2021

Analyst Non-GAAP Reporting: When Do Equity Analysts Adjust GAAP Earnings – an Application of Text Based Firm Complexity

Patrick Chen

Claremont McKenna College

Analyst non-GAAP reporting: When do equity analysts adjust GAAP earnings – an application of text based firm complexity

submitted to
Professor George Batta

by
Patrick Chen

for
Senior Thesis
2021 Spring
May 03, 2020

Abstract

Equity analysts adjust GAAP earnings to report non-GAAP figures that they believe better represent a firm's current and future performance. The importance of such non-GAAP(street) figures has been determined and accepted by the literature, but the process in which analysts arrive at such figures is less understood. I apply a newly defined measure of firm complexity using text based analysis of firm annual reports and investigate its effect. When complexity increases, it becomes harder to individuate different components of a system. When that system is a firm, complexity then increases the difficulty in determining an accurate measure of performance. The effect of this type of firm characteristic on analysts has yet to be explored. I find that firms with higher complexity exhibit a greater difference between GAAP earnings and analyst adjusted non-GAAP earnings. A cross sectional test featuring earnings volatility is also found to increase the marginal effect of complexity on the difference. This suggests that firm complexity plays an important role in influencing an analyst's decision to adjust GAAP earnings.

Contents

1 Introduction

Publicly traded companies listed on exchanges within the United States are mandated by the SEC to follow US GAAP financial reporting guidelines as issued by the FASB. However, managers, financial analysts, individual investors, and other stakeholders are observed to evaluate a company's earnings performance through methods that in fact stray away from GAAP rules. These methods are characterized by the exclusion of certain items from GAAP earnings that are deemed nonrecurring or irrelevant to determining a firm's true value. So-called "non-GAAP" earnings have garnered a significant amount of literature investigating it's behavior, with the majority of studies centered around the role of firm management and equity analysts in its definition. Management issues guidance and discloses non-GAAP measures in their quarterly and annual reports in addition to required GAAP measures. Analysts adjust GAAP earnings in their forecasts which are then consolidated by financial data providers like I/B/E/S.

I will be focusing on the topic of non-GAAP earnings from the analyst standpoint. This subsection of the non-GAAP literature has adopted the term "street earnings" to signify equity analyst adjusted earnings (Bradshaw and Sloan 2002; Abarbanell and Lehavy 2007). While prior research suggests the importance of street earnings to investors, the process in which equity analysts exclude or include items from GAAP earnings to create street earnings is less understood. I investigate the effects of certain predictor variables on the magnitude of difference between GAAP and analyst adjusted street earnings. The first of such predictor variables will be a text based measure of firm complexity as defined by a 2020 paper by Loughran and McDonald. In addition to firm complexity, I will utilize proxies of phenomena that have been found to have significant effect on the difference between GAAP and street earnings from previous literature. These phenomena include analyst ability, analyst incentive, management guidance (Gu and Chen 2004; Baik, Farber, and Petroni 2009; Christensen et al. 2011). Although each variable has been investigated separately, it is to my knowledge that a holistic study combining and controlling for all of these hypotheses has not yet been attempted. Firm complexity measures a distinct and important characterization that is unique to each individual firm (Loughran and McDonald 2020). However, the measure itself has rarely been featured in literature due to inconsistent definitions and difficulty in measurement. The usage of word count in an annual report has been a relatively popular definition of firm complexity (Lehavy, Li, and Merkley 2011; Loughran and McDonald 2014; Dyer, Lang, and Stice-Lawrence 2017). Number of business segments and the existence of foreign sales as given by Compustat has also been used as an indicator of firm complexity (Doyle, Ge, and McVay 2007; Ge and McVay 2005; Ashbaugh-Skaife, Collins, and Lafond 2009; Cohen and Lou 2012). Other definitions include number of XBRL segments, fractional percentage of intangible assets relative to total assets, and derivative usage (R. Hoitash and U. Hoitash 2018; Gomes, Gorton, and Madureira 2007; Chang, Donohoe, and Sougiannis 2016). The complexity measure that I will utilize is a newly defined measure as created by Loughran and McDonald. In their 2020 paper, Loughran and McDonald define a 374 word lexicon of words representative of complexity. The number of unique complex word occurrences on a firm's 10-K report is recorded and normalized by the total number of complex words (374) to output a complexity score. Efficacy of the complexity score is determined as it is found to be highly associated with a firm's audit fees, which itself is determined by the size of a firm and how complex the audit will be.

Complex systems like firms are best understood as a collection of predictable components. When the complexity of a system increases, it becomes more difficult to disentangle one component from another. Since the interactions between different components of a firm are unpredictable and chaotic, predictability becomes more challenging (Loughran and McDonald 2020). This is the motivation to include firm complexity in the exploration of analyst non-GAAP reporting. As complexity increases, the permanent and transient factors of a firm's earnings become harder to measure and may stray from defined GAAP systems. Accordingly, I predict as my first hypothesis that equity analysts will be more likely to make their own adjustments as to what consists of a firm's true value as complexity increases, increasing the difference between analyst-adjusted non-GAAP earnings and GAAP earnings. In order to strengthen the potential findings of the relationship between complexity and difference, I also predict as my second hypothesis that for more volatile firms, complexity will have a greater effect on the difference than for less volatile firms.

