

Extension Analysis

Lauren Goodwin

Advisor: Steve Cherry

Spring 2015

Contents

1	Introduction and Background	2
1.1	Filing Basics	2
2	Objectives and Questions	3
3	Methods	4
3.1	Data Collection	4
3.2	Preliminary Analysis	4
3.2.1	Extension rate by taxonomy	6
3.2.2	Extension rate by year	7
3.2.3	Extension rate by filing status	7
3.2.4	Interactions	8
3.3	Analysis	9
4	Results	9
4.1	10-K Model	9
4.1.1	Model Diagnostics	12
4.2	10-Q Model	13
4.2.1	Model Diagnostics	15
4.3	10-Q Model with Quarterly Data	15
5	Conclusion	16
5.1	Scope of Inference	16
6	Future Analysis	17
7	Acknowledgments	18

1 Introduction and Background

One of the basic features and strengths of eXtensible Business Reporting Language (XBRL) is the ability to create custom elements or unique tags, referred to as extensions. With the U.S. Securities and Exchange Commission's (SEC) implementation of XBRL, filers may create extensions within certain guidelines. However, it has been observed that filers create extensions which are materially identical to existing standard elements. Unnecessary extensions impair the usability of XBRL data by introducing meaningless variability. A benefit of XBRL is that it enables company comparisons by creating a data set shared at least two companies. However, if a company creates an extension then direct comparison of the company to others is difficult. For instance, if a company files an extended element that is identical to existing standard elements then this company's element can not be directly compared to other companies who filed the existing standard element. So comparison between company filings is not possible.

Several studies have shown that approximately 50% of extensions are unnecessary. A recent SEC staff observation found that a large portion of extensions are reasonable and appropriate, and the SEC has stated that extension rates will be reduced as filers learn more about XBRL. However, Workiva's marketing innovation analysis team suggests this is not the case. Workiva created and maintains Wdesk, which is a cloud-based platform for companies to collect, manage, report, and analyze business data.

This analysis seeks to understand the trend of extensions over time and the influential variables of extensions.

1.1 Filing Basics

"XBRL, eXtensible Business Reporting Language, is an XML-based markup language used to communicate financial and business data electronically."^[3] XBRL is a reporting language that encodes financial statements with context so that information is computer-readable. These statements can then be read automatically and managed more efficiently. For example, a company wants to report net sales of 131,383. However, more information is required, such as currency, scaling, over what time period, etc. XBRL will tag net sales of 131,383 along with the required context. To be able to tag items the user needs a set of defined elements, referred to as a XBRL US Generally Accepted Accounting Practices (GAAP) taxonomy. When preparing a financial statement, the company will use a XBRL US GAAP taxonomy to match each piece of information to an element. An element represents the financial reporting concept. (Figure 1)

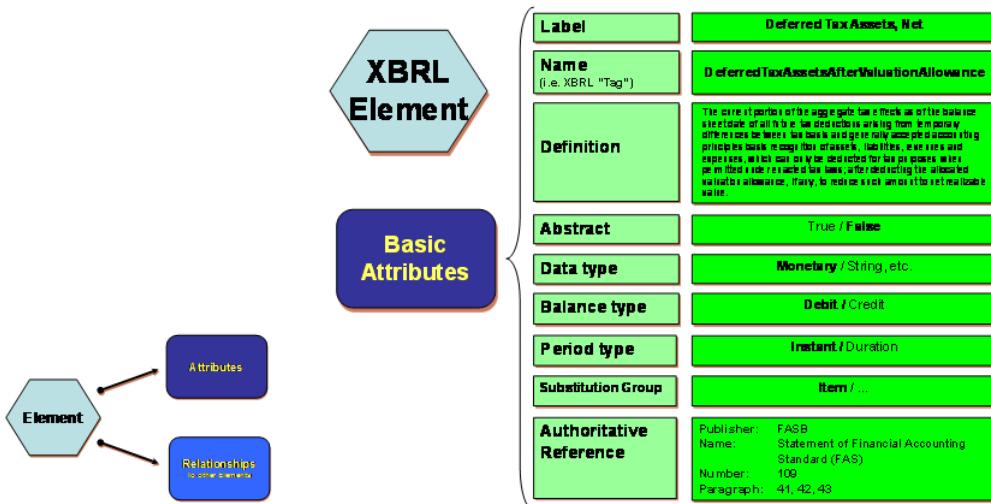


Figure 1: An element and it's break down

The matching process is referred to as mapping. During this mapping process, preparers are able to *extend*

the US GAAP taxonomy by creating new elements. There are several reasons for extending the taxonomy. For example, an element a company would like to use to tag a piece of their financial information may not exist within the US GAAP taxonomy, so XBRL allows the preparer to create a custom element. The extension then must be validated as well as the entire financial report.

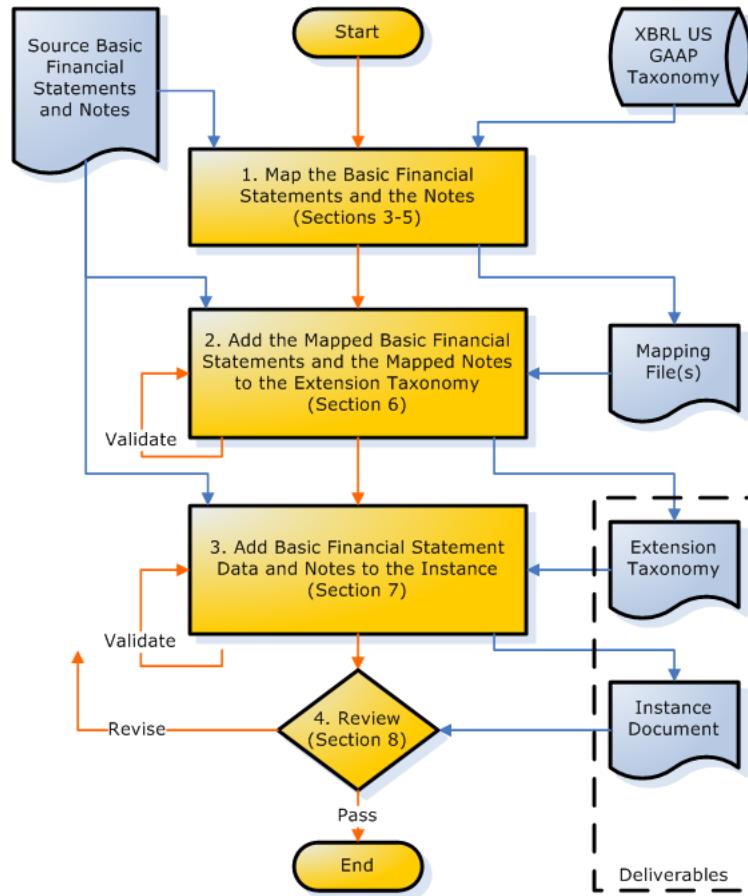


Figure 2: XBRL financial report process

When a company extends an element direct comparisons of two companies are impossible. Not all extensions are necessary and some filers are aware they are unnecessary but extend anyways.

