms in the corpus. It is not surprising that non-GAAP is most frequently mentioned term. With the term counts, we can also identify some other meaningful terms, like "forward looking", "income loss", "income taxes" and etc, while we also notice that a great portion of these terms are not meaningful, like "december 31", "per share", "year ended" and etc. This indicates that I may need to further develop my stop word dictionary adding these meaningless words so that they can be removed before model training. This is a time consuming process, which

I plan to update in my next draft. Results from this less refined corpus may be considered as the baseline for all my tests and validations.

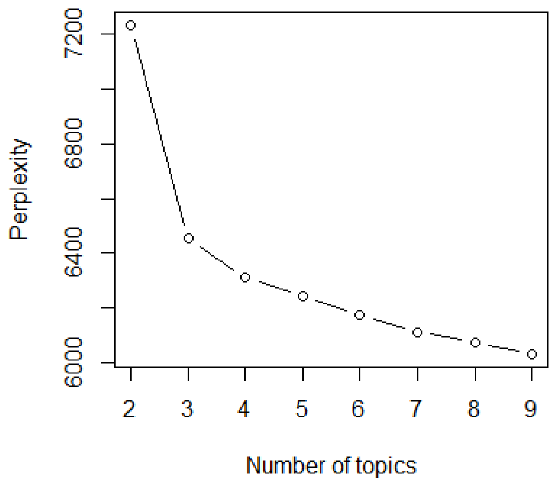

Figure 3: Perplexity of Different Number of topics

The results of LDA method (Figure 3) shows that when nine topics are assigned to the model, it generates the lowest perplexity, which indicates that nine is the optimal number of topics for my corpus.

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6
[1,]	"non gaap"	"fourth quart"	"net incom"	"net sal"	"ended decemb"	"cash flow"
[2,]	"financial measur"	"non gaap"	"non gaap"	"fourth quart"	"per shar"	"months end"
[3,]	"gaap financi"	"continuing oper"	"fiscal year"	"per shar"	"common shar"	"operating incom"
[4,]	"gaap net"	"net incom"	"per shar"	"net earn"	"year end"	"ended decemb"
[5,]	"based compens"	"per dilut"	"fourth quart"	"prior year"	"investment incom"	"per shar"
[6,]	"per shar"	"diluted shar"	"quarter fisc"	"full year"	"fair valu"	"earnings p"
[7,]	"stock bas"	"pre tax"	"diluted shar"	"earnings p"	"per common"	"non gaap"
[8,]	"net incom"	"per shar"	"per dilut"	"operating incom"	"three month"	"income tax"
[9,]	"gaap oper"	"income tax"	"gaap net"	"non gaap"	"net incom"	"free cash"
[10,]	"intangible asset"	"full year"	"forward look"	"operating profit"	"net invest"	"three month"

	Topic 7	Topic 8	Topic 9
[1,]	"fourth quart"	"fourth quart"	"net incom"
[2,]	"per shar"	"natural ga"	"adjusted ebitda"
[3,]	"net incom"	"net incom"	"fourth quart"
[4,]	"full year"	"gains loss"	"ended decemb"
[5,]	"income tax"	"combined ratio"	"income loss"
[6,]	"ended decemb"	"operating incom"	"adjusted net"
[7,]	"special item"	"per shar"	"year end"
[8,]	"non gaap"	"income loss"	"income tax"
[9,]	"months end"	"prior year"	"three month"
[10,]	"operating expens"	"full year"	"depreciation amort"

Figure 4: Most Frequent Words in Each Topic

Figure 4 presents the top ten words by counts for each of the topics. From the results we can find that there are significant varieties of key words across topics, although the corpus still suffers from frequent meaningless words. With these key words of each topic, we can roughly decide the meaning of each topic. LDA model will assign weights to different terms in each press release with respect to different topics. The sum of weights of different topics enable us to identify the main topics of each press release and therefore be used to link topics with the output in sLDA model. Figure 5 presents the top two topics of the first ten press releases in my dataset. The column is the index of press releases in the dataset and the first- and second row are top two topic numbers for each press release respectively. From this example, we can see that topic six seems to be a popular topic in these press releases.

	297	1236	1218	1737	421	1495	915	1393	152	323
[1,]	4	1	5	2	5	2	1	2	6	7
[2,]	6	6	8	6	8	6	6	6	7	5

Figure 5: Top Two Topics for First 10 Press Releases

5.2 Model Results

```
Call:
lm(formula = annotations ~ z.bar. + 0)

Residuals:
    Min      1Q  Median      3Q     Max
-331.72  -16.70   -7.04    9.47  419.81

Coefficients:
        Estimate Std. Error t value Pr(>|t|)
z.bar.1   17.850      3.316   5.384 8.17e-08 ***
z.bar.2   17.855      3.690   4.838 1.41e-06 ***
z.bar.3   17.182      2.047   8.393  < 2e-16 ***
z.bar.4   20.888      5.140   4.064 5.02e-05 ***
z.bar.5   16.184      6.828   2.370   0.0179 *
z.bar.6    3.741      5.913   0.633   0.5271
z.bar.7   18.426      4.265   4.320 1.64e-05 ***
z.bar.8   14.033      3.424   4.099 4.32e-05 ***
z.bar.9    8.917      5.014   1.778   0.0755 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 40.56 on 1991 degrees of freedom
Multiple R-squared:  0.141,     Adjusted R-squared:  0.1371
F-statistic:  36.3 on 9 and 1991 DF,  p-value: < 2.2e-16
```

Figure 6: Coefficient Estimates on Each Topic

The results for sLDA model with nine topics are presented in Figure 6. Except topic 6 and 9, all the coefficients on other topics are significant. It seems that my trained sLDA model is useful in predicting the firm specific ERC.