To test the effect of firm complexity on the differences in GAAP and equity analyst-adjusted non-GAAP earnings, I will first define the difference in earnings measures as the magnitude of the difference between IBES actual EPS (non-GAAP) and CapIQ diluted EPS excluding extraordinary items (GAAP). Earnings data will come from these two data sources for firms between the years 2000 and 2021, and feature firms that have been consistently a component of the S&P 500 since 2000. The measure of complexity will not come directly from Loughran and McDonald as the data they provide only features data up until 2018. Instead, a self written scraping and parsing algorithm will collect the data following the same conceptual background. Control variables from the aforementioned phenomena will also be collected from various sources like CapIQ, IBES, and CRSP.

In a cluster-robust regression of complexity on the difference, we find that firms with higher complexity do indeed exhibit a higher difference in earnings definitions. These results are enhanced by adding firm fixed effects and controlling for time based relationships in creating dummy variables for each year. Although the significance of the relationship between complexity and difference decreases, significance of the positive relationship is maintained at the 5% level. With regards to the second hypothesis, an interesting pattern emerges. Another regression with cluster robust standard errors is run with volatility defined as either a continuous variable or a binary variable split on median volatility. In both models, the individual volatility variable has a negative effect on difference. However, the interaction term between complexity and volatility has a much larger positive effect. The marginal effect of higher volatility is therefore positive as well, although it must be noted that only the interaction terms between complexity and the volatility variables have statistical significance.

2 Literature Review

Various aspects of equity analyst non-GAAP (street) earnings have been researched in past literature. Researchers have first explored the importance of street earnings as a worthwhile metric for consideration. Bradshaw and Sloan (2002) investigates the two alternative definitions of earnings as GAAP earnings and street earnings. They find that there has been a steady increase in the difference between the two earnings definitions since 1985. Street earnings have been increasingly associated with stock prices while GAAP earnings have declined in association, suggesting that the market put more weight on street earnings definitions rather than GAAP earnings. Brown and Sivakumar (2003) structures a similar study comparing the value relevance of IBES street earnings versus Compustat GAAP earnings. Street earnings are found to be more value relevant than GAAP earnings, echoing the findings of Bradshaw and Sloan.

These studies have laid a firm groundwork for the apparent importance of street earnings. However, the very process in which analysts make adjustments to GAAP earnings is not entirely understood in the literature. Gu and Chen (2004) find that the items analysts included in their earnings definitions, which are rolled up into the street earnings definition, are more persistent than the items in which they exclude. Curiously, analysts were also found to have excluded items which were evidently persistent and value relevant. This suggests that analyst expertise serves as a factor for analysts to selectively elect inclusions and exclusions of nonrecurring line items. Baik, Farber, and Petroni (2009) offers empirical evidence for economic incentives influencing analyst's decisions of inclusions and exclusions by investigating glamour stocks. They find that analysts are more likely to exclude expense items from earnings for these glamour stocks. Christensen et al. (2011) moves away from analyst-centric predictors and investigates the role of management in the adjustment of GAAP earnings by analysts. For firms whose management has issued specific earnings guidance during the fiscal period, there is a much higher amount of incremental exclusion by equity analysts.

In attempting to answer the question of how street earnings are determined, we can see that previous literature has focused on separate and distinct predictors. However, it is unrealistic to assume that equity analysts are only ever influenced by one factor at a time. In this paper I will seek not only to model out studied predictors as control variables, but also new factors that would all work in conjunction in application.

3 Hypothesis

The role of an equity analyst is to determine the true value of a firm's stock and predict future financial performance by looking at and analyzing the financial data that a firm provides. Financial modeling and forecasting is a liquid process that has no cookie cutter template if done with care and expertise. This makes sense as all firms are constantly changing, evolving, and reacting to certain developments and events. Looking at GAAP earnings data as provided in 10-K documents, it exists as a more standardized figure created from a given set of rules as defined by the FASB. However a common critique of GAAP is that the rules maintain a picture of a firm that is too conservative. One such example of GAAP conservatism is the phenomenon that large one time non-cash expenses are much more prevalent in earnings calculations than one time non-cash gains (Heflin, Hsu, and Jin 2015). This results in firms experiencing negative income shocks due to transitory events like impairments and restructuring charges. Thus one way for analysts to reach a "truer" picture of firm performance and value is to make their own adjustments by excluding this transitory information in GAAP financials which is deemed irrelevant to a firm's future performance. Research has indeed determined that these supplemental exclusions which analysts elect are informative with value added (Gu and Chen 2004; Heflin, Hsu, and Jin 2015).

With regards to complexity, the lexicon of 374 of complexity words as compiled by Loughran and McDonald seeks to include words that a reader of a firm's annual report would find to signal an increasing difficulty in forecasting future cash flows or creating audited financial disclosures. As an example, if an annual report uses language that describes leases, intangible assets, international operations, or acquisitions, forecasting the operating performance of a firm would prove to become more challenging (Loughran and McDonald 2020). The complexity lexicon was designed to take all of such cases into account when determining the complexity score. When complexity is high, analysts would be more inclined to look into this complexity and determine whether or not certain items as retained by GAAP are representative of a firm's true value. This serves as the motivation for my first hypothesis.

Hypothesis 1 (H1): *The more complex the firm, the larger the difference will be between GAAP earnings and equity analyst adjusted non-GAAP earnings.*

This hypothesis may not hold if this measure of complexity does not accurately determine the "true" complexity of a firm. Although Loughran and McDonald have robustly tested the efficacy of such a measure, there is no direct test which verifies the relationship between complexity as defined by the 374 word lexicon versus the indicators that equity analysts are attuned to like impairments and restructuring charges. Another reason is that even if the complexity measure functions as intended, equity analysts may instead stick to GAAP earnings when complexity is high. Similarly to how equity analysts are more likely to exclude when management issues guidance, high complexity may be cautionary signal to stick to a more conservative outlook on firm performance.