2 Objectives and Questions

Objectives:

- To analyze extension rates in XBRL filings across various dimensions to discover trends in rates over time.

Questions:

- What factors have a strong association with extensions?
- Are extension rates changing over time?

3 Methods

3.1 Data Collection

Data were collected from all companies publicly traded in the United States. The data were gathered from the SEC via FTP and RSS feed. The data are initially encoded in XBRL. Once obtained the data are put into a PostgreSQL database based on the Arelle data model (Figure 3)(the code can be provided upon request).

The data contains information on filings from years 2008 to 2014. However, not all companies were required to file until the year 2012. Accordingly, only data from 2012 to 2014 will be analyzed. Furthermore, the implementation of XBRL initially was rough so the data are inconsistent in earlier years.

The response variable is defined as,

$$\text{extension rate} = \frac{\text{number of extended elements}}{\text{total elements}}$$

Where *total elements* is the sum of the counts of extended elements and the counts of standard elements. Extended elements are defined as elements that were custom made due to the element not existing or other reasons. A Standard element is defined as an element that is recognized by GAAP. Therefore, we can consider an element either extended or not (a standard element). Thus extension rate is a proportion based on binary counts.

The following explanatory variables were initially collected:

- Standard Industry Code: Four digit code to classify industry area of the company. Classification was based on the SIC code website. For example Safeway would be classified as retail and thus have a SIC code associated with retail.[4]
- Creation Software: Software used to create filing; for example Workiva (formally known as Webfilings) Wdesk products.
- Filer status: large accelerated filers (LAF, $n = 27275$), accelerated filers (AF, $n = 18182$), non accelerated filers (NAF, $n = 10351$), and small reporting companies (SRC, $n = 45053$).
- Year: Calendar year of filing 2012 ($n = 32315$), 2013 ($n = 31213$), or 2014 ($n = 15152$).
- Taxonomy: Taxonomy year used for filing. Taxonomy is defined as a set of rules that a company must abide by when filing. A company can choose to file with the current taxonomy year or the past two versions. For example in 2014 companies could file a 2014, 2013, or 2012 taxonomy. In this data set there are taxonomy years 2009 ($n = 7938$), 2011 ($n = 32740$), 2012 ($n = 36351$), 2013 ($n = 23786$), and 2014 ($n = 26$). The 2010 taxonomy does not exist.
- Form Type: If filing a 10-K ($n = 21826$) or 10-Q ($n = 67604$). There are other form types, this analysis is only concerned with these two.

3.2 Preliminary Analysis

The preliminary stage of the analysis was to plot the raw data to get an idea of what trends were apparent and to get an idea of what explanatory variables were to be included in the model. It was already determined that there would be two models since 10-K are only filed once a year and 10-Q filed quarterly. So the 10-K summarizes information from the 10-Q filings.

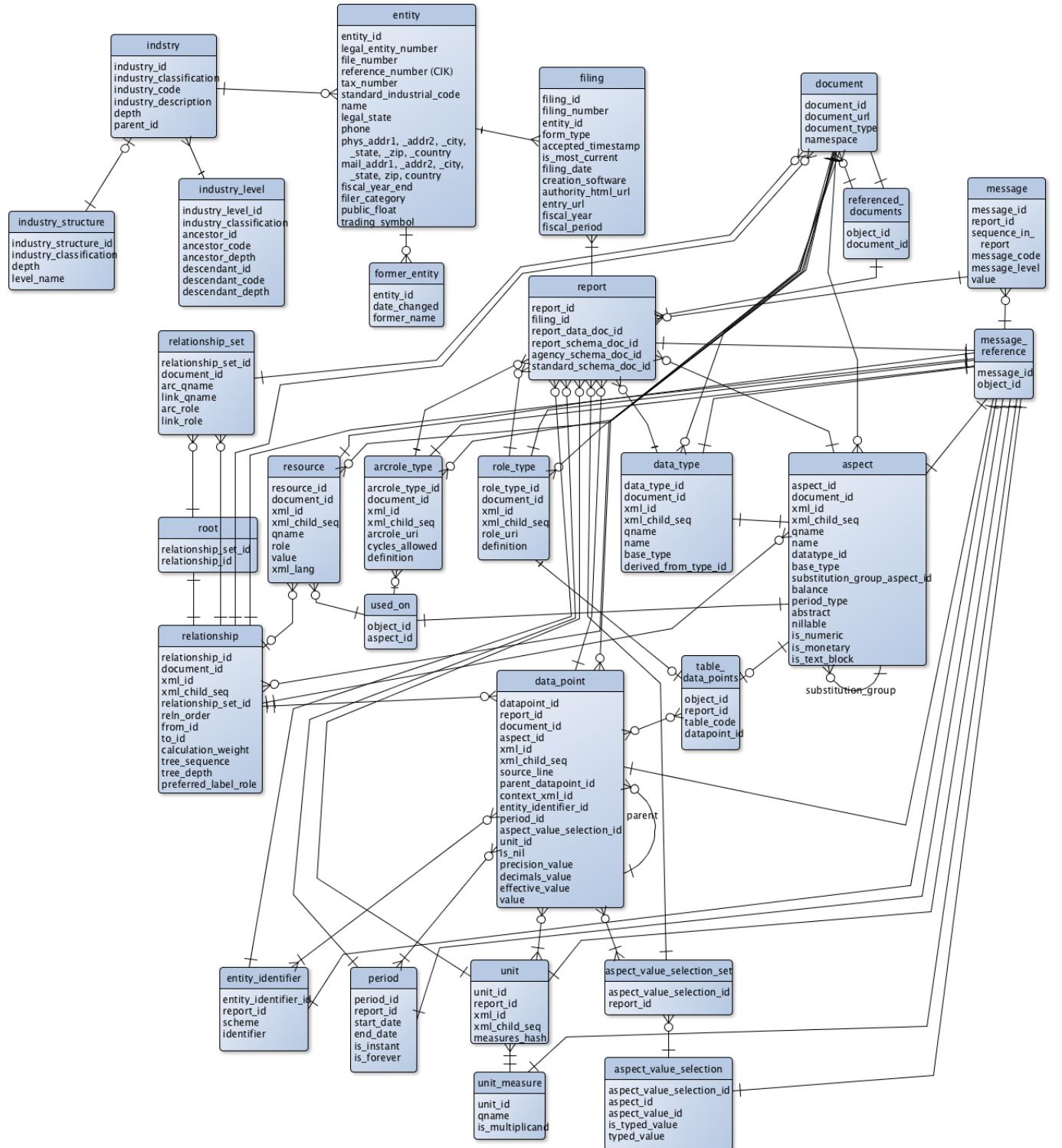


Figure 3: Arelle data model

3.2.1 Extension rate by taxonomy

There appears to be an decreasing trend in extension rate as taxonomy year increases from taxonomy 2012 to 2014 for both the 10-K and 10-Q forms.(Figure 4) When the plot is separated by the four filer categories we do not see one consistent trend. In addition, the NAF filer category at taxonomy 2009 only has one data point. This data point is an outlier and since linear regression is not resistant to outliers this point potentially could alter the estimated model coefficients. Therefore, this point will be removed.