To further validate my results. I follow [DL21], regressing firm specific ERCs on the predicted ERCs (yhat). The model summary in Figure 7 indicates that the coefficient on predicted ERC is positive and significant at 1% level, suggesting that the qualitative information in earnings press releases captures non-GAAP earnings quality measured by ERCs.

```
Call:
lm(formula = erc ~ yhat)

Residuals:
    Min      1Q  Median      3Q     Max
-331.94  -16.62   -7.04    9.44  419.90

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.3026     5.5256   0.055  0.95633
yhat          0.9813     0.3370   2.912  0.00363 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 40.49 on 1998 degrees of freedom
Multiple R-squared:  0.004226,	Adjusted R-squared:  0.003728
F-statistic:  8.48 on 1 and 1998 DF,  p-value: 0.00363
```

Figure 7: Out-of-Sample Validation

6 Conclusion

I test whether machine learning method can extract qualitative information in non-GAAP earnings press releases to predict non-GAAP earnings quality. I train sLDA model to identify textual information that explains non-GAAP earnings quality. The model can be further applied to all firms with non-GAAP earnings press releases disclosures. Using holdout samples, I verify that my non-GAAP earnings quality measure has a significant association with firms' non-GAAP earnings quality as measured by firm specific non-GAAP ERCs.

One potential drawback to using machine learning methods is that these methods could identify phrases that relate statistically but not economically related to non-GAAP earnings quality. Especially in my current test, there still are a lot of meaningless key words identified in the corpus. Possibly, noise could be introduced into my measure. I may need to tackle this issue in future drafts.

My study adds to the growing body of research using machine learning method to extract information from firm's qualitative disclosure to explain key outcomes (e.g. credit risks, misreporting, fraud and etc.) In addition, my research adds to the research that examines qualitative information in disclosures (e.g. tone, complexity, words using and etc.). While investors use both quantitative and qualitative disclosure for their investment decisions, extant research on non-GAAP disclosures mainly focuses on non-GAAP earning numbers. My study helps fill this void by developing a comprehensive, less subjective and scalable measure of non-GAAP earnings quality. Information users can use my measure to supplement other methods to generate a more comprehensive and independent estimate of non-GAAP earnings quality.

References

[Bar12] I. D. Gow D. J. Taylor Barth, M. E. Why do pro forma and street earnings not reflect changes in gaap? evidence from sfas 123r. *Review of Accounting Studies*, 2012.

[BC09] D. E. Black and T. E. Christensen. Us managers' use of 'pro forma' adjustments to meet strategic earnings targets. *Journal of Business Finance Accounting*, 2009.

[BE20] R. M. Crowley Brown, N. C. and W. B. Elliott. What are you saying? using topic to detect financial misreporting. *Journal of Accounting Research*, 2020.

[BJ03] A. Ng Blei, D. M. and M. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 2003.

[BJ07] D. Blei and Mcauliffe J. Supervised topic models. *Advances in Neural Information Processing Systems*, 2007.

[BL03] E. L. Black T. E. Christensen Bhattacharya, N. and C. R. Larson. Assessing the relative informativeness and permanence of pro forma earnings and gaap operating earnings. *Journal of Accounting and Economics*, 2003.

[BS02] M. T. Bradshaw and R. G. Sloan. Gaap versus the street: An empirical assessment of two alternative definitions of earnings. *Journal of Accounting Research*, 2002.

[BW18] T. E. Christensen K. H. Gee Bentley, J. W. and B. C. Whipple. Disentangling managers' and analysts' non-gaap reporting. *Journal of Accounting Research*, 2018.

[CK89] D. W. Collins and S. P. Kothari. An analysis of intertemporal and cross-sectional determinants of earnings response coefficients. *Journal of Accounting and Economics*, 1989.

[CW14] S. E. McVay Curtis, A. B. and B. C. Whipple. The disclosure of non-gaap earnings information in the presence of transitory gains. *The Accounting Review*, 2014.

[CY21] Y.-J. Lee S.-Y. Lo Chen, H.-C. and Y. Yu. Qualitative characteristics of non-gaap disclosures and non-gaap earnings quality. *Journal of Accounting and Economics*, 2021.

[DL21] J. Jennings-K. Koharki Donovan, J. and J. Lee. Measuring credit risk using qualitative disclosure. *Review of Accounting Studies*, 2021.

[DS12] J. M. Piger Davis, A. K. and L. M. Sedor. Beyond the numbers: Measuring the information content of earnings press release language. *Contemporary Accounting Research*, 2012.

[FL16] J. Jennings Frankel, R. and J. Lee. Using unstructured and qualitative disclosures to explain accruals. *Journal of Accounting and Economics*, 2016.

[HZ14] S. H. Teoh Huang, X. and Y. Zhang. Tone management. *The Accounting Review*, 2014.

[KM08] C. A. Marquardt Kolev, K. and S. E. McVay. Sec scrutiny and the evolution of non-gaap reporting. *The Accounting Review*, 2008.

[Li10] F Li. The information content of forward-looking statements in corporate filings—a naïve bayesian machine learning approach. *Journal of Accounting Research*, 2010.

[LM04] B. A. Lougee and C. A. Marquardt. Earnings informativeness and strategic disclosure: An empirical examination of "pro forma" earnings. *The Accounting Review*, 2004.

[LM16] T. Loughran and B. Mcdonald. Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, 2016.

[TW96] W. R. Teets and C. E. Wasley. Estimating earnings response coefficients: Pooled versus firm-specific models. *Journal of Accounting and Economics*, 1996.