# Information Disclosure and Firm Value: Empirical Evidence for MILA

Luisa F. Restrepo[†]

Advisor:
Diego F. Tellez, Ph.D.[‡]

School of Economics and Finance
Universidad EAFIT

Medellin, Colombia
Dic 1, 2019

[†]M.Sc. in Finance student at Universidad EAFIT, Colombia. E-mail: lfrestrepc@eafit.edu.co
[‡]Professor of Finance at Universidad EAFIT, Colombia. E-mail: dftellezf@eafit.edu.co

# Information Disclosure and Firm Value: Empirical Evidence for MILA

Luisa Restrepo*

*School of Economics and Finance,*
Universidad EAFIT, Medellin, Colombia

December 1, 2019

---

**Abstract**

Information disclosure is an important aspect of a company's governance policy, affecting investor's decision-making and firm value. The purpose of this study is to estimate the effect of information disclosed on firm value for firms in the Integrated Latin American Market (MILA) countries over the period 2011-2017. We use Structural Equations Modelling (SEM), where our latent variable "Disclosure Quality" is measured using five textual analysis variables as indicators. We find a positive and statistically significant effect of "Disclosure Quality" on Firm Value.

*Key words:* Information asymmetry; disclosure; firm value; textual analysis; structural equations modelling; SEM

*JEL Classification:* D82; G32; C38

*MSc in Finance student (Universidad EAFIT), Medellin. lfrestrepc@eafit.edu.co

# 1 Introduction

Information disclosure is an important aspect of a company's governance policy, this affects the investor's decision-making, asset prices and, in turn, the firm value. When companies commit to higher levels of disclosure, this reduces the possibility of information asymmetry between the company and its shareholders (Leuz and Verrecchia, 1999) also reducing the costs involved. Consequently, a firm's disclosure policy influences the amount of information contained in stock returns (Haggard et al., 2008), risk and firm value.

A substantial literature in accounting and finance has focused on information disclosure issues. There is an important strand of disclosure literature that explores the role of disclosure practices on firm variables (Banghoj and Plenborg, 2008; Botosan, 1997; Brown and Hillegeist, 2007; Dutta and Nezlobin, 2017; Hassan, 2018). The analytical foundation of these studies is the proposal that greater disclosure reduces information asymmetry. Diamond (1985) seminal study shows that the information disclosure policy improves the well-being of investors due to the reduction of information costs and the reduction of risk-sharing. Subsequently, Diamond and Verrecchia (1991) state that higher levels of disclosure lead to a reduction in information asymmetry between managers and investors. In the same vein, Easley and O'hara (2004) argue that investors demand higher returns when information is private, so the higher the disclosure, the greater the amount of public information and the lower the risk of investors owning a share.

Although several studies show substantial economic benefits from disclosure, there is little understanding of emerging markets (Chauhan and Kumar, 2018), including Latin American. Most studies related to information disclosure focus on the U.S. market, where the variation in disclosure levels is not large because mandatory disclosure is already high (Lopes and de Alencar, 2010). Following the theory of agency costs, (Jensen and Meckling, 1976) which establishes the consequences of agency costs on firm value, the characteristics of Latin American markets such as high information asymmetry and ownership concentration magnify the impact of agency costs. This makes it particularly interesting to evaluate the effect of corporate governance policies as disclosure in these markets.

Furthermore, given that in a scenario of high information asymmetry, the effect of corporate governance practices is greater, a vulnerable investor environment in Latin American countries creates an opportunity for firms to differentiate from others (Garay and González, 2008). Hence, the economic consequences of information disclosure on firm value are particularly important in a region characterized by high levels of information asymmetry as Latin America where it is more likely to observe a high disclosure level variation (Lopes and de Alencar, 2010). In this regard, empirical evidence on the impact of disclosure policies in Latin American markets has found positive effects (González et al., 2019).

In this sense, a Latin American market as MILA, created in 2010 seeking integration and diversification of the Latin American region and currently the most important center of investment in Latin America (MILA, 2017), is interesting to explore the effects of information disclosure over high levels of information asymmetry, since the purposes under which this market was created depend primarily on liquidity and transaction costs, which are in fact directly affected by information asymmetry and therefore by disclosure.

The purpose of this study is to estimate the effect of disclosure on firm value for firms in MILA

over the period 2011-2017. We estimate the effect of disclosure on firm value using Structural Equations Modelling (SEM), where our latent variable "Disclosure Quality" is measured using five textual analysis variables.