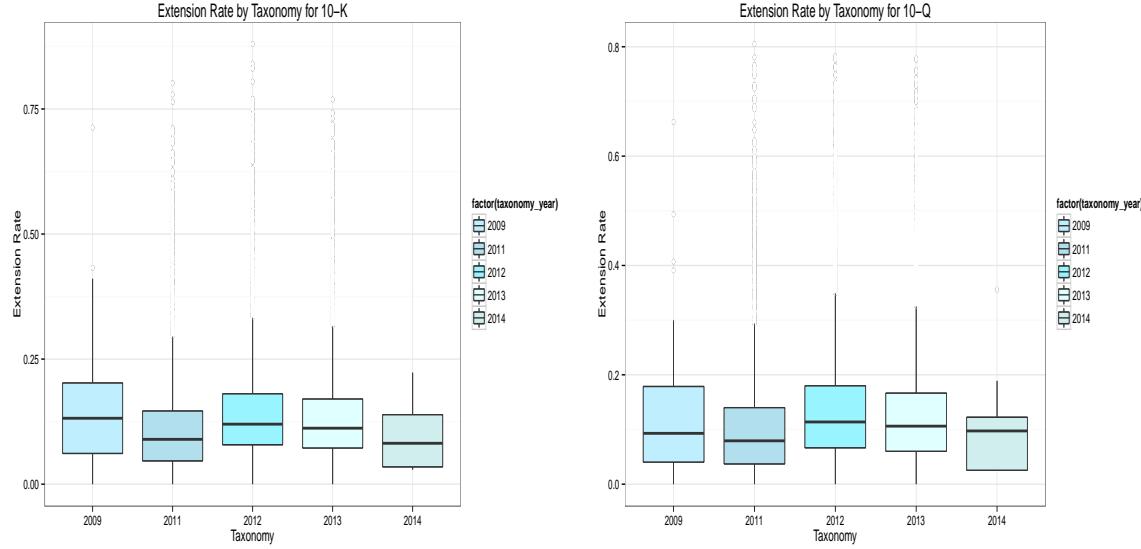


Figure 4: Extension rate by taxonomy for 10-K and 10-Q

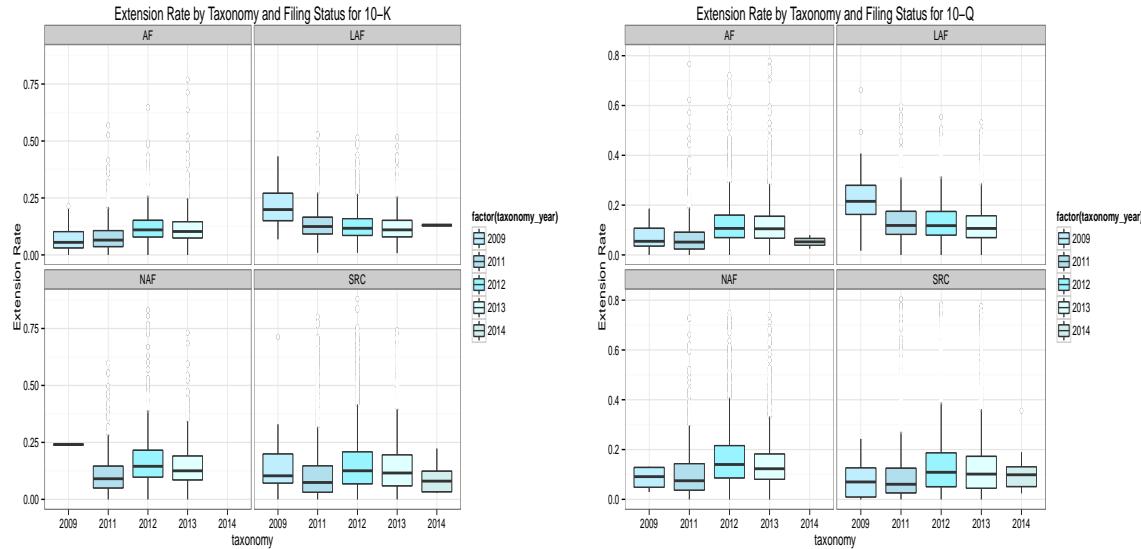


Figure 5: Extension rate by filing status and taxonomy for 10-K and 10-Q

3.2.2 Extension rate by year

There does not appear to be an overwhelming trend in extension rate as year increases for both the 10-K and 10-Q forms.(Figure 6) However, if the plot is separated by the four filer categories we see that extension rates are remaining roughly constant as year increases.(Figure 7)