3.1 Earnings Volatility

In order to further investigate the relationship between complexity and the difference in earnings definitions, I will be conducting a cross-sectional test featuring earnings volatility. A widely held belief among management is that earnings volatility is negatively related to the predictability of a firm's future earnings (Dichev and Tang 2009). This means that if the standard deviation of a firm's earnings is higher, it will be harder to determine how that firm will perform in the future. Considering how I expect complexity to exhibit the same behavior, the interaction between volatility and complexity should then increase that effect. This serves as the motivation for my second hypothesis.

Hypothesis 2 (H2): *The marginal effect of complexity on the difference in earnings definitions will be more pronounced in firms with high earnings volatility compared to firms with low earnings volatility.*

A concern for this hypothesis is that volatility itself might have and inverse relationship to difference in earnings definitions. Considering earnings volatility as a singular predictor, equity analysts may stick closer to GAAP earnings due to the wild swings in earnings that is expected of volatile firms.

4 Data

Earnings and predictor data was collected for S&P 500 component companies with a sample period spanning from the start of 2000 until the beginning of March 2021. However, since the component list of S&P 500 companies is subject to constant change, I have restricted the sample data set to only include firms that have been a component company since the start of the sampling period, and have continued to stay as a component until the date that I collected the data. This ensures that for each firm included there is continuity in the data. Out of 505 current S&P 500 component companies, only 201 remained after filtering.[1] Figure 1 shows the distribution of industry sectors for the firms that have remained.

4.1 Difference in Earnings

GAAP earnings are defined as diluted earnings per share excluding extraordinary items as given by S&P Capital IQ Financials. While the vast majority of non-GAAP literature has featured Compustat as the GAAP data provider of choice, I use CapIQ Financials specifically as a more accurate measure of the information that equity analysts interact with. Data collection for Capital IQ Financials is facilitated by the use of the Excel add-on. With 201 component companies spanning 22 years, the CapIQ_financials dataset contains 4422 observations. It must be noted however that many firms have not reached their fiscal year end for 2021, so the vast majority of 2021 data is missing.

[1]Historical S&P 500 component data collected from: https://github.com/fja05680/sp500

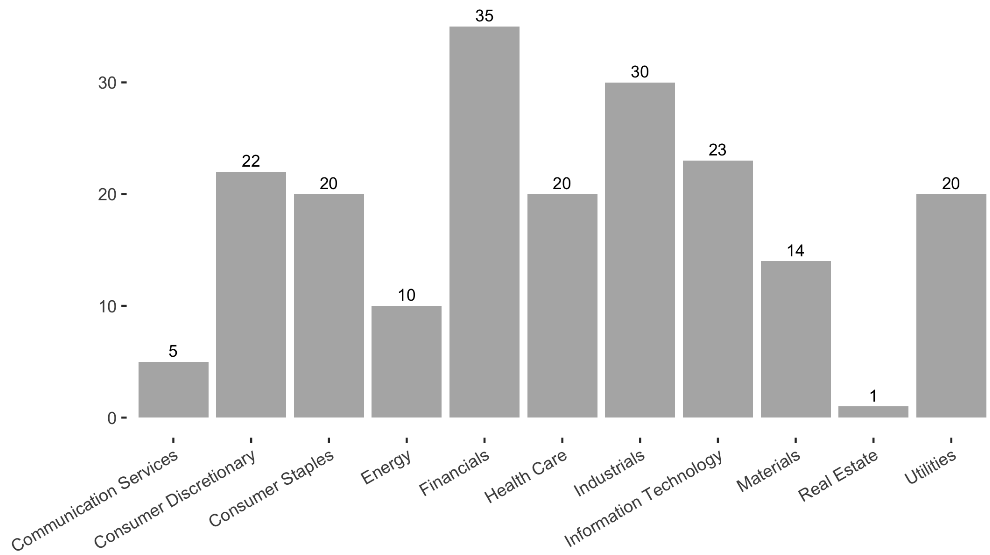

Figure 1: Distribution of Industry Sectors

Analyst non-GAAP earnings is defined using I/B/E/S "Actuals" EPS accessed using WRDS. The I/B/E/S actual value for a given firm in a given year represents the earnings per share value after adjustments as defined by the majority of equity analysts. The data contains 3958 observations.

In order to create the difference variable, both CapIQ GAAP EPS and I/B/E/S non-GAAP EPS are first deflated by the last sale price on the fiscal year end date to give a market cap adjusted measure of firm earnings. The difference value is then defined as follows:

$$\text{diff} = abs(\text{eps\_ibes} - \text{eps\_ciq})$$

Here we take the magnitude of the difference since we are only worried about if there is a difference present between the two measures, not if the difference is positive or negative. These differences are then winsorized at the 1% and 99% levels to mitigate the effect of extreme outliers.