## 2 Information Disclosure and Firm Value

The main mechanism by which information disclosure affects firm value is by reducing the information asymmetry between the firm and its stakeholders. The role of information asymmetry in the investor decision making is a critical issue for market efficiency, as investors with different levels of information face different price curves that lead to information asymmetry being reflected in prices (Armstrong et al., 2011). Also, environments of high information asymmetry lead to estimation and valuation errors from investors when assessing investment opportunities.

These problems related to investor decisions in information asymmetry environments affect the price of the firm. Easley and O'hara (2004) find that private information leads to higher returns demanded by investors who have an informational disadvantage and higher estimation risk. This type of investor drives down the price of securities with greater information asymmetry. In summary, this reflects important economic consequences of information asymmetry, such as the increased risk of investors holding stocks and inaccurate investor estimates that affect their decisions and behavior about a firm.

Nonetheless, there are different ways to reduce these consequences of information asymmetry through disclosure. Diamond and Verrecchia (1991) states that greater levels of disclosure reduce information asymmetry between managers and investors due to increased liquidity of securities. On the other hand, Diamond (1985) shows that the information disclosure policy improves the well-being of the investor since information costs and risk sharing implied by information asymmetry are reduced. Empirically, Greenstein and Sami (1994) provide evidence about the negative association between the level of disclosure and information asymmetry. They find that bid-ask spread proxy for information asymmetry is reduced with higher segment disclosure.

The above supports the idea that corporate information disclosure reduces estimation risks and improves the valuation accuracy of investors. As Chauhan and Kumar (2018) state, the benefits of corporate disclosure come from the fact that investors can access information that allows them to assess the companys profitability in different time periods, which enhances valuation accuracy and then, the returns on investments. Consequently, the benefits of information disclosure are directly reflected in firm cash flows and cost of capital, which are fundamental determinants of firm value (Damodaran, 2012). First, a higher corporate disclosure can influence firm value directly through pure cashflow effects by reducing agency costs (Lang et al., 2003). Second, information disclosure reduces a firm cost of capital (Mangena et al., 2016; Schreder, 2018) through the reduction of estimation risks that implies a lower rate of return required by investors (Lambert et al., 2007), improving firm value.

Empirical evidence regarding the relationship between disclosure and firm value suggests a positive association. Examining whether corporate environmental disclosure affects firm value using a sample of S&P Global 1200 companies from 2010 to 2015, Hassan (2018) finds a

positive and significant relationship. Likewise, Chauhan and Kumar (2018) study the effect of non-financial disclosure on firm value for Indian firms and find a positive relationship. They attribute this result to lower cost of funds and higher operating cash flow due to disclosure. Also, for Latin American countries, Garay et al. (2013) and González et al. (2019), examine the effect of corporate disclosure index on firm value for Latin American stock markets and find that there is a positive association. Given these positive valuation effects associated with different measures of disclosure, as Patel et al. (2002) suggest, this evidence is consistent with the idea that there is a market premium on companies with lower asymmetric information. Therefore, better practices of information disclosure benefit a firm by increasing its value through the reduction of information asymmetry that reduces the cost of capital and enhances cash flows.

Table 2.1 Common Measures of Disclosure in Empirical Literature

| Authors | Disclosure Index |
|---|---|
| Chauhan & Kumar (2018) | Bloomberg Score |
| Hassan ( 2018) | Bloomberg Score |
| Patel et al. (2002) | Annual Report Attributes |
| Botosan & Plumlee (2002) | Annual Report Attributes |
| Garay et al. (2013) | Internet Attributes |
| Gonzles et al. (2019) | Anual Report/Internet Attributes |
| Lang et al. (2012) | Financial Variables |

The main challenge in this study is to measure the quality of information disclosed. Since the firm's level of disclosure is not directly observable, it is necessary to use proxy variables to measure it. In the empirical literature, such as the studies we listed above, the most common methodology is proxy disclosure using an index based on a checklist of items contained in annual reports. In Table 1.1 we have examples of different indexes used to measure disclosure. González et al. (2019) and Garay et al. (2013) built indexes as a checklist of information available on the firm's annual reports and the internet corporate web page for Latin American firms. Also, other indexes built on annual reports information are found in Patel et al. (2002) and Botosan and Plumlee (2002).