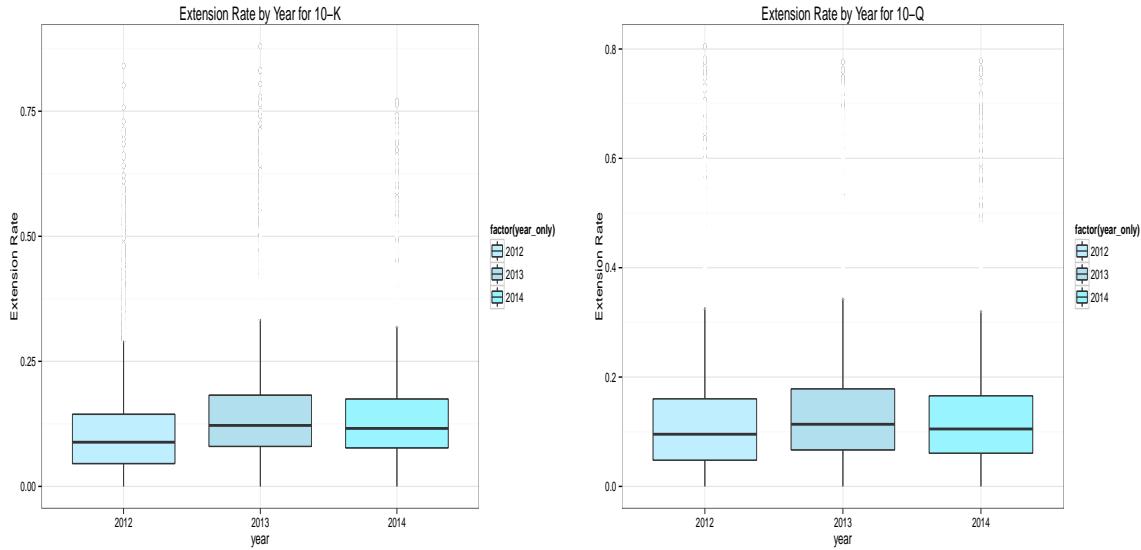


Figure 6: Extension rate by year for 10-K and 10-Q

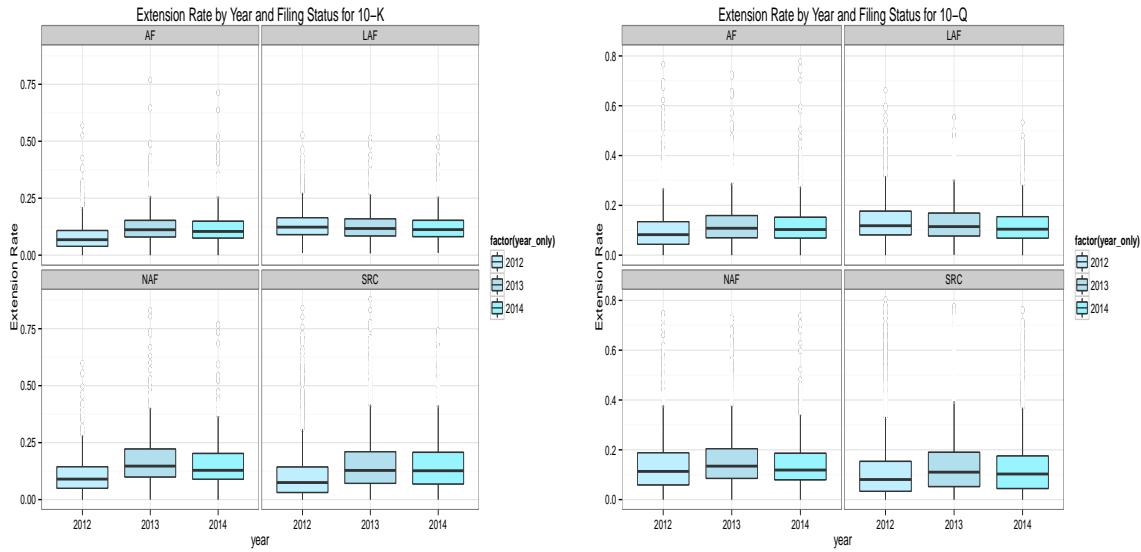


Figure 7: Extension rate by filing status and year for 10-K and 10-Q

3.2.3 Extension rate by filing status

It appears that all four categories have around the same rates of extension for both 10-K and 10-Q.(Figure 8)

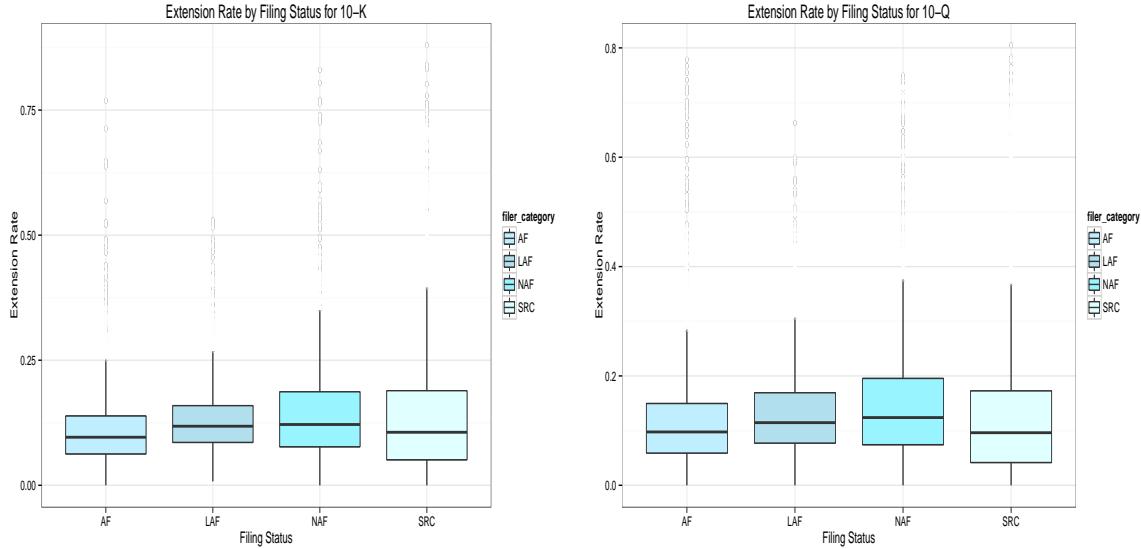


Figure 8: Extension rate by filing status for 10-K and 10-Q

3.2.4 Interactions

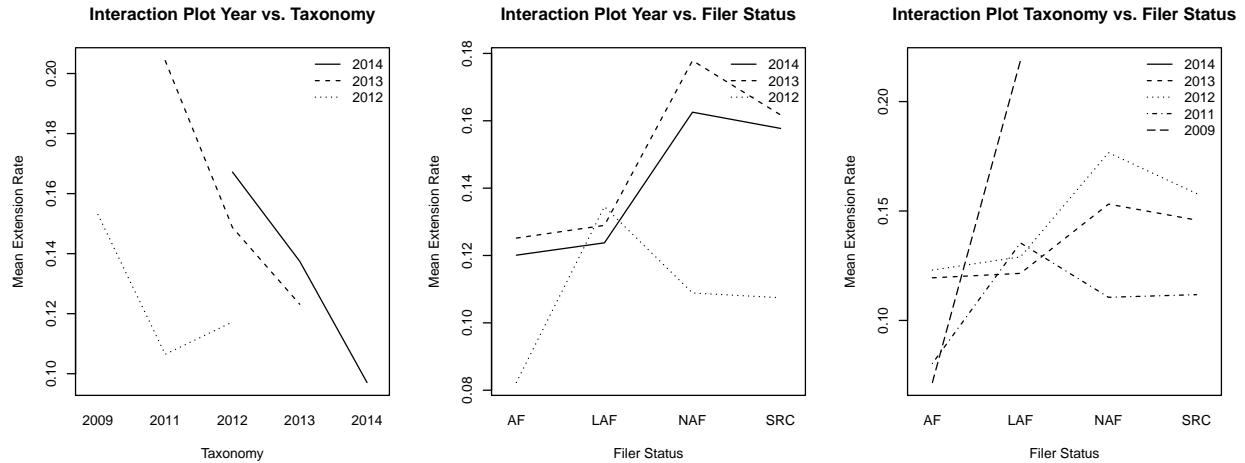


Figure 9: Interaction plots

To determine if interaction terms are needed we will use the plots (Figure 9) to investigate. From the above interaction plot (the first plot) we see that interactions between year and taxonomy is not needed. The need for interactions is indicated by the lack of parallel lines. The lines in the first plot appear to be parallel enough that an interaction term between year and taxonomy is not necessary. In other words, this indicates that year does not depend on taxonomy year and vice versa. In the other two above interaction plots there is an obvious lack of parallelism thus interaction terms for taxonomy and filer status, and year and filer status will need to be included in the model.

3.3 Analysis

For the analysis we consider multiple linear regression with a logit transformation on the response, extension rate. The benefit of this transformation is that it takes the data constrained in the interval 0 to 1, and transforms the data to values between negative infinity and positive infinity. The data are transformed by $\text{logit}(\text{extension_rate}) = \log(\text{extension_rate}/(1 - \text{extension_rate}))$, which is the logarithm of the ratio of proportion of extended elements to proportion of non extended elements. We could define this ratio as the extension ratio. Extension ratio will provide the same type of information as extension rate.

Then, we proceed with multiple linear regression, using $\log(\text{extension_ratio})$ as the response under the assumption of normality. Interpretation is then done as usual for a log transformation of the response.

Since 10-K are only filed once a year and 10-Q filed quarterly then 10-K summarizes information from the 10-Q filings. So the data for each will be considered two separate data sets. Thus, two models will be considered. Both models will have the structure:

$$\begin{aligned}\log(\text{extension_ratio}_i) &= \\ \beta_0 + \beta_1 \text{taxonomy} + \beta_2 \text{year} + \beta_3 \text{filer_status} + \beta_4 \text{taxonomy} * \text{filer_status} + \beta_5 \text{year} * \text{filer_status} + \epsilon_i\end{aligned}$$

where $\epsilon_i \sim \text{IIDN}(0, \sigma^2)$ and $\text{extension_ratio} = \frac{\text{extension_rate}}{1 - \text{extension_rate}}$

Note, that the bolded coefficients correspond to a vector of coefficients associated with the levels of the categories within that explanatory variable.

Initially, logistic regression for binomial count data was used to fit the data. However, there were issues with the fit. The model resulted in standardized deviance residuals as large as 15,000. The large residuals indicated a serious lack of fit, thus a different approach was considered.

4 Results

When fitting a log transformed response the estimated coefficients are on an additive scale. Therefore, the coefficients can be exponentiated to get back to extension ratio. However, back transforming to the extension ratio scale the estimated coefficients are now on an multiplicative scale and are discussed in terms of changes in the median extension ratio.

4.1 10-K Model

The following model is fit using 10-K form data.

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-1.9586	0.0227	-86.23	0.0000
factor(year_only)2013	0.0760	0.0553	1.37	0.1696
factor(year_only)2014	0.1548	0.0728	2.13	0.0336
factor(tax_year)2009	0.5901	0.1884	3.13	0.0017
factor(tax_year)2012	-0.1422	0.0553	-2.57	0.0102
factor(tax_year)2013	-0.2882	0.0739	-3.90	0.0001
factor(tax_year)2014	-0.0931	0.8802	-0.11	0.9157
filer_catAF	-0.7746	0.0336	-23.07	0.0000
filer_catNAF	-0.4549	0.0426	-10.68	0.0000
filer_catSRC	-0.4939	0.0286	-17.24	0.0000
factor(year_only)2013:filer_catAF	0.2552	0.0860	2.97	0.0030
factor(year_only)2014:filer_catAF	0.1949	0.1100	1.77	0.0764
factor(year_only)2013:filer_catNAF	0.6129	0.1085	5.65	0.0000
factor(year_only)2014:filer_catNAF	0.7167	0.1424	5.03	0.0000
factor(year_only)2013:filer_catSRC	0.4908	0.0650	7.55	0.0000
factor(year_only)2014:filer_catSRC	0.6515	0.0864	7.54	0.0000
factor(tax_year)2009:filer_catAF	-0.6409	0.2767	-2.32	0.0206
factor(tax_year)2012:filer_catAF	0.4804	0.0862	5.57	0.0000
factor(tax_year)2013:filer_catAF	0.5321	0.1136	4.68	0.0000
factor(tax_year)2012:filer_catNAF	0.1994	0.1087	1.83	0.0667
factor(tax_year)2013:filer_catNAF	-0.0335	0.1474	-0.23	0.8203
factor(tax_year)2009:filer_catSRC	0.0192	0.3082	0.06	0.9503
factor(tax_year)2012:filer_catSRC	0.1745	0.0651	2.68	0.0074
factor(tax_year)2013:filer_catSRC	-0.0663	0.0878	-0.76	0.4502
factor(tax_year)2014:filer_catSRC	-0.8257	0.9418	-0.88	0.3807

Table 1: Summary of fit for 10-K data with response log(Extension Ratio)

	Taxonomy 2009	Taxonomy 2011	Taxonomy 2012	Taxonomy 2013	Taxonomy 2014
LAF	0.25	0.14	0.12	0.11	0.13
AF	0.06	0.07	0.09	0.08	
NAF		0.09	0.09	0.06	
SRC	0.16	0.09	0.09	0.06	0.03

Table 2: Extension ratio by taxonomy and filer status

	2012	2013	2014
LAF	0.14	0.15	0.16
AF	0.07	0.09	0.09
NAF	0.09	0.18	0.21
SRC	0.09	0.15	0.19

Table 3: Extension ratio by year and filer status

Since we are working with census data the following summary of the results do not include standard errors. However, if the model is used to predict extension ratios we need to incorporate error.

Tables 2 and 3 give the fitted extension ratios from the model. To obtain these fitted extension ratios a linear combination is taken from the estimated coefficients from the model and then exponentiated. For example, to obtain the fitted extension ratio for AF in taxonomy 2012,

$$\text{extension_ratio}_{AF,\text{tax}2012} = \exp\{-1.9586 + -0.1422 + -0.7746 + 0.4804\} = 0.0912.$$

For LAF the estimated median extension ratio for taxonomy 2011 in year 2012 is 0.14. The median extension ratio for taxonomy 2009 is 1.8043 times larger than taxonomy 2011. The median extension ratio for taxonomy 2012 is 0.8674 times smaller than taxonomy 2011. The median extension ratio for taxonomy 2013 is 0.7496 times smaller than taxonomy 2011. The median extension ratio for taxonomy 2014 is 0.9111 times smaller than taxonomy 2011. Therefore, we see a decreasing trend in the extension ratio as taxonomy increases. However, the extension ratio increased in taxonomy 2014 compared to taxonomy 2013. This may be due to the small sample size of 26 in taxonomy 2014 so once more filers use taxonomy 2014 we may see the same decreasing trend. The median extension ratio for year 2013 is 1.0789 times larger than the extension ratio in year 2012. The median extension ratio for year 2014 is 1.1674 times larger than the extension ratio in year 2012. We see a slight increasing trend in extension ratio as year increases.

For AF the estimated median extension ratio for taxonomy 2011 in year 2012 is 0.07. The median extension ratio for taxonomy 2009 is 0.9505 times smaller than taxonomy 2011. The median extension ratio for taxonomy 2012 is 1.4024 times larger than taxonomy 2011. The median extension ratio for taxonomy 2013 is 1.2762 times larger than taxonomy 2011. There is slight increasing trend in the extension ratio as taxonomy increases. The median extension ratio for year 2013 is 1.3926 times larger than the extension ratio in year 2012. The median extension ratio for year 2014 is 1.4186 times larger than the extension ratio in year 2012. From year 2012 to 2013 extension ratio increases but from years 2013 to 2014 the extension ratio stays about the same.

For NAF the estimated median extension ratio for taxonomy 2011 in year 2012 is 0.09. The median extension ratio for taxonomy 2012 is 1.0588 times larger than taxonomy 2011. The median extension ratio for taxonomy 2013 is 0.7249 times smaller than taxonomy 2011. From taxonomy years 2011 and 2012 extension ratio is about the same, then from taxonomy 2012 to 2013 extension ratio decreases. The median extension ratio for year 2013 is 1.9914 times larger than the extension ratio in year 2012. The median extension ratio for year 2014 is 2.3905 times larger than the extension ratio in year 2012. There is an increasing trend in extension ratio as year increases.

For SRC the estimated median extension ratio for taxonomy 2011 in year 2012 is 0.09. The median extension ratio for taxonomy 2009 is 1.8392 times larger than taxonomy 2011. The median extension ratio for taxonomy 2012 is 1.0328 times larger than taxonomy 2011. The median extension ratio for taxonomy 2013 is 0.7015 times smaller than taxonomy 2011. The median extension ratio for taxonomy 2014 is 0.399 times smaller than taxonomy 2011. Therefore, we see a slight decreasing trend in the extension ratio as taxonomy increases. The median extension ratio for year 2013 is 1.7626 times larger than the extension ratio in year 2012. The median extension ratio for year 2014 is 2.2395 times larger than the extension ratio in year 2012. There is an increasing trend in extension ratio as year increases.

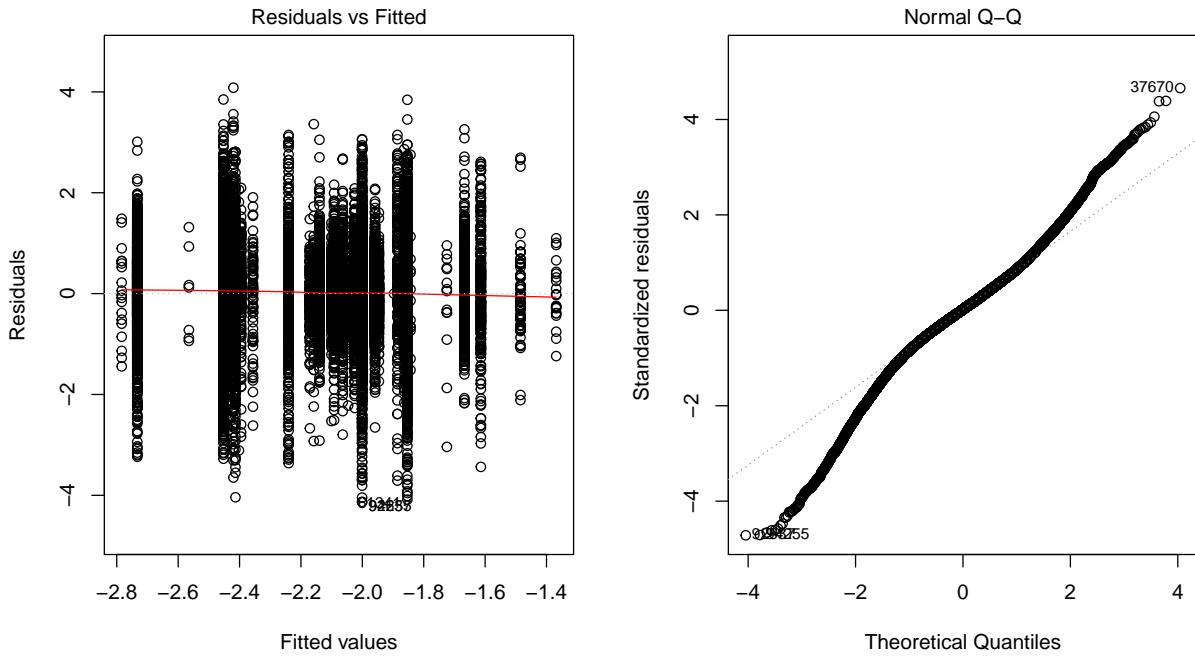


Figure 10: Diagnostic plots for the 10-K model

4.1.1 Model Diagnostics

Multiple linear regression assumes the residuals are normal, independent, and identically distributed, with constant variance. The residuals versus fitted values plot of the 10-K model indicates that the constant variance assumption is adequately met since the variance of the residuals appears to be constant (Figure 9). The normal Q-Q plot indicates some deviation from normality due to the points deviating from the line on the ends (Figure 9). However, since the sample size is large and multiple linear regression is robust it is reasonable to assume the normality of errors assumption is adequately met.

4.2 10-Q Model

The following model is fit using the 10-Q form data.

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-1.3513	0.1855	-7.29	0.0000
factor(year_only)2013	-0.0157	0.0255	-0.62	0.5374
factor(year_only)2014	-0.0643	0.0379	-1.69	0.0901
factor(taxonomy_year)2011	-0.6498	0.1863	-3.49	0.0005
factor(taxonomy_year)2012	-0.6709	0.1866	-3.60	0.0003
factor(taxonomy_year)2013	-0.7555	0.1884	-4.01	0.0001
factor(taxonomy_year)2014	-0.2596	0.3434	-0.76	0.4497
filer_catAF	-1.3908	0.2650	-5.25	0.0000
filer_catNAF	-1.1852	0.4994	-2.37	0.0176
filer_catSRC	-0.9813	0.2711	-3.62	0.0003
factor(year_only)2013:filer_catAF	0.0886	0.0360	2.46	0.0138
factor(year_only)2014:filer_catAF	0.1231	0.0545	2.26	0.0240
factor(year_only)2013:filer_catNAF	0.1077	0.0434	2.48	0.0131
factor(year_only)2014:filer_catNAF	0.0914	0.0656	1.39	0.1638
factor(year_only)2013:filer_catSRC	0.2684	0.0310	8.67	0.0000
factor(year_only)2014:filer_catSRC	0.3086	0.0458	6.73	0.0000
factor(taxonomy_year)2011:filer_catAF	0.4502	0.2666	1.69	0.0912
factor(taxonomy_year)2012:filer_catAF	1.2246	0.2664	4.60	0.0000
factor(taxonomy_year)2013:filer_catAF	1.2533	0.2691	4.66	0.0000
factor(taxonomy_year)2014:filer_catAF	-0.1035	0.7650	-0.14	0.8924
factor(taxonomy_year)2011:filer_catNAF	0.6716	0.5007	1.34	0.1798
factor(taxonomy_year)2012:filer_catNAF	1.3399	0.5004	2.68	0.0074
factor(taxonomy_year)2013:filer_catNAF	1.2909	0.5025	2.57	0.0102
factor(taxonomy_year)2011:filer_catSRC	0.4139	0.2720	1.52	0.1282
factor(taxonomy_year)2012:filer_catSRC	0.7669	0.2721	2.82	0.0048
factor(taxonomy_year)2013:filer_catSRC	0.6161	0.2740	2.25	0.0245

Table 4: Fit for 10-Q Response log(Extension Ratio)

	Taxonomy 2009	Taxonomy 2011	Taxonomy 2012	Taxonomy 2013	Taxonomy 2014
LAF	0.26	0.14	0.13	0.12	0.20
AF	0.06	0.05	0.11	0.11	0.04
NAF	0.08	0.08	0.15	0.14	
SRC	0.10	0.08	0.11	0.08	

Table 5: Extension ratio by taxonomy and filer status

	2012	2013	2014
LAF	0.26	0.25	0.24
AF	0.06	0.07	0.07
NAF	0.08	0.09	0.08
SRC	0.10	0.12	0.12

Table 6: Extension ratio by year and filer status

For LAF the estimated median extension ratio for taxonomy 2009 in year 2012 is 0.2589. The median extension ratio for taxonomy 2011 is 0.5222 times smaller than taxonomy 2009. The median extension ratio for taxonomy 2012 is 0.5112 times smaller than taxonomy 2009. The median extension ratio for taxonomy 2013 is 0.4698 times smaller than taxonomy 2009. The median extension ratio for taxonomy 2014 is 0.7714 times smaller than taxonomy 2009. Therefore, we see a decreasing trend in the extension ratio as taxonomy increases. However, the extension ratio increased in taxonomy 2014 compared to taxonomy 2013. This may be due to the small sample size of 26 in taxonomy 2014 so once more filers use taxonomy 2014 we may see the same decreasing trend. The median extension ratio for year 2013 is 0.9844 times smaller than the extension ratio in year 2012. The median extension ratio for year 2014 is 0.9378 times smaller than the extension ratio in year 2012. Extension ratio remains roughly constant as year increases.

For AF the estimated median extension ratio for taxonomy 2009 in year 2012 is 0.0644. The median extension ratio for taxonomy 2011 is 0.8191 times smaller than taxonomy 2009. The median extension ratio for taxonomy 2012 is 1.7396 times larger than taxonomy 2009. The median extension ratio for taxonomy 2013 is 1.6451 times larger than taxonomy 2009. The median extension ratio for taxonomy 2014 is 0.6955 times smaller than taxonomy 2009. From taxonomy 2009 to 2012 extension ratio decreases, however, from taxonomy 2011 to 2012 extension ratio increases then remains approximately the same for taxonomy 2013. From taxonomy 2013 to 2014 the extension ratio decreases. The median extension ratio for year 2013 is 1.0756 times larger than the extension ratio in year 2012. The median extension ratio for year 2014 is 1.0606 times larger than the extension ratio in year 2012. Extension ratio remains roughly constant as year increases.

For NAF the estimated median extension ratio for taxonomy 2009 in year 2012 is 0.0791. The median extension ratio for taxonomy 2011 is 1.0221 times larger than taxonomy 2009. The median extension ratio for taxonomy 2012 is 1.9522 times larger than taxonomy 2009. The median extension ratio for taxonomy 2013 is 1.7082 times larger than taxonomy 2009. From taxonomy 2009 to 2012 extension ratio is approximately the same, however, from taxonomy 2011 to 2012 extension ratio increases. From taxonomy 2012 to 2013 the extension ratio decreases. The median extension ratio for year 2013 is 1.0963 times larger than the extension ratio in year 2012. The median extension ratio for year 2014 is 1.0275 times larger than the extension ratio in year 2012. Extension ratio remains roughly constant as year increases.

For SRC the estimated median extension ratio for taxonomy 2009 in year 2012 is 0.0970. The median extension ratio for taxonomy 2011 is 0.7899 times smaller than taxonomy 2009. The median extension ratio for taxonomy 2012 is 1.1008 times larger than taxonomy 2009. The median extension ratio for taxonomy 2013 is 0.8699 times smaller than taxonomy 2009. From taxonomy 2009 to 2012 extension ratio decreases, however, from taxonomy 2011 to 2012 extension ratio increases. From taxonomy 2012 to 2013 the extension ratio decreases. The median extension ratio for year 2013 is 1.2875 times larger than the extension ratio in year 2012. The median extension ratio for year 2014 is 1.2768 times larger than the extension ratio in

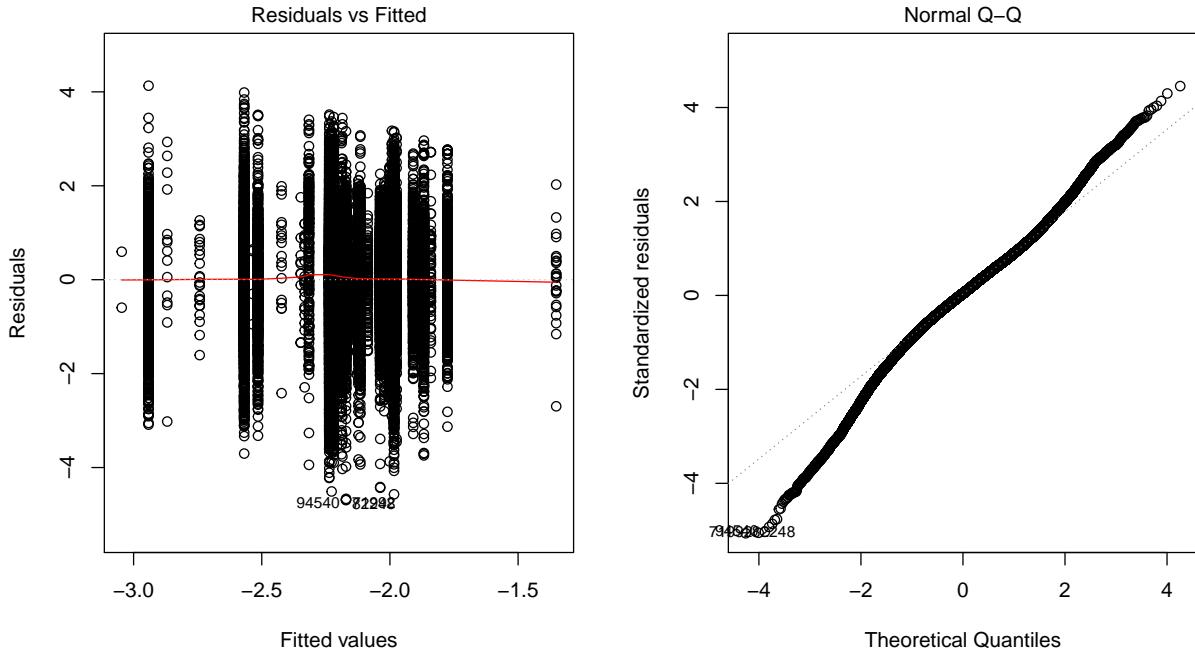


Figure 11: Diagnostic plots for the 10-Q model

year 2012. There is a slight increase in extension ratio from year 2012 to 2013 but from year 2013 to 2014 extension ratio remains roughly the same.

