# The Determinants of the Length of Annual Reports:

## Evidence from 20-F Filings

Nguyen Thi Thuy Phuong

Akihisa Kimura

### 1. Introduction

According to efficient market theory, stock prices reflect publicly available information, while annual reports are considered as the main source of information. Understanding the crucial role of annual reports to users, SEC and FASB have promulgated a series of rules and instructions for firms to file the annual reports. However, the extant research records the substantial increase in the complexity of annual reports which causes high financial and time cost for investors to analyze (Bloomfield, 2002).

In detail, previous research, as well as reports of professional organizations, has reported a dramatic increase in the number of words in 10-K (KPMG, 2012; Cazier and Pfeiffer, 2016). They showed that operating complexity, redundancy, and firm residual disclosures cause the difference in the length of 10-K. Recently, Li (2017) recorded the rapid increase in the repetitive disclosures of 10-K; however, Li said that repetitive disclosures do not mean less informative like SEC's suggestion. Most of the other research assumes that the increase in the length of annual reports and the decline in the readability of annual reports undoubtedly create negative impacts on users along with stock markets (Li, 2008; Miller, 2010; You and Zhang, 2008, Lee, 2012; Lehavy et al., 2011). This requires SEC and other standard setters to spend more efforts to understand the length of annual reports and what determinants of length of annual reports are.

Continuing the previous research relating to the length of annual reports, our research uses the sample set of 20-F rather than 10-K to test the length of annual reports and its respective determinants. Our findings show some interesting contributions to practice and academic research. In this research, we choose the sample set of the annual reports of foreign firms listed on the US Stock Exchange, which is not a popular selection in previous research. Understanding the length of 20-F forms is expected to widen the knowledge on the length of annual reports. Our research adds one more piece to the whole picture of the significant increase in the length of annual reports occurring not only in 10-K but also in 20-F. The increase in the length of 20-F is more serious. In addition to operating complexity and firm performance, we also record three more new and crucial determinants of length of annual reports, namely volatility, accounting standards, and language. Realizing new determinants help users and standard setters to deeply understand the behaviors of firms in issuing their annual reports.

The article is organized as follows. The following section displays the change in the length of annual

reports as well as the impacts of the increase in the length of annual reports on stock markets and users. Section 3 presents the sample selection and shows how to measure the length of annual reports. Section 4 identifies what determinants of 20-F length are and predicted signs. Section 5 presents our findings and discussion. Finally, Section 6 gives some concluded remarks.

## 2. Literature Review

Each firm has frequent disclosures through the fiscal year with the hope of updating more timely information to investors; among of firms' disclosures, the annual reports contain more price-relevant information (You and Zhang, 2009). According to SEC's requirements, firms must annually file form 10-K for the US firms and form 20-F for the foreign firms listed on the US Stock Exchange, respectively. Those forms are strictly regulated by SEC to protect investors; however, the length of annual reports has significantly increased over time with overloaded information and redundancy (Cazier and Pfeiffer, 2016; Li, 2017).

### 2. 1 The significant increase in the length of annual reports

To make sufficient, accurate information available to investors for making decisions, SEC "requires public companies to disclose meaningful financial and other information to the public. This provides a common pool of knowledge for all investors to use to judge for themselves whether to buy, sell or hold a particular security" (SEC, 2013). Besides, SEC, as well as FASB, force all firms to strictly obey the mandatory rules specifically to the financial statements, related notes, and footnotes. They also concern about the language used in disclosures of firms. They compel the use of plain English in communication to users which emphasizes to "clarity, not brevity" and leads to "increase the length of particular sections" (SEC, 1998). These suggestions of SEC mean that the longer annual reports are, the more information annual reports provide.

A significant increase in the length of annual reports has been recorded by researchers and institutional organizations. Miller (2010) reports that the number of words in 10-K increased from 25,515 in 1995 to 40,579 in 2006. Similarly, KPMG (2012) records a 16% increase in the length of annual reports and a 28% increase in the length of notes in the period of 2004 to 2010. More recently, Cazier and Pfeiffer (2016) present the average number of words in 10-K during the period from 2003 to 2012 at 55,335.

Unfortunately, the rapid increase in the length of annual reports is assumed not to provide more information like SEC's expectation. Li (2017) reports the percentage of information in notes repeated in MD&A gradually rises from 19% in 1995 to over 26% in 2013. Li also presents a higher percentage of repetitive disclosures in the significant accounting policy, litigation and revenue notes. The reasons for the increase in the repetition of disclosures are assumed as the appearance of special events such as new CEO, issued equity or missed prior year's earnings benchmark. In addition, Cazier and Pfeiffer (2016) provide evidence for reasons of the significant increase in the length of 10-K. They showed that the operating complexity, redundant disclosures between SEC and FASB, and residual disclosures of firms cause such increase of 10-K.

The increase in the length of annual reports due to the repetitive disclosures has different views. According to SEC, the repetition in disclosures of annual reports does not provide more information to investors and makes investors lose the concentration in other material information (SEC, 2003). In contrast, the succession hypothesis has the different view about the repetition of communication. This hypothesis

supposes the repetition is informative since investors do not read all sections of annual reports and spend around ten to thirty minutes for reviewing the annual reports (SEC, 2008; Arnold, Bedard, Phillips and Sutton, 2010), therefore, the repetitive disclosures in different sections of annual reports help investors not to miss the important information. Additionally, the repetitive disclosures in annual reports improve the information processing ability of investors since the repetition enhance the understanding of investors when information is displayed in different ways (Li, 2017).

### 2. 2 Negative impacts of the increase in the length of annual reports.

The dramatic increase in the length of annual reports has attracted the interest of researchers under controversial views. On the one hand, some research considers the length of annual reports as the readability of annual reports in the sense that the increase in the length of annual reports means the decrease in the readability of annual reports, thereby leading to negative impacts on stock markets and users. On the other hand, other research supposes the increase in the number of words in annual reports improves the readability of these reports. In fact, readability of annual reports is extremely important to external users and managers. For investors and shareholders, readability of annual reports prevents them not only from omitting important information of firms, but also from misunderstanding the current business performance of firms. Most importantly, readability of annual reports helps them save time and cost for acquiring information. Lower readability of annual reports causes investors to spend more money to buy analyst services and time for further search (Lehavy, Li, Merkley, 2011). In terms of managers, they must follow SEC's regulations about how to disclose information which requires annual reports of firms to "be readable and easy to understand to all shareholders" (SEC, 1998a). Additionally, the previous research shows that the way of disclosure that managers strategically decide can reduce the unfavorable views on negative economic consequences of investors to firms and prevent transparency.

Li (2008) reports that firms with poor performance provide the longer annual reports with the hope to hide adverse information to investors, hence the annual reports with more words prevent small investors from trading (Lawrence, 2013; Miller, 2010). Some research also records the more underreaction of the stock market when annual reports become longer (Callen, Khan, & Lu, 2010; Lee, 2012; You & Zhang, 2009). The increase in the length of annual reports also causes greater dispersion, lower accuracy, greater uncertainty in analysts' earnings forecasts (Lehavy et al., 2011).

However, some research records that the increase in the length of annual reports improves the readability of annual reports which is highly appreciated by investors, such as Cheung and Lau (2016) and Lundholm et al. (2014). They show that the longer annual reports due to different expressions upgrade the readability of annual reports. In other research, we also provide the evidence that foreign firms with better adhering plain English improve the readability of their annual reports in regardless of the significant increase in the length of 20-F (Nguyen and Kimura, 2017).

## 3. Sample selection

In contrast to existing research, we use the annual reports of foreign firms listed on the US Stock Exchange. As SEC's requirements, foreign firms listed on the US Stock Exchange, normally called "foreign private issuers", trade less than 50% of their shares on the US Stock Exchange, and annually fill the 20-F forms

**Table 1 Sample selection**

We use Python to collect all the 20-F filings from 2004 to 2013 from EDGAR's website. We remove all filings in the finance and insurance sectors and filings with fiscal year on the different date of December 31st from our sample. We only choose filings of firms having RIC codes (ie. .N or .O) available on Thomson Reuters and filings which have no missing data on Thomson Reuters.

| | |
|---|---|
| Total of observations collected from EDGAR | 7,588 |
| Number of firms | 1,475 |
| Number of firms whose fiscal years end on December 31 | 966 |
| Number of firms whose RIC codes available on Thomson Reuters | 264 |
| **Number of observations without missing data** | **1,527** |
| *In which:* | |
| Agriculture, forestry, and fishing | 2 |
| Mining | 118 |
| Construction | 5 |
| Manufacturing | 758 |
| Transportation, communication, electric, gas, and sanitary services | 385 |
| Wholesale trade | 23 |
| Retail trade | 13 |
| Services | 223 |

rather than 10-K like the US firms. We use Python to automatically collect all the 20-F filings in the period from 2004 to 2013 available on EDGAR (Electronic Data Gathering, Analysis, and Retrieval) and find 7,588 filings. We remove all filings in the finance and insurance industries and filings with fiscal year on the different date of December 31$^{st}$. We only choose filings of firms listed on NYSE or NASDAQ which they have RIC codes (i.e., .N or .O) available on Thomson Reuters Datastream. We eliminate all filings which have missing data. We also accept filings in text files which are compatible with using Perl language to measure the number of words. Finally, our sample has 1,527 observations (see Table 1).