4.2 Complexity

Although I am referencing the Loughran and McDonald paper on text based firm complexity, I will not be directly using their figures as reported from their findings. The mechanics of returning the variable complexity will be the same conceptually, but different procedurally. Like Loughran and McDonald, I will be using the same 374 word complexity lexicon. However, the way in which each 10-K document is parsed to look for the complexity words is marginally different. An algorithm is run to first scrape through the SEC EDGAR web fillings to retrieve the files of choice for the 201 S&P 500 component companies. After a 10-K file is located, I run a text parsing algorithm to extract all human readable text from the document. A number of companies store a significant portion of their MD&A section in Exhibit 13 of a separate annual report, so for such companies Exhibit 13 is also parsed. With this parsed text, a complexity score can be calculated by dividing the number of complex words that appear in the 10-K at least once by 374, the total number of complex words. That is:

$$\text{complexity} = \frac{\sum \text{unique complex word occurrences}}{\text{length of the complexity lexicon}}$$

Therefore, each complexity value represents the percentage of complex words from the complexity lexicon that exist within the human readable text of a 10-K document.

Due to the limitations of EDGAR, this algorithm was run for firms that filed their documents starting from 2001, and all the way until present day. Original Loughran and McDonald figures were only run until 2018. There are minimal differences between the complexity values, so all complexity values in this study will originate from my algorithms for consistency. Some manual cleaning was also required. For example, Clorox Co filed their 2003q2 10-Q as a 10-K file, which was treated as a 10-K file in the algorithm. Irregularities in fiscal year period end corresponding with leap year characteristics also required manual cleaning. See Appendix 7.1 for the complexity data.

4.3 Base Sample

For the base regression, we end up with a sample of 3,658 observations featuring 200 firms. In this sample, most years are populated with close to 200 firm observations, which the exception of 2000 and 2020 (see Appendix 7.3). The cause of data loss in 2000 is due to the complexity measure and the usage of SEC EDGAR. For fillings submitted before 2001 on EDGAR, there is no consistent working method to access 10-K information. Therefore, the remaining 84 firms are firms with 2000 fiscal year ends who have filed their reports in 2001. For the data loss in 2020, the main reason is the lack of population of I/B/E/S actuals earnings data for firms. Since data was collect on March 15 of 2021, many firms have yet to file for their 2020 fiscal year end. From a industry sector breakdown we see that there is an representation of firms that is heavily weighted towards Financials and Industrials. This is mostly due to the fact that I have only kept firms in the sample that have been a S&P 500 component since the start of 2000. Table 1 shows the process of sample selection for the base regression and where observations and firms are lost.

Table 1: Sample Selection for Base Regression

Sample Selection	Observations	Firms
Starting annual observations for 201 firms for 22 years (2000-2021)	4,422	201
Observations after removing missing I/B/E/S earnings (defined as of March 15, 2021)	3,951	201
Observations with price data	3,930	200
Observations with complexity data	3,658	200

Table 2 displays the summary statistics for the variables that are included in the base regression. On average, *eps_ibes* is 0.056 while *eps_ciq* is 0.040. This is expected as equity analyst adjusted earnings will paint a more optimistic picture of earnings than GAAP. GAAP earnings does have more variability, as standard deviation is higher for *eps_ciq* at 0.384 compared to *eps_ibes* at 0.221. We then see that the average *diff* is relatively the difference of the average earnings definitions. On average, *complexity* is 0.269, which means that 26.9 percent of complexity words from the complexity lexicon are used in a 10-K document.

Table 2: Summary Statistics for Base Regression

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
eps_ibes	3,658	0.056	0.221	−12.682	0.047	0.076	0.330
eps_ciq	3,658	0.040	0.384	−22.429	0.038	0.070	0.487
complexity	3,658	0.269	0.052	0.061	0.233	0.305	0.404
diff	3,658	0.016	0.033	0.000	0.001	0.016	0.226

4.4 Control Variables

Four control variables are collected to be included in the regression.

The first control variable is firm size as determined by the market capitalization for a given firm at a given period end date. Market cap data is collected from CapIQ financials the same way that GAAP EPS is collected. The market cap figure is logged to achieve a more normally distributed distribution.

The second control variable relates to the phenomenon in which managers play an active role in influencing the composition of street earnings via earnings guidance (Christensen et al. 2011). To control for such a relationship, I collect guidance data from CapIQ financials as a dummy variable. If management has issued earnings guidance in a given year, the guidance variable equals 1. If not, the guidance variable equals 0.

The third control variable relates to the phenomenon in which analyst expertise affects the adjustments that are made to GAAP earnings (Gu and Chen 2004). To control for analyst expertise, I utilize the Detail History file of I/B/E/S which includes analyst codes. Keeping track of the number of years that a specific analyst code has issued an earnings forecast since the beginning of I/B/E/S data availability, I can calculate the average analyst experience for a given firm in a given year. Average analyst experience is then winsorized at the 1% and 99% levels to mitigate the effect of extreme outliers.

The last control variable relates to the phenomenon that analysts covering firms with glamour stock status will be more motivated to exclude items from GAAP earnings to paint a better picture of future performance. To control for glamour stock status, I will utilize a measure of stock turnover following previous literature (Baik, Farber, and Petroni 2009). That is:

$$\text{turnover} = \frac{\text{average monthly trading volume}}{\text{number of shares outstanding}}$$

Trading volume and shares outstanding data are both collected from CRSP on WRDS. The turnover figure is also logged to achieve a more normally distributed distribution.

4.5 Control Sample

For the regression with control variables included, there is data attrition due to missing data in a given control variable. In the I/B/E/S data, there were 12 observations where analyst experience could not be calculated and returned an N/A Value. When including the turnover figure from CRSP, one firm returns no information is therefore is omitted. Volatility data is also included in this table, although we use it in the cross-sectional test. The loss of one observation is due to the one firm only having one observation in the data, leading to an N/A standard deviation calculation. Table 3 shows the process of sample selection for the control and cross-sectional regression and where observations and firms are lost.