4.2.1 Model Diagnostics

The residuals versus fitted values plot of the 10-Q model indicates that the constant variance assumption is adequately met since the variance of the residuals appears to be constant (Figure 10). The normal Q-Q plot indicates some deviation from normality due to the points deviating from the line on the ends (Figure 10). Since the sample size is large and multiple linear regression is robust it is reasonable to assume the normality of errors assumption is adequately met.

4.3 10-Q Model with Quarterly Data

Results from the 10-K and 10-Q models indicated that the extension ratio is roughly constant over the years. In the 10-Q model, the extension ratio for each year is an aggregated measure over quarters within that year. To get a better grasp of extension ratio over time each year can be broken down into quarters. Quarterly estimates of extension ratio within a year can be obtained. This could give a better indication of a trend of extension ratio over time. For example, it could be that extension ratio remains roughly the same over the first three quarters but spikes during the fourth quarter. Information on the quarterly level is lost when the data is analyzed on the year level. Therefore, the 10-Q model can be fit with a time variable that has information on the year and quarter instead of year alone.

The 10-Q model above is fit with a time variable that indicates the year and quarter. For this paper estimates of extension ratio over years by quarters are only examined for LAF.

The lowest extension ratios occur during the first quarter, in 2012 the estimated extension ratio is .22, 2013

	2012 Q1	2012 Q2	2012 Q3	2012 Q4	2013 Q1	2013 Q2	2013 Q3	2013 Q4	2014 Q1	2014 Q2
LAF	0.22	0.27	0.28	0.29	0.22	0.28	0.29	0.30	0.23	0.29

Table 7: Extension ratio by Year and Quarter

it is .22, and in 2014 it is .23. Extension ratio slightly increases as quarter increases. (Table 7) The ratio is the highest during the fourth quarter, in 2012 extension ratio is estimated to be .29, in 2013 it is .3, and we do not have data on the fourth quarter for 2014. (Table 7) For future analysis, estimates for extension ratio over years by quarters should be obtained for AF, NAF , and SRC.

5 Conclusion

The factors which impact extension rate is form type, year, taxonomy year, filer status, and the interactions between year, taxonomy, and filer status. With the 10-K data, we conclude that for all four filer status types there is an slight increasing trend in extension ratio as year increases. In addition, for LAF and SRC there is a decreasing trend in extension ratio as taxonomy year increases. With the 10-Q data, we conclude that for LAF there is a decreasing trend in extension ratio as taxonomy year increases. However, for AF and NAF extension ratio increases as taxonomy year increases. For SRC extension ratio is approximately the same for all taxonomy years. In addition, for all four filer status types as year increases extension ratio remains approximately constant. However, when year is broken down by quarter extension ratio is the lowest during the first quarter and increases the remaining three quarters.

The only factor that is associated with decreasing extension ratios is taxonomy. This would be expected since the SEC attempts to improve upon the last taxonomy when constructing a newer version. In addition, there have been claims that filers are “learning” over time. If filers were learning over time we would expect extension ratios to be decreasing as the years increase regardless of taxonomy. However, this is not the case as year increases extension ratio is either increasing or remaining constant.

To reduce extension rates, careful consideration of taxonomy is required. Furthermore, to reduce extension rates when filing, the most recent version of the taxonomy should be used. The most recent taxonomy contain more and improved standard elements, thus filers are less likely to make extensions. For instance, a filer that uses the most recent taxonomy can tag an element with a standard element, whereas previously they would have to extend the same element if an older taxonomy version was used.

Since extension ratio is remaining close to constant as year increases some intervention may be needed. Intervention should occur during the first quarter since extension ratios are the lowest during this time, and will only increase the remaining three quarters. Possible intervention could be making the process of validating extensions stricter or requiring companies to use an outside source such as Wdesk (Workiva’s filing product) to create filings. In regards to filer status, LAF tend to have the higher extension ratios compared to the other 3 filer categories. Therefore, intervention could be piloted just on LAF to see if intervention is effective.

5.1 Scope of Inference

This is an observational study because explanatory variable values were not randomly assigned to treatments (and could not be). Therefore, it can only be stated that changes in the explanatory variable were associated with changes in the odds of an extension. Data was collected from all publicly traded companies in the United States (a census of filings from publicly traded companies in US) therefore inferences apply to all publicly traded companies United States.

6 Future Analysis

As mentioned in section 4.3 estimates for extension ratio over years by quarters should be obtained for AF, NAF , and SRC. The LAF estimates showed an increasing trend of quarter increases, however, we need to investigate the trends across the other categories. In addition, we can investigate the effects of using software such as Wdesk on extension ratio. If filers using software tend to have lower extension ratios then the SEC could make a requirement to use a software to reduce extension rates.

Results from the 10-K and 10-Q models indicated that the extension ratio across all four filer categories is roughly the same. Within filer category we can separate filers industries indicated by SIC. Extension ratio for each filer category is an aggregated measure over all SICs (or over all industries). To get a better grasp of extension ratio over filer categories each category can be broken down into industries indicated by SIC. Estimates of extension ratio by industry with filer category can be obtained. This could give a better indication of which type of industries tend to make more extension than others.

All the analysis was performed on data submitted in the year 2014. As filings are coming in, analysis of data from 2015 and 2014 10-K will need to be done. Since the 2014 taxonomy had only 26 observations more recent data (which will hopefully have more observations for the 2014 taxonomy) will provide better model estimates for this taxonomy.

7 Acknowledgments

Dr. Steve Cherry - Writing Project Chair
Allen Bross, M.S. - Data Analyst, Workiva
Susan Yount -Workiva
Workiva

References

- [1] Fred Ramsey and Daniel Schafer. *The Statistical Sleuth: A Course in Methods of Data Analysis*. Duxbury Press; 2 edition, 2002.
- [2] Andrew Gelman and Jennifer Hill. *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Cambridge University Press; 1 edition (December 18, 2006).
- [3] XBRL US GAAP Taxonomy Preparers Guide. <http://xbrl.us/preparersguide/Pages/default.aspx>.
- [4] SIC code / NAICS code. *The Leader in SIC and NAICS Codes*. <http://siccode.com/en/>.