To identify the length of annual reports, we need to clean the raw text files downloaded from EDGAR which contains both images, tables, figures. We follow the guidelines of Bonsall et al. (2017), Li (2008), and Loughran and McDonald (2014a, 2014b) for cleaning (Appendix 1). We use the Lingua::EN:: Fathom package and other packages of Perl language to calculate the number of words of filings after cleaning. Perl language is proved as the better way for analyzing the text of the large sample sets as compared to different solutions (Li, 2008).

## 4. Determinants of 20-F length

Cazier and Pfeiffer (2016) showed that there are three determinants of 10-K length which are operating complexity, redundancy, and residual disclosure. Applying findings of Cazier and Pfeiffer (2016), Li (2008), Lundholm et al. (2014) into our setting for 20-F filings, we use the specification in Equation (1) to identify the determinants of 20-F length.

$$NUMWORDS_{it} = \beta_0+\beta_1 SIZE_{it}+\beta_2 DEBT\_RATIO_{it}+\beta_3 LOSS_{it}+\beta_4 PTB_{it} +\beta_5 VOLATILITY_{it}+\beta_6 IFRS_{it}+\beta_7 ENGLISH_{it} \quad (1)$$

Overall, we identify there are seven determinants in five groups explaining the difference in the length of 20-F filings. The definitions and expected signs of variables are summarized in Appendix 2.

### 4. 1 Operating Complexity

Based on suggestion of Cazier and Pfeiffer (2016), we assume that firms with more transactions and more complicated operations have more disclosures at the end of fiscal year since they have more explanations about their business activities to investors under disclosure regulations of SEC and FASB, therefore, it is difficult for those firms to shorten their annual reports. To measure the operating complexity of firms, we use two different measurements:

We use the size as the proxy of operating complexity since firms with bigger size mean more business activities or more frequently complicated transactions or more operating complexity. In fact, there are several ways to measure the operating complexity of firms such as size, age, business or geographic segments but SIZE is the most popular proxy for operating complexity that is used in most of the prior research (Li, 2008; Cazier, Pfeiffer, 2016). More importantly, Size is easy to get reliable data while others are so difficult to acquire in Thomson Reuters Datastream. SIZE is identified by the natural logarithm of the market value of equity at the end of fiscal year.

Information on debts of firms is extremely sensitive to different users and necessary to disclose more details under regulations. We expect annual reports of firms with higher debt ratio are longer. DEBT_RATIO is equal to total liabilities scaled by total assets at the same year.

### 4. 2 Poor Performance

According to Li (2008), firms with poor performance often provide longer annual reports with fewer self-referential, more causation words, more future tense verbs to hide the adverse information as compared to reports of firms performing well. Therefore, we use LOSS as an indicator variable set equal to 1 when firms have positive earnings, and 0 otherwise. Additionally, price-to-book is also the other indicator of the performance of firms. PTB is equal to the market price per share divided into book value per share at the end of fiscal year. We expect firms with loss and lower price to the book have longer annual reports to explain more details about business activities or obfuscate the disadvantage information.

### 4. 3 Uncertainty

The uncertainty in stock returns is expected as another critical determinant of the length of annual reports. Under the efficient market theory, stock prices reflect the available information to markets, therefore firms with more information cause the volatility in stock prices. Meanwhile, annual reports are considered as the official resources of information to investors. It means that if stock prices of firms become more volatile, the annual reports are projected to become longer with more information or more explanations. VOLATILITY is equal to the daily standard deviation of adjusted stock returns during a year. We expect there is a positive association between the volatility of stock returns and the length of annual reports.

### 4. 4 Accounting Standard Systems

To improve the comparability of annual reports issued by firms globally, the harmonization and convergence of global accounting systems have occurred in recent years. Since 2007, SEC has recognized that US-GAAP and IFRS are converging sufficiently. Therefore, it allows foreign firms to eliminate a reconciliation to US-GAAP if they adhere to IFRS. The SEC has also started to consider allowing the US firms to file their annual reports by adopting IFRS (Topic Gateway Series, 2008). Kim et al. (2012) find that such SEC's elimination does not cause the negative impacts on stock markets and users. Firms adhering to IFRS do not increase more disclosures or the frequency of disclosures.

However, there have been some material differences existed between US-GAAP and IFRS which demand to have more declarations from firms (PwC, 2016). Therefore, we suppose annual reports of foreign firms under IFRS are assumed to be longer than annual reports following US-GAAP since these firms want to increase their comparability with the US-firms.

SEC allows foreign firms to choose IFRS or US-GAAP for their annual reports provided that firms must declare which accounting standard system is applied to their annual reports at the beginning of 20-F. Based on firms' announcement, we sort our filings into two different groups, namely IFRS and US-GAAP. IFRS is an indicator of which accounting standard is followed by firms and equal to 1 if firms adopt IFRS, and 0 otherwise.

### 4. 5 Language

Lundholm et al. (2014) find that foreign firms in non-English-speaking countries issue the longer MD&A section than foreign firms located in English-speaking countries. Following Lundholm et al. (2014), we expect the length of the annual reports issued by foreign firms located in English-speaking countries is longer than those provided by foreign firms in non-English speaking countries, implying that the advantage in the language of native speakers can shorten their annual reports.

Based on firms' declaration on 20-F, we identify the executive locations of firms. We follow Lundholm et al. (2014) for sorting filings into English native speaking countries and others. ENGLISH is equal to 1 if firms located in English speaking countries, and vice versa equal to 0. The list of countries and their language are mentioned in Appendix 3.

## 5. Findings and Discussion

### 5. 1 How does the length of annual reports change over time?

The previous research, as well as reports of professional organizations, have recorded the significant increase in the length of annual reports. Our research has the similar result with a dramatic increase in the length of 20-F at the average rate of 32% during 10 years from 2004 to 2013 which is equivalent to 19,508 words rising.

On average, each 20-F filing contains 61,908 words in 2004. Surprisingly, the number of words in 20-F has gradually risen over 10 years and reached to 81,416 words in 2013. Figure 1 describes the change in the length of annual reports over 10 years. Such increase in the length of annual reports does not mean to cause the decline in the readability of 20-F over time (not tabular), the change in the writing styles is considered as the reason for this tendency. Foreign firms become better adoption to the suggestions of SEC in using

**Figure 1 The length of reports from 2004 to 2013**

Figure 1 describes the length of 20-F over the 10 years. The length of annual reports is based on analyzing the text of the 20-F forms downloaded from EDGAR by using the Perl language.

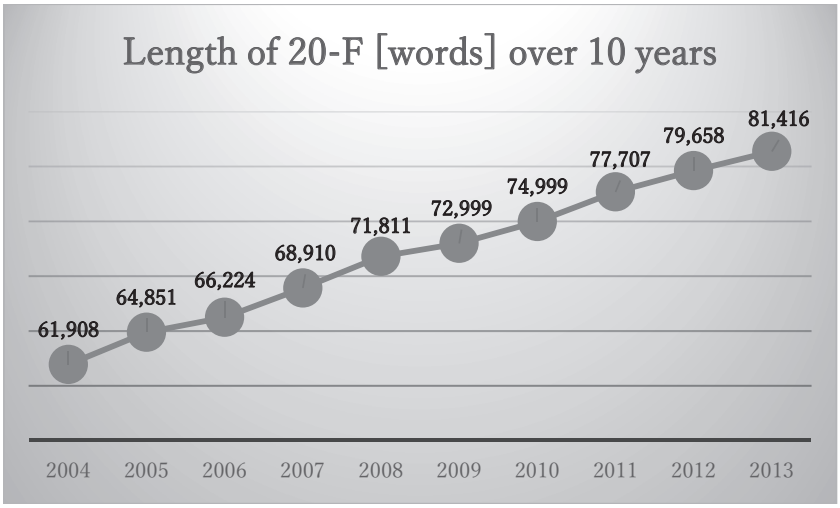


**Table 2 Descriptive Statistics**

| | N | Mean | Std. Dev | Q1 | Median | Q3 |
|---|---|---|---|---|---|---|
| *NUMWORDS* | 1,527 | 73,097 | 28,417 | 54,989 | 68,639 | 84,510 |
| *SIZE* | 1,527 | 7,07 | 2.52 | 5.20 | 7.00 | 9.13 |
| *DEBT_RATIO* | 1,527 | 0.23 | 0.21 | 0.04 | 0.19 | 0.35 |
| *LOSS* | 1,527 | 0.74 | 0.44 | 0 | 1 | 1 |
| *PTB* | 1,527 | 2.55 | 13.09 | 1 | 1.66 | 2.62 |
| *VOLATILITY* | 1,527 | 0.0314 | 0.02 | 0.0192 | 0.0269 | 0.0393 |
| *IFRS* | 1,527 | 0.5029 | 0.52 | 0 | 0 | 1 |
| *ENGLISH* | 1,527 | 0.4119 | 0.49 | 0 | 0 | 1 |

plain English for communication. In detail, they write more sentences but shorter sentences to improve the readability of 20-F over time.

### 5. 2 Descriptive statistics

Table 2 illustrates the descriptive statistics of all variables in Equation (1). The average number of words in 20-F form is 73,097 words, more than double those in 10-K which is reported at around 37,000 words in the previous research (Li, 2008; Bonsall et al., 2017). However, extremely long annual reports do not mean to be more informative since they contain more repetitive disclosures (Li, 2017; Cazier and Pfeiffer, 2016). The difference in the length of 20-F and 10-K is explained by the fact that foreign firms located in different countries, different legal and business environment so that they have to disclose more information to the US

**Table 3 Pearson correlation between variables**

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| **Numwords (1)** | 1 | | | | | | | |
| **Size (2)** | 0.39 | 1 | | | | | | |
| **Debt_Ratio (3)** | 0.20 | 0.03 | 1 | | | | | |
| **Loss (4)** | − 0.04 | 0.36 | − 0.06 | 1 | | | | |
| **PTB (5)** | − 0.07 | 0.01 | − 0.01 | − 0.04 | 1 | | | |
| **Volatility (6)** | − 0.10 | − 0.54 | 0.03 | 0.02 | − 0.32 | 1 | | |
| **IFRS (7)** | 0.29 | 0.5 | 0.04 | 0.01 | 0.20 | − 0.23 | 1 | |
| **English (8)** | − 0.26 | − 0.26 | − 0.15 | − 0.02 | − 0.14 | 0.08 | − 0.24 | 1 |

investors (Lundholm et al., 2014). Foreign firms with more explanation hope to increase the comparability and reliability of their annual reports to the US investors who underweight the equity of foreign firms (French & Poterba, 1991).

The average size of firms in our sample is more than 12,425 thousand dollars equivalent to the natural logarithm of 7.07. The mean of debt ratio is around 23% which means that around 23% of total assets are financed by debts. The mean of LOSS indicates that there are around 74% observations in our sample owning positive earnings. Price to book value and Volatility of stock returns are 2.55 and 0.0314 on average, respectively. The number of filings in our sample adopting IFRS is approximately equal to the number of filings adhering US-GAAP. There are around 42% of annual reports filed by English native speakers.

Table 3 shows the Pearson correlation between variables in Equation (1). There is a significant correlation between NUMWORDS and independent variables. The length of annual reports has the positive correlation with the size of firms, debt ratio, accounting standards, in contrast, the negative correlation between the length of 20-F with the loss, price-to-book value, volatility, and English. Although other variables are correlated with each other, the correlation magnitudes are not large. Especially, SIZE shows the negative correlation with the volatility. Our result is consistent with previous research (Fu, Kraft, and Zhang, 2012; Amiram, Owens and Rozenbaum, 2016). This negative correlation is explained by leverage effect and time-varying risk premia (Chueng and Ng, 1992; Duffee, 1995; Perez-Quirross and Timmermann, 2000).

### 5. 3 Empirical results

Table 4 summarizes the regression results of Equation (1). Column 1 shows the regression results of Equation without industry fixed effects and year fixed effects. Column 2 and Column 3 present the coefficients of Equation (1) when either industry fixed effects or year fixed effects are included in the regression. Column 4 illustrates the results when these both fixed effects are included in the regression. Overall, the coefficients and significance among models are not so different. Adjusted R square is improved over 4 models; however, the degree of freedom is declined when we include fixed effects into regressions.

As shown in Table 4, the length of annual reports is positively associated with the operating complexity. The coefficients of SIZE and DEBT_RATIO are 4,825 with t-statistic of 14.06 and 20,150 with t-value of 6.58

**Table 4 Regression results**

| Independent variables | Dependent variable | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| *Intercept* | 39,453***<br>(11.15) | 84,037***<br>(4.91) | 23,924***<br>(5.48) | 63,512***<br>(3.74) |
| *Size* | 4,825***<br>(14.06) | 4,270***<br>(12.04) | 5,189***<br>(14.59) | 4,613***<br>(12.46) |
| *Debt_Ratio* | 20,150***<br>(6.58) | 19,712***<br>(5.71) | 17,255***<br>(5.83) | 16,355***<br>(4.82) |
| *Loss* | − 12,633***<br>( − 8.11) | − 12,319***<br>( − 8.03) | − 11,340***<br>( − 7.40) | − 11,075***<br>( − 7.33) |
| *PTB* | − 180.47***<br>( − 3.78) | − 172***<br>( − 3.70) | − 149***<br>( − 3.18) | − 141**<br>( − 3.08) |
| *Volatility* | 169,496***<br>(3.95) | 150,284***<br>(3.51) | 265,998***<br>(5.35) | 242,761***<br>(4.87) |
| *IFRS* | 5,604***<br>(3.99) | 4,034**<br>(2.65) | 6,253***<br>(4.52) | 4,855**<br>(3.23) |
| *English* | − 8,143***<br>( − 6.06) | − 9,068***<br>( − 6.75) | − 7,183***<br>( − 5.50) | − 8,009***<br>( − 6.08) |
| *Industry fixed effects* | No | Yes | No | Yes |
| *Year fixed effects* | No | No | Yes | Yes |
| *Adjusted R2* | 0.26 | 0.30 | 0.30 | 0.33 |
| *Observations* | 1,527 | 1,527 | 1,527 | 1,527 |

respectively. The positive coefficients of size and debt ratio mean that firms with bigger size and/or higher debt ratio issue longer annual reports than other firms, ceteris paribus.

In contrast, the number of words in 20-F is negatively correlated with the performance of firms. This result is consistent with the results of Li (2008) and Aymen, Mhamed, and Badreddine (2016). They assume firms with negative earnings issue longer annual reports. The coefficients of LOSS and PTB are negative but significant at 99% confidence level. Firms getting a loss or lower price-to-book value provide longer reports since they want to obfuscate the adverse information to investors.

The regression result also supposes for the fact that daily stock returns of firms are more volatile, firms disclose more information in their annual reports which leads to longer annual reports. The coefficient of VOLATILITY is 169,469 with t-statistic at 3.95. Other words, the difference in the length of annual reports results from the variation in the volatility of stock returns.

Interestingly, filings following IFRS are extremely longer than filings under US-GAAP. The coefficient of IFRS is positive and significant at 99% confidence. Foreign firms listed on the US Stock Exchange adopting IFRS have to provide more explanation of the difference between IFRS and US-GAAP to investors which lengthen their annual reports. Another reason is that IFRS is principle-based and have a broad guideline, while US-GAAP is rule-based with a list of detailed rules. Principle-based accounting requiring more disclosures than rule-based accounting, which leads to a substantial increase in the length of annual reports issued by firms

adopting IFRS. In addition, more disclosures are expected to attract more US investors who are familiar with US-GAAP.

Table 4 also shows that annual reports of foreign firms located in English speaking countries are significantly shorter than those of firms in non-English speaking countries. ENGLISH has a negative coefficient. This result assumes that native speakers with advantage language are likely to shorten their annual reports under the same formats and same regulations, however, the length of sentences in those reports are longer than reports issued by firms in non-English speaking countries. Our result is consistent with Lundholm et al. (2014) who find that MD&A of 20-F in non-English speaking countries are longer than MD&A of 20-F in English speaking countries. However, the increase in the length of 20-F in non-English speaking countries does not worsen the readability of those reports (not tabular).

Our results do not significantly change when including industry fixed effects and/or year fixed effects, however, the significant level of IFRS decrease from 99% to 95% when we insert industry fixed effects into models. As concluded, the results over different models remain. Adjusted R squared of our models are from 26% to 33%. Other words, our independent variables can explain 26% to 33% the change of dependent variable.

Overall, the empirical results support for the expected signs mentioned above. In conclusion, the length of 20-F becomes longer when firms become bigger, higher debt-ratio, higher volatility. Conversely, firms with positive earnings and higher P/B shorten their annual reports. Additionally, adopted accounting standards and native language substantially affect the length of annual reports.

## 6. Conclusion

Annual reports are considered as the official communication channel of firms to investors, therefore, SEC, as well as FASB, seriously concern the format and contents of annual reports. They set a series of regulations and guidelines for firms to annually file forms. The more complicated business environment and various regulations make annual reports become more difficult-to-read for investors. The previous research recognizes the dramatic increase in the length of annual reports, moreover, the repetitive disclosures also rise rapidly. Similarly, our research records the length of 20-F forms issued by foreign firms listed on the US stock exchange rapidly increase over 10 years by 32% which is more serious than 10-K forms. Foreign firms increase the length of their reports due to shortening the sentences which is followed the SEC's suggestions.

In addition to operating complexity and firm performance, we realize that the volatility, accounting standards that firms apply to their annual reports and issuers' native language crucially determine the length of annual reports. This result is expected to assist the users as well as standard setters in understanding which factors determine the length of annual reports. From that, they can adjust their behaviors to adapt to the complicated tendency of current annual reports.

Despite our interesting findings, our research has still contained some open issues that need to deeply be investigated. Our research concentrates on how the number of words in 20-F change over time and the determinants of the length of annual reports, but we do not test whether longer annual reports are better/informative or not. Also, we have not yet concerned about the repetitive disclosures in 20-F forms which are shown as the serious problem in 10-K. We leave these issues to future work.

**References**


Amiram, D., Owens, E., & Rozenbaum, O. (2016). Do information releases increase or decrease information asymmetry? New evidence from analyst forecast announcements. *Journal of Accounting and Economics*, 62 (1), 121−138.

Arnold, V., Bedard, J., Phillips, J., Sutton, S., (2010). Where do Investors Prefer to Find Nonfinancial Information? *Journal of Accountancy*. https://www.journalofaccountancy.com/news/2010/aug/20102682.html Access 28 February 2018.

Aymen, A., Mhamed, L., & Badreddine, M. (2016). Guiding through the Fog: Does annual report readability reveal earnings management? *Research in International Business and Finance*, 38, 509−516.

Bloomfield, R. J. (2002). The "Incomplete revelation hypothesis" and financial reporting. *Accounting Horizons*, 16 (3), 233−243. https://doi.org/10.2308/acch.2002.16.3.233.

Bonsall, S. B. V, Leone, A. J., & Miller, B. P. (2017). A plain English measure of financial reporting readability. *Journal of Accounting and Economics*, 63 (2−3), 329−357.

Callen, J. L., Khan, M., & Lu, H. (2013). Accounting quality, stock price delay, and future stock returns. *Contemporary Accounting Research*, 30 (1), 269−295.

Cazier, R. A., & Pfeiffer, R. J. (2016). Why are 10-K filings so long? *Accounting Horizons*, 30 (1), 1−21. https://doi.org/10.2308/acch-51240.

Chueng. Yin-Wong, Ng. K. Lilian. (1992). Stock Price Dynamics and Firm Size: An Empirical Investigation. *The Journal of Finance*, 47 (5), 1985−1997.

Duffee.G.R. (1995). Stock Returns and Volatility: A firm-level analysis. *Journal of Financial Economics*, 37, 399−420.

Fu, R., Kraft, A., & Zhang, H. (2012). Financial reporting frequency, information asymmetry, and the cost of equity. *Journal of Accounting and Economics*, 54 (2−3), 132−149.

French, R. K., & Poterba, M. J. (1991). Investor Diversification and International Equity Markets. *American Economic Review*, 81 (2), 222−226.

Kim, Y., Li, H., & Li, S. (2012). Does eliminating the Form 20-F reconciliation from IFRS to U.S. GAAP have capital market consequences? *Journal of Accounting and Economics*, 53 (1−2), 249−270. https://doi.org/10.1016/j.jacceco.2011.05.001.

KPMG and Financial Executives Research Foundation. (2012). Disclosure overload and complexity: Hidden in plain sight. *Report*.

Lawrence, A. (2013). Individual investors and financial disclosure. *Journal of Accounting and Economics*, 56 (1), 130−147.

Lee, Y-J. (2012). The effect of quarterly report readability on information efficiency of stock prices. *Contemporary Accounting Research*, 29 (4), 1137−1170.

Lehavy, R., Li, F., & Merkley, K. (2011). The effect of annual report readability on analyst following and the properties of their earnings forecasts. *Accounting Review*, *86* (3), 1087−1115. https://doi.org/10.2308/accr.00000043.

Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics*, 45 (2−3), 221−247. https://doi.org/10.1016/j.jacceco.2008.02.003.

Li, H. (2017). Repetitive Disclosures in the MD&A. *Working Paper*. https://doi.org/10.2139/ssrn.2924193.

Loughran, T., & Mcdonald, B. (2014). Measuring readability in financial disclosures. *Journal of Finance*, 69 (4), 1643–1671. https://doi.org/10.1111/jofi.12162.

Lundholm, R. J., Rogo, R., & Zhang, J. L. (2014). Restoring the tower of babel: How foreign firms communicate with U.S. Investors. *Accounting Review*, 89 (4), 1453−1485. https://doi.org/10.2308/accr-50725.

Miller, B. P. (2010). The effects of Reporting Complexity on Small and Large Investor Trading. *The Accounting Review*, 85 (5), 1227−53.

Nguyen, T. T. P., Kimura. A., (2017). Are Longer Annual Reports Less Readable? Evidence from Foreign Firms Listed on the United States Stock Exchange. *Working Paper.*

Perez-Quiros, G., & Timmermann, A. (2000). Firm Size and Cyclical Variation in Stock Returns. *The Journal of Finance*, 55 (3), 34.

PwC. (2016). IFRS and GAAP: Differences and similarities, (November). Retrieved from www.pwc.com/usifrs%0Awww.pwc.com/usifrs.

SEC. Staff Legal Bulletin No. 7. Plain English disclosure (1998a). https://www.sec.gov/interps/legal/slbcf7.htm Accessed 1 October 2017.