Table 3: Sample Selection for Control Regression

Sample Selection	Observations	Firms
Starting observations from base regression	3,658	200
Observations with analyst experience figure	3,646	200
Observations with turnover figure	3,640	199
Observations with earnings volatility figure. (One firm with only one observation)	3,639	198

Table 4 displays the summary statistics for new variables that are introduced in the control regression, in addition to the variables that are included in the base regression. The number of observations has decreased, which slightly changes the summary statistics for those variables included in the base regression. However the overarching relationships and behaviors of *eps* figures, *complexity*, and *diff* stay the same. The logged market cap has a average of 10.080 with a standard deviation of 1.128. Guidance has a mean of 0.585, which means that on average firm management issues guidance 58.5% of the time. The average experience for equity analysts covering a firm is 9.751 years with a standard deviation of 2.757. The logged turnover value has an average of 0.358 and is quite variable with a standard deviation of 0.542.

Table 4: Summary Statistics for Control Regression

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
eps_ibes	3,640	0.057	0.221	−12.682	0.047	0.076	0.330
eps_ciq	3,640	0.040	0.385	−22.429	0.038	0.070	0.487
diff	3,640	0.016	0.033	0.000	0.001	0.016	0.228
complexity	3,640	0.269	0.052	0.061	0.233	0.305	0.404
lmk_cap	3,640	10.080	1.128	6.925	9.296	10.777	14.468
guidance	3,640	0.585	0.493	0	0	1	1
avg_exp	3,640	9.751	2.757	3.980	7.770	11.528	16.926
lturnover	3,640	0.358	0.542	−1.740	−0.013	0.688	3.368

5 Results

5.1 Correlation

Table 5 provides both the Spearman correlations and Pearson correlations of the variables included in the regression. Consistent with my expectations and hypothesis, complexity is positively correlated both linearly and monotonically with the difference in earnings definitions. Logged market cap is negatively correlated with the difference. This correlation is much more significant linearly than monotonically. Guidance has no linear correlation to the difference, but it will change along with the difference together although not at the same rate. Both analyst experience and logged turnover also show a positive correlation to the difference at highly significant levels.

Table 5: Spearman(Pearson) Correlations above(below) diagonal

	diff	complexity	lmk_cap	guidance	avg_exp	lturnover
diff		0.31****	-0.04*	0.05**	0.10****	0.07****
complexity	0.19****		0.22****	0.12****	0.22****	0.07****
lmk_cap	-0.12****	0.22****		0.17****	0.15****	-0.36****
guidance	0.00	0.12****	0.16****		0.18****	0.12****
avg_exp	0.06***	0.24****	0.15****	0.17****		-0.07****
lturnover	0.18****	0.10****	-0.37****	0.10****	-0.08****	

****$p < 0.0001$; ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$

5.2 Regression

The equation of the base regression in Model 1 is as follows:

$$diff_{i,t} = \beta_0 + \beta_1 complexity_{i,t} + \epsilon_{i,t}$$

I then control for previously studied phenomenon of guidance, analyst experience, and glamour status in Model 2 with the regression equation below:

$$diff_{i,t} = \beta_0 + \beta_1 complexity_{i,t} + \beta_2 lmk\_cap_{i,t} + \beta_3 guidance_{i,t} + \beta_4 avg\_exp_{i,t} + \beta_5 lturnover_{i,t} + \epsilon_{i,t}$$

Introducing firm fixed effects in Model 3:

$$\begin{aligned} diff_{i,t} = {} & \beta_1 complexity_{i,t} + \beta_2 lmk\_cap_{i,t} + \beta_3 guidance_{i,t} \\ & + \beta_4 avg\_exp_{i,t} + \beta_5 lturnover_{i,t} + firmfixedeffects_i + \epsilon_{i,t} \end{aligned}$$

And introducing year dummy variables in Model 4:

$$\begin{aligned} diff_{i,t} = {} & \beta_1 complexity_{i,t} + \beta_2 lmk\_cap_{i,t} + \beta_3 guidance_{i,t} \\ & + \beta_4 avg\_exp_{i,t} + \beta_5 lturnover_{i,t} + \beta_{6-25} dummy(t) + firmfixedeffects_i + \epsilon_{i,t} \end{aligned}$$

Table 6 features the regression output for these four models.

In the base regression of Model 1, the *complexity* coefficient is statistically significant at the 0.1 percent level with a value of 0.1203. An increase in *complexity* by 1 percent increases the magnitude of difference in price deflated earnings definitions by 0.0012. Although this value means little in the practical sense, it gives a solid baseline that *complexity* does positively affect the *diff* as our first hypothesis predicts. To determine a practical significance, one standard deviation increase in *complexity* (0.052) increases *diff* by 0.0063 ($0.052 * 0.1203$). This is a 19 percent ($0.0063/0.033$) increase in the standard deviation of *diff*. Here the R^2 value is low at 0.0353, so there is a lot of unexplained variation left over.

Table 6: Model 1 is a cluster robust linear regression between complexity and diff. Model 2 features the introduction of control variables. Model 3 introduces firm fixed effects. Model 4 introduces year dummy variables to control for time trends.

	Model 1	Model 2	Model 3	Model 4
(Intercept)	-0.0161^{**}	0.0097		
	(0.0051)	(0.0107)		
complexity	0.1203^{***}	0.1248^{***}	0.1132^{**}	0.1022^{*}
	(0.0203)	(0.0206)	(0.0351)	(0.0439)
lmk_cap		-0.0033^{**}	-0.0129^{***}	-0.0168^{***}
		(0.0010)	(0.0023)	(0.0028)
guidance		-0.0017	0.0002	-0.0002
		(0.0014)	(0.0012)	(0.0011)
avg_exp		0.0005	0.0012^{**}	0.0001
		(0.0004)	(0.0004)	(0.0004)
lturnover		0.0080^{***}	0.0092^{***}	0.0111^{**}
		(0.0023)	(0.0026)	(0.0034)
R^2	0.0353	0.0761	0.3229	0.3516
Adj. R^2	0.0350	0.0749	0.2829	0.3093
Num. obs.	3658	3640	3640	3640
RMSE	0.0326	0.0320	0.0282	0.0277
N Clusters	200	199	199	199

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

In Model 2, *complexity* maintains statistical significance at the 0.1 percent level with a coefficient of 0.1248. With the control variables added into the regression equation, the effect of *complexity* on *diff* actually increases. One standard deviation increase in *complexity* increases *diff* by 0.0065 ($0.052 * 0.1248$), or 20 percent ($0.0065/0.033$) of its standard deviation. Two out of the four control variables are also statistically significant. An increase in *lmk_cap* by 1 decreases *diff* by 0.0033, with this effect reaching statistically significance at the 1 percent level. Glamour stock status as proxied by *lturnover* increases *diff* by 0.008 and is statistically significant at the 0.1 percent level. *Guidance* has a negative effect on *diff* and *avg_exp* has a positive effect on *diff*, but both of these variables are not statistically significant. This model still does a poor job of explaining our independent variable of *diff* with an R^2 value of only 0.0761.

By introducing firm fixed effects in Model 3, the time invariant characteristics across firms are controlled for. With this model, *complexity* is still statistically significant although at the lower 1 percent level. The coefficient has also decreased to 0.1132. To consider practical significance, one standard deviation increase in *complexity* increases *diff* by 0.0059 ($0.052 * 0.1132$), or 18 percent of its standard deviation ($0.0059/0.033$). *Lmk_cap* and *lturnover* maintain their statistical significance with coefficients of -0.0129 and 0.0092, respectively. With firm fixed effects analyst experience *avg_exp* also becomes statistically significant at the 1 percent level with a coefficient of 0.0012. The effect of *guidance* on *diff* becomes positive, but still statistically insignificant. With firm fixed effects, the R^2 value jumps to 0.3229, meaning that about a third of the unexplained variation is explained by our model.

Finally to account and control for time trends, 20 year dummy variables from 2000 to 2019 are included in the model. These variables are not shown in the regression output of Table 6 in consideration of spacing, but can be found in Appendix 7.2. With time trends controlled for, the statistical significance of *complexity* still holds, but decreases to the 5 percent level. The coefficient also decreases again to 0.1022. One standard deviation increase in *complexity* increases *diff* by 0.0053 ($0.052 * 0.1022$), or 16 percent ($0.0053/0.033$) of its standard deviation. *Lmk_cap* is statistically significant at the 0.1 percent level with a coefficient of -0.0168. *Lturnover* is statistically significant at the 1 percent level with a coefficient of 0.0111. The effects of *guidance* and *avg_exp* become miniscule with no statistical significance. We end with an R^2 value of 0.3516 meaning some of the unexplained variation was indeed picked up by the year dummy variables.

Through these four models, Hypothesis 1 is affirmed. The higher a firm's complexity is, the larger the difference will be between GAAP earnings and equity analyst adjusted non-GAAP earnings. This relationship is statistically significant at the 5 percent level in a cluster robust regression with firm fixed effects controlled for other studied phenomenon and time trends.

5.2.1 Cross-Sectional Test

In the cross sectional test, I include the continuous firm earnings volatility variable in Model 1. Since volaility is fixed for each firm, the regression is run without firm fixed effects as given by the following equation:

$$\begin{aligned} diff_{i,t} = {} & \beta_1 complexity_{i,t} + \beta_2 lmk\_cap_{i,t} + \beta_3 guidance_{i,t} + \beta_4 avg\_exp_{i,t} \\ & + \beta_5 lturnover_{i,t} + \beta_{6-25} dummy(t) + \beta_{26} vol_i + \beta_{27} vol_i * complexity_{i,t} + \epsilon_{i,t} \end{aligned}$$

In Model 2, I instead use the binary volatility variable split into high and low volatility by median volatility. Like the previous model, the regression is run without firm fixed effects.

$$\begin{aligned} diff_{i,t} = {} & \beta_1 complexity_{i,t} + \beta_2 lmk\_cap_{i,t} + \beta_3 guidance_{i,t} + \beta_4 avg\_exp_{i,t} \\ & + \beta_5 lturnover_{i,t} + \beta_{6-25} dummy(t) + \beta_{26} vol\_bin_i + \beta_{27} vol\_bin_i * complexity_{i,t} + \epsilon_{i,t} \end{aligned}$$

Table 7 features the regression output for these two models. Most specifically I will focus on the effects of *vol*, *vol_bin*, and their interaction terms with complexity. In Model 1 the individual volatility variable has a coefficient of -0.0555 but is not statistically significant. However when it is interacted with complexity, the coefficient becomes positive at 0.4455 and is statistically significant at the 5 percent level. Model 2 shares the same behavior. The binary volatility variable has a negative relationship with a coefficient of -0.0105 and is not statistically significant. When interacted with complexity, the coefficient is 0.0768 and is statistically significant at the 5 percent level. *Complexity* and *lmk_cap* remain statistically significant at the 0.1 percent level in these models.

Marginally, this means that a 1 percent increase in *complexity* will increase *diff* by 0.00155 (0.00078 + 0.00077) for high volatility firms compared to 0.00078 (0.00078 + 0) for low volatility firms. Practically, one standard deviation increase in *complexity* when

	Model 1	Model 2
(Intercept)	0.0241	0.0241
	(0.0135)	(0.0150)
complexity	0.0856***	0.0778***
	(0.0165)	(0.0156)
lmk_cap	−0.0033***	−0.0036***
	(0.0009)	(0.0010)
guidance	0.0010	0.0014
	(0.0013)	(0.0013)
avg_exp	0.0005	0.0006
	(0.0005)	(0.0005)
lturnover	0.0054*	0.0062*
	(0.0022)	(0.0026)
vol	−0.0555	
	(0.0720)	
complexity:vol	0.4455*	
	(0.2162)	
vol_bin		−0.0105
		(0.0094)
complexity:vol_bin		0.0768*
		(0.0354)
R^2	0.1532	0.1248
Adj. R^2	0.1469	0.1182
Num. obs.	3639	3639
RMSE	0.0308	0.0313
N Clusters	198	198

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 7: Cross Sectional Test Models

volatility is high increases *diff* by 0.008 $((0.0778 * 0.052) + (0.0768 * 0.052 * 1))$, or 24 percent (0.008/0.033) of its standard deviation. When volatility is low, one standard deviation increase in *complexity* increases *diff* by 0.004 $(0.0778 * 0.052)$, or 12 percent (0.004/0.033) of it's standard deviation. High volatility doubles the practical effect of *complexity* on *diff*.

The findings of these models affirms Hypothesis 2. The effect of *complexity* on *diff* is higher in firms with high earnings volatility compared to firms with low earnings volatility. This cross-sectional test serves to strengthen the relationship between *complexity* and *diff*.

6 Conclusion

US GAAP is a standardized set of rules which seeks to improve the clarity, consistency, and comparability of financial information. As such, all publicly traded companies listed on an exchange within the United States are mandated to follow GAAP reporting guidelines in their financial reports. However, there are many groups and individuals who look to report firm earnings in a way that differs from GAAP. One of such groups is equity analysts, who report adjusted non-GAAP earnings (street earnings) measures that they feel better represent the true conditions of a firm. Previous literature has readily established the importance of analyst adjusted non-GAAP earnings (Bradshaw and Sloan 2002; Brown and Sivakumar 2003). Although the literature has also found phenomenon that have affected the determination of these earnings (Gu and Chen 2004; Baik, Farber, and Petroni 2009; Christensen et al. 2011), the process is still obscure and less understood.

In order to better understand this process, I introduce a text based measure of firm complexity as defined by Loughran and McDonald (2020). Complexity is calculated as the percentage of unique occurrences of complexity indicating words in a certain 10-K document to a complete lexicon of 374 complexity indicating words. As complexity increases, the harder it becomes to disentangle the interactions between different parts of a firm, therefore decreasing clarity and predictability. Given this conceptual background, it is expected that equity analysts will face complexity in the process of determining their non-GAAP figures.

The results of my models show that complexity does indeed have a significantly positive effect on the difference between equity analyst non-GAAP earnings and GAAP earnings. That is, if the complexity for a firm is high, analysts are more likely to restate GAAP earnings figures with their own figures. This relationship is strengthened by a cross sectional test with earnings volatility, as the marginal effects when considering the interaction between complexity and volatility increases the difference furthermore.

Since this measure of complexity was only just recently defined, this thesis is one of the first applications to my knowledge. This has implications for future research as a new readily defined variable that can be used in other contexts. Text based analysis like the complexity measure in this study are a precursor to the field that is natural language processing. The current complexity measure relies on human based decisions to create the 374 word complexity lexicon which can lead to a misrepresentation of "true" complexity. Although it surpassed it scope of this study, using machine natural language processing could provide a more objective definition of complexity. The problem to tackle then would be to accurately create a set of training data as complexity itself does not have an objective measure. I do believe that with time even more sophisticated measures of such phenomenon will develop and lead to greater insights.

7 Appendix

7.1 Complexity Data

Find complexity data here at: `https://github.com/chenpatrickc/thesisdata`

7.2 Additional Regression Output

	Model 1	Model 2	Model 3	Model 4
(Intercept)	-0.0161^{**}	0.0097		
	(0.0051)	(0.0107)		
complexity	0.1203^{***}	0.1248^{***}	0.1132^{**}	0.1022^{*}
	(0.0203)	(0.0206)	(0.0351)	(0.0439)
lmk_cap		-0.0033^{**}	-0.0129^{***}	-0.0168^{***}
		(0.0010)	(0.0023)	(0.0028)
guidance		-0.0017	0.0002	-0.0002
		(0.0014)	(0.0012)	(0.0011)
avg_exp		0.0005	0.0012^{**}	0.0001
		(0.0004)	(0.0004)	(0.0004)
lturnover		0.0080^{***}	0.0092^{***}	0.0111^{**}
		(0.0023)	(0.0026)	(0.0034)
dummy2000				-0.0151
				(0.0094)
dummy2001				-0.0149
				(0.0087)
dummy2002				-0.0195^{*}
				(0.0080)
dummy2003				-0.0260^{***}
				(0.0075)
dummy2004				-0.0246^{***}
				(0.0069)
dummy2005				-0.0249^{***}
				(0.0068)
dummy2006				-0.0231^{***}
				(0.0065)
dummy2007				-0.0248^{***}

	Model 1	Model 2	Model 3	Model 4
				(0.0063)
dummy2008				-0.0157^{*}
				(0.0071)
dummy2009				-0.0281^{***}
				(0.0069)
dummy2010				-0.0254^{***}
				(0.0067)
dummy2011				-0.0258^{***}
				(0.0065)
dummy2012				-0.0205^{**}
				(0.0065)
dummy2013				-0.0194^{**}
				(0.0059)
dummy2014				-0.0170^{**}
				(0.0060)
dummy2015				-0.0147^{*}
				(0.0061)
dummy2016				-0.0131^{*}
				(0.0060)
dummy2017				-0.0037
				(0.0059)
dummy2018				-0.0111
				(0.0058)
dummy2019				-0.0101
				(0.0058)
R^2	0.0353	0.0761	0.3229	0.3516
Adj. R^2	0.0350	0.0749	0.2829	0.3093
Num. obs.	3658	3640	3640	3640
RMSE	0.0326	0.0320	0.0282	0.0277
N Clusters	200	199	199	199

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

7.3 Observation Frequency Tables

	Year	n	prop
1	2000	83	0.02
2	2001	113	0.03
3	2002	178	0.05
4	2003	178	0.05
5	2004	180	0.05
6	2005	181	0.05
7	2006	185	0.05
8	2007	188	0.05
9	2008	188	0.05
10	2009	190	0.05
11	2010	188	0.05
12	2011	192	0.05
13	2012	194	0.05
14	2013	191	0.05
15	2014	187	0.05
16	2015	194	0.05
17	2016	195	0.05
18	2017	195	0.05
19	2018	196	0.05
20	2019	199	0.05
21	2020	45	0.01

Table 9: Frequency tables of observations by year

	Sector	n	prop
1	Communication Services	81	0.02
2	Consumer Discretionary	420	0.12
3	Consumer Staples	350	0.10
4	Energy	187	0.05
5	Financials	625	0.17
6	Health Care	355	0.10
7	Industrials	560	0.15
8	Information Technology	413	0.11
9	Materials	260	0.07
10	Real Estate	19	0.01
11	Utilities	370	0.10

Table 10: Frequency tables of observations by sector

8 References

Abarbanell, J. and R. Lehavy (2007). Letting the “tail wag the dog”: The debate over GAAP versus street earnings revisited. *Contemporary Accounting Research* 24, pp. 675–723.

Ashbaugh-Skaife, H., D. W. Collins, and R. Lafond (2009). The effect of SOX internal control deficiencies on firm risk and cost of equity. *Journal of Accounting Research* 47, pp. 1–43.

Baik, B., D. Farber, and K. Petroni (2009). Analysts’ incentives and street earnings. *Journal of Accounting Research* 47, pp. 45–69.

Bradshaw, M. and R. Sloan (2002). GAAP versus the street: An empirical assessment of two alternative definitions of earnings. *Journal of Accounting Research* 40, pp. 41–66.

Brown, L. and K. Sivakumar (2003). Comparing the value relevance of two operating income measures. *Review of Accounting Studies* 8, pp. 561–572.

Chang, H. S., M. Donohoe, and T. Sougiannis (2016). Do analysts understand the economic and reporting complexities of derivatives? *Journal of Accounting and Economics* 61, pp. 584–604.

Christensen, T. et al. (2011). Do managers use earnings guidance to influence street earnings exclusions? *Review of Accounting Studies* 16, pp. 501–527.

Cohen, L. and D. Lou (2012). Complicated firms. *Journal of Financial Economics* 104, pp. 383–400.

Dichev, I. and V. Tang (2009). Earnings volatility and earnings predictability. *Journal of Accounting and Economics* 47, pp. 160–181.

Doyle, J., W. Ge, and S. McVay (2007). Determinants of weaknesses in internal control over financial reporting. *Journal of Accounting and Economics* 44, pp. 193–223.

Dyer, T., M. Lang, and L. Stice-Lawrence (2017). The evolution of 10-K textual disclosure: Evidence from latent Dirichlet allocation. *Journal of Accounting and Economics* 64, pp. 221–245.

Ge, W. and S. McVay (2005). The disclosure of material weaknesses in internal control after the Sarbanes-Oxley Act. *Accounting Horizons* 19, pp. 137–158.

Gomes, A., G. Gorton, and L. Madureira (2007). SEC regulation fair disclosure, information, and the cost of capital. *Journal of Corporate Finance* 13, pp. 300–334.

Gu, Z. and T. Chen (2004). Analysts' treatment of nonrecurring items in street earnings. *Journal of Accounting and Economics* 38, pp. 129–170.

Heflin, F., C. Hsu, and Q. Jin (2015). Accounting conservatism and street earnings. *Review of Accounting Studies* 20, pp. 674–709.

Hoitash, R. and U. Hoitash (2018). Measuring accounting reporting complexity with XBRL. *Accounting Review* 93, pp. 259–287.

Lehavy, R., F. Li, and K. Merkley (2011). The effect of annual report readability on analyst following and the properties of their earnings forecasts. *Accounting Review* 86, pp. 1087–1115.

Loughran, T. and B. McDonald (2014). Measuring readability in financial disclosures. *Journal of Finance* 69, pp. 1643–1671.

— (2020). Measuring firm complexity. *Working Paper*. DOI: `https://dx.doi.org/10.2139/ssrn.3645372`.