Differently from that approach, we measure information disclosure not from the quantity disclosed using indexes but from the quality of information captured by textual variables. Information disclosure measured merely as a quantity does not capture aspects beyond how much information is disclosed. It is important to characterize the information disclosed in order to capture the signaling effect of information and its true impact on stakeholders. We argue that some textual and language characteristics of the president's letter section in annual reports reflect the quality of information disclosed by the firm. Thus, we propose to measure the quality of information disclosed as a latent variable "Disclosure Quality" that is manifest in textual measures, given that the way a firm sends information to its shareholders is reflected in what is said and how it is said. To capture the quality aspects of information disclosed we use textual analysis techniques that allow us to capture different features related to the text, the language used and intention of information contained in annual reports.

Textual analysis is a relatively new area in accounting and finance (Loughran and Mcdonald,

2016) and it typically focuses on using computational resources to extract meaning from a set of texts. It is considered a subset of qualitative analysis. Some studies in accounting and finance using textual analysis methods to examine the impact of qualitative information on equity valuations are Antweiler and Frank (2004); Frazier et al. (1984); Tetlock (2007). To the best of our knowledge, studies related to disclosure and firm value (Chauhan and Kumar, 2018; Garay et al., 2013; Hassan, 2018; Patel et al., 2002) have not focused on examining narrative content to proxy disclosure, but on checking whether the reports contain certain financial and accounting information or information related to other corporate areas, these studies have not implemented textual measures to capture textual aspects that lead to a characterization of the content of interest. Nevertheless, González et al. (2019) characterize the information disclosed by Latin American firms using language textual measures, in this line, this is the closest research to our work, which argue that the tone is a language feature of information disclosed that affects firm value.

Considering the above, our input to measure the company "Disclosure Quality" in this research is the text from the presidents' letter section in the firm's annual report. Unlike most studies that examine the information content of financial statements, and more related to studies such as Abrahamson and Amir (1996) and González et al. (2019), we focus on the president's letter section of the annual report. Since the president's letter section is a narrative portion of the annual report that contains the presidents perspectives and opinions on the firm current and future performance, we intend to capture relevant non-financial and voluntary information. Based on this information, we use Cosine Similarity, Readability, Positive and Negatives tones and Length textual analysis variables to extract the quality of information disclosed and infer the latent variable "Disclosure Quality". Then, given the latent nature of our variable of interest, we estimate the effect of disclosure on firm value using a SEM Methodology that enables us to explore the relationships between latent and observed variables.

We contribute to extending the literature on the consequences of increased disclosure on emerging markets by providing empirical evidence useful for policy design about market transparency and disclosure levels that depends on firm incentives related to value. Furthermore, we contribute as we use a different methodology based on textual analysis techniques to measure disclosure level and structural equations model to estimate the relationship between disclosure and firm value. Our findings on the quality of information disclosed in the president's letter give important insights to explore the effects of narrative informatio on the firm value.

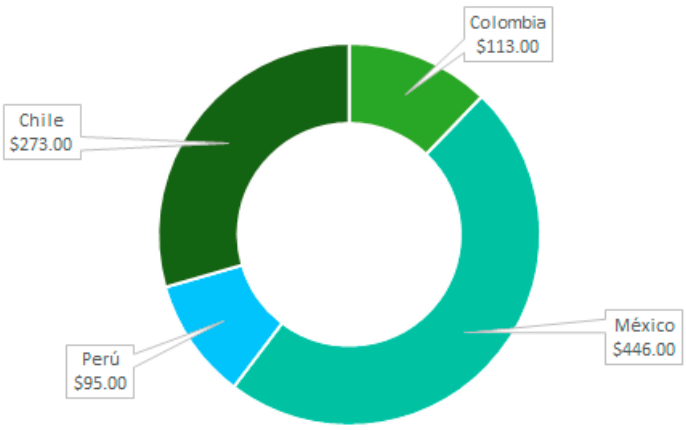
The remainder of the paper is as follows: In the Data and Methodology section, we describe the sample, the control, dependent and independent variables, the textual analysis we perform and the SEM model. In the Results section, we report the estimations obtained and discussion of them. Finally, we summarize and conclude.

# 3 Data and Methodology

MILA is a Latin American market created in 2010 seeking integration and diversification of the Latin American region and currently is the most important center of investment in Latin America (MILA, 2017). Figure 2.1 shows the market capitalization of markets in MILA as September 2017. The largest markets are Mexico and Chile followed by Colombia and Peru. MILA Market is an important investment center in Latin America that has great diversifica-

tion and also, an important potential for the creation of new business and commercial relations (MILA, 2017). Thus, the academic interest in this market has grown and has been mainly focused on the analysis of integration and market dynamics (e.g., Barari, 2004; Espinosa-Méndez et al., 2017; Mellado and Escobari, 2015).