SEC. Commission Guidance Regarding Management's Discussion and Analysis of Financial Condition and Results of Operation (2003). https://www.sec.gov/rules/interp/33-8350.htm Access 2 March 2018.

SEC. Mandatory Disclosure Documents Telephone Survey (2008). https://www.sec.gov/pdf/disclosuredocs.pdf Access 2 March 2018.

SEC. What we do (2013). https://www.sec.gov/Article/whatwedo.html Accessed 1 October 2017.

Topic Gateway Series. Convergence of Accounting Standards (2008). http://www.cimaglobal.com/Documents/ImportedDocuments/cid_tg_accounting_standards_feb08.pdf.pdf Accessed 10 October 2017.

You, H., & Zhang, X. jun. (2009). Financial reporting complexity and investor underreaction to 10-k information. *Review of Accounting Studies*, 14 (4), 559–586. https://doi.org/10.1007/s11142-008-9083-2.

**Appendix 1: Cleaning the raw text files of the 20-F forms**

To clean the raw text files downloaded from EDGAR, we follow the instructions of and Bonsall *et al.* (2017), Li (2008), and Loughran and McDonald (2014a, 2014b). We use the packages File::Slurp; HTML::Format text; and HTML::TreeBuilder; HTML::Entities; Text::Unidecode to clean the raw 20-F files, following these steps:

1. Remove format design in the raw files.
2. Remove all content between the <XBRL> and </XBRL> tags.
3. Remove all tables with more than 15% numeric characters.
4. Remove all markup tags.
5. Remove other textual expressions such as newline and underscored characters.

**Appendix 2: Variable Definitions and Expected Signs**

| Variable Name | Expected Sign | Definition |
|---|---|---|
| *NUMWORDS* | N/A | A number of words acquired by Lingua::EN::Fathom package of Perl language for 20-F after cleaning. |
| *SIZE* | (+) | Natural logarithm of the market value of equity at the end of fiscal year. |
| *DEBT_RATIO* | (+) | Total liability scaled by total assets. |
| *LOSS* | (–) | An indicator variable which is equal to 1 if firms have positive net income, 0 otherwise. |
| *PTB* | (–) | Market price to book value per share. |
| *VOLATILITY* | (+) | Daily standard deviation of stock returns during a year. |
| *IFRS* | (+) | An indicator variable which is equal to 1 if firms adopt IFRS, 0 if firms adhere to US-GAAP. |
| *ENGLISH* | (–) | An indicator variable which is equal to 1 if firms locate in English speaking countries, 0 if firms residence in non-English speaking countries. |

**Appendix 3: List of English-speaking and non-English-speaking countries**

Based on the declaration in the first part of the 20-F forms, the executive locations of firms are identified. The classification of Lundholm *et al.* (2014) is used to sort filings of foreign firms into the group of firms locating in English speaking countries and the group of firms in non-English speaking countries.

| ***Country*** | ***English*** | ***Country*** | ***English*** |
|---|---|---|---|
| *Anguilla* | 1 | *Liberia* | 1 |
| *Argentina* | 0 | *Luxembourg* | 0 |
| *Australia* | 1 | *Mexico* | 0 |
| *Austria* | 0 | *Monaco* | 0 |
| *Bahamas* | 1 | *Netherlands* | 0 |
| *Belgium* | 0 | *New Zealand* | 1 |
| *Brazil* | 0 | *Norway* | 0 |
| *Bulgari* | 0 | *Panama* | 0 |
| *Canada* | 1 | *Papua New Guinea* | 1 |
| *Chile* | 0 | *Peru* | 0 |
| *Columbia* | 0 | *Philippine* | 1 |
| *Denmark* | 0 | *Portugal* | 0 |
| *Finland* | 0 | *Russia* | 0 |
| *France* | 0 | *Singapore* | 1 |
| *Germany* | 0 | *South Africa* | 1 |
| *Greece* | 0 | *Spain* | 0 |
| *Hong Kong* | 1 | *Sweden* | 0 |
| *India* | 1 | *Switzerland* | 0 |
| *Indonesia* | 0 | *Taiwan* | 0 |
| *Ireland* | 1 | *Turkey* | 0 |
| *Israel* | 1 | *United Kingdom* | 1 |
| *Italy* | 0 | *Uruguay* | 0 |
| *Japan* | 0 | *Venezuela* | 0 |
| *Korea* | 0 | | |

［グェン　ティ　テュイ　フォン　横浜国立大学大学院国際社会科学府博士課程後期］
［きむら　あきひさ　横浜国立大学大学院国際社会科学研究院准教授］