Figure 3.1: Market Capitalization MILA Markets (Billion USD)




**Source**: MILA, 2017

The current study aims to examine the effect of information disclosure on firm value for firms in MILA. For this purpose, we start with the list of firms registered in MILA by country. Then, we hand-collect the annual reports for each firm during the period spanning 2010 to 2017. The source for the compilation of the annual reports was the website of the stock exchanges of each country and the websites of the companies. We start the period of study in 2011 as MILA began to operate in this year. We use the annual reports to, mainly, construct our text measures. We then complete our database with financial information from other sources such Bloomberg and Eikon. Due to lack of financial information for some companies, some annual reports were excluded from the total sample of 1631 annual reports. Then, our final sample is composed of 1412 observations representing 198 firms from which we were able to get annual reports and financial information required. For the analysis of textual variables, we take into account the entire sample of 1631 annual reports.

In Table 2.1 we have sample observations by country. Chile and Mexico are the countries with the greatest contribution as this are the greatest markets of MILA. With 48% of participation, observations from Chile composes almost half of complete sample, followed by Mexico with 22%, and Peru and Colombia which have the lowest contribution with 17% and 14% respectively.

Table 3.1 Sample by Country

| Country | Freq. | Percent | Cum. |
|---|---|---|---|
| Chile | 674 | 47.73 | 47.73 |
| Mexico | 311 | 22.03 | 69.76 |
| Peru | 233 | 16.5 | 86.26 |
| Colombia | 194 | 13.74 | 100 |
| Total | 1,412 | 100 | |

In Table 2.2 we have observations by sector. Sector data was collected from Economatica database. We get The North American Industry Classification System (NAICS) code as it is a much more consistent indicator of a company industry. Most companies of the sample are from Manufacturing Industry and Financial and Insurance Services with 27% and 25% respectively. Other relevant sectors in the sample are those related to utilities, mining and retail, which are less than and equal to 10% of sample. For further analysis of variables by sector, only these five representative sectors are taken into account.

Table 3.2 Sample by Sector

| Sector | Freq. | Percent | Cum. |
|---|---|---|---|
| Manufacturing Industries | 380 | 26.91 | 26.91 |
| Financial and Insurance Services | 351 | 24.86 | 51.77 |
| Electricity, Gas and Water Companies | 148 | 10.48 | 62.25 |
| Mining, Quarrying and Oil and Gas Extraction | 101 | 7.15 | 69.41 |
| Retail Trade | 96 | 6.8 | 76.2 |
| Agriculture, Livestock, Forestry, Fishing and Hunting | 69 | 4.89 | 81.09 |
| Construction | 64 | 4.53 | 85.62 |
| Transport, Post and Warehousing | 55 | 3.9 | 89.52 |
| Information in Mass Media | 38 | 2.69 | 92.21 |
| Corporate | 24 | 1.7 | 93.91 |
| Real Estate and Rental Services of Movable and Intangible Property | 24 | 1.7 | 95.61 |
| Health and Social Services | 23 | 1.63 | 97.24 |
| Cultural and Sporting Recreation and Other Recreational Services | 16 | 1.13 | 98.37 |
| Wholesale Trade | 10 | 0.71 | 99.08 |
| Temporary Accommodation and Food and Beverage Preparation Services | 7 | 0.5 | 99.58 |
| Business Support Services and Waste and Waste Management, and Remediation Services | 6 | 0.42 | 100 |
| Total | 1,412 | 100 | |

## 3.1 Dependent Variable

Given that our variable of interest is the firm value, we use Tobin's Q as it is the common variable used in the literature to measure firm value. Several studies related to firm value use Tobin's Q, some examples are studies that as well as we seek to measure the effects of corporate governance practices on firm value (e.g., Ammann et al., 2011; Jo and Harjoto, 2011). Tobin's Q is measured as the market value of assets divided by the book value of assets (La Porta et al., 2002). Given that low average liquidity of Latin American markets can make Tobin's Q less informative (Garay et al., 2013), we also use ROA as a robustness test to our model. ROA is

measured as the ratio of firm's Net Income to Total Assets.

## 3.2 Independent Variables

We propose to measure disclosure through the variable "Disclosure Quality" which is a non observable variable. In this work we argue that the quality of information disclosed is manifested in textual features of information disclosed, textual and language characteristics of president's letter section in annual reports reflects the quality of what a firm disclose. The way a firm sends information to its shareholders is reflected in what is said and how it is said. To measure the contribution of textual features on the quality of information disclosed, we use five textual measures, Similarity, Readability, Positive and Negative tones, and Length. We use this indicators because they allow us to characterize the type of text contained in the president's letter from text size, language attributes and differences in information disclosed from year to year.

First, we use Similarity also known as Cosine Similarity. This is a measure for narrative disclosure introduced by Brown and Tucker (2011). This measure allows us to capture the degree by which the president's letter section differs from the previous year using Vector Space Model (VSM) methodology. The Vector Space Model represents a document as a vector in an n-dimensional Euclidean space, where n is the number of unique words in all documents in the sample and the value of each vector element is the frequency of a particular word in that document Brown and Tucker (2011). With the vectors for each document, we calculate the similarity score which is measured as the angle between the two vectors representing the documents.

Suppose the sample has n unique words. Each document is represented as a n-dimension vector. The Similarity Score or Cosine Measure is defined as:

$$Sim = cos(\theta) = \frac{v_1}{\|v_1\|} . \frac{v_2}{\|v_2\|} = \frac{v_1 . v_2}{\|v_1\| \|v_2\|} \tag{3.1}$$

Where,

$v_1$ is the vector for document 1 and $v_2$ is the vector for document 2.

$\theta$ is the angle between $v_1$ and $v_2$

$\|v_1\|$ and $\|v_2\|$ are the vector length of $v_1$ and $v_2$ respectively

The score is bounded between 0 and 1, with a higher score indicating more similarity.

Second, we use the FLesch-Szigriszt Readability Index (IFSZ) for Spanish texts proposed by Szigriszt (1993). The score is bounded between 0 and 100, where a value between 0-50 indicates that the text is difficult; between 55-65 is normal and a score higher than 70 indicates that the text is easy.

$$IFSZ = 206.835 - 62.3 \frac{S}{P} - \frac{P}{F} \tag{3.2}$$

Where, $S$ it the total number of syllables, $P$ the total number of words and $F$ the total number of phrases.

Third, we use word content analysis to define positive and negative language in the President's letter section of the annual report. We use the positive and negative dictionaries for Spanish proposed by González et al. (2019). The procedure is to search inside the document for a list of words "bags of words" that are likely to be associated with positive and negative language.

$$score_i^{tf.idf} = \frac{1}{(1+log\ a_i)} \sum_{j=1}^{J} w_{ij}^{tf.idf} \tag{3.3}$$

Where, $a$ is the total number of words in document i, $w_{ij}^{tf.idf}$ is the weight of each term in the document and $J$ is the total number of words.

Finally, we include Length. This is a textual measure that is built from the word count of a text. Following Loughran and Mcdonald (2014) argument, that traditional measures of readability like the Fog Index are poorly specified in financial applications, we add Length as a measure of robustness for readability since the authors propose to measure readability with the file size which they show is a good proxy.

$$Length = Length(Text) \tag{3.4}$$

## 3.3 Control Variables

We follow the literature and propose the following control variables to our model setting. Firm Size is measured as the logarithm of total revenue. Firm size is expected to be positively associated with firm value indicating a size effect by which larger firms have higher Tobin's Q (Mak and Kusnadi, 2005). Leverage is expected to have a negative effect on firm value as high levels of leverage are expected to have underinvestment consequences (Myers, 1977). Growth Opportunities is measured as the growth rate of revenue, and it is negatively associated with firm value following the argument given by Myers and Majluf (1984) by which firms with high growth opportunities have greater levels of information asymmetry.

## 3.4 Descriptive Statistics

As mentioned above, the annual reports of the companies that make up our sample are the input for the calculation of the textual measures that we use in this study. The annual reports referred to here are those documents in which a company reports on its annual performance, not specific regulatory financial reports. The annual reports are generally composed of information related to all relevant areas of the company such as finance, risks, social responsibility, environment, etc., reporting the main results and expectations on the issues surrounding these areas. We are particularly interested in the section of the document where the president of the company gives a message to the shareholders. This section is usually named Letter to shareholders or the President's message.

Different names are used to refer to the annual reports containing the president's letter, depending on the country. In Chile and Peru, these reports are by the name "Memoria Anual" in Colombia "Informe de gestión" and in Mexico "Informe Anual". In most cases, these reports are not mandatory, as are the specific financial reports that are reported to each country's regulatory bodies. This is one limitation of our sample data, not all companies share annual reports

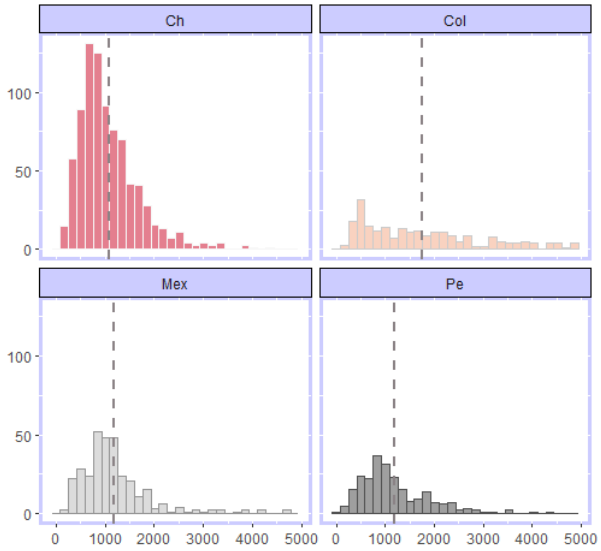
that are not exclusively mandatory.

For the collection of annual reports, we had to access different web pages as required. For companies registered on the Chilean stock exchange, the website of the financial market commission (www.cmfchile.cl) contains the annual reports of each company. By searching for the name of the company we accessed all the registered information of the company and also to the section "Memoria Anual" from where we downloaded the annual reports. For companies registered on the Peruvian stock exchange, on the page of the stock exchange of Lima (www.bvl.com.pe) when looking for a company name, in the section "Información Corporativa" we find the associated annual memories. In the case of Colombia and Mexico, for companies registered on the stock exchanges, it was necessary to access one by one to the web pages of the companies since these reports are not consolidated in a single source as in the case of Chile and Peru. Mexico's stock exchange web page contains only financial regulatory reports.

Table 3.3 Number of words by country

| Country | Mean | Min | Max | Number of reports |
|---|---|---|---|---|
| Chile | 1071 | 158 | 4342 | 836 |
| Mexico | 1167 | 179 | 4618 | 323 |
| Peru | 1180 | 64 | 4372 | 237 |
| Colombia | 1709 | 124 | 4795 | 235 |
| Total | | | | 1631 |

Figure 3.2: Number of words by country




Note: Histograms of the number of words in the president's letter grouped by country. For each histogram, dashed lines represents the mean of number of words in annual reports.

In Table 2.3 we have descriptive statistics of president's letters by country. Although Colombia and Peru are the countries that contribute the least to the sample of reports, they have the highest average words with reports containing an average of 1180 and 1709 words. Chile is the country

that contribute the most to the sample of annual reports with 836 reports. Average number of words in Chile reports is 1071 words. The smallest annual report in our sample has 64 words and is from Peru and the largest annual report with 4795 words if from Colombia. In Figure 2.2 we have histograms of number of words in annual reports by country. In general, the annual reports do not have a standard size per country, the number of words varies in a considerable range of words. Colombia's reports have the lower variation compared to the other countries.

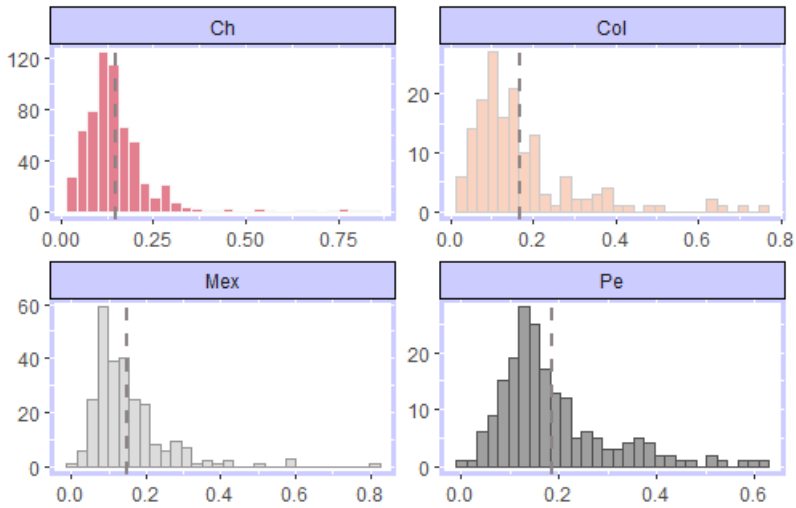
The descriptive statistics for the financial and textual variables are give in Table 2.4. The input to calculate textual measures is the text of the president's letter in annual reports. Textual measures are obtained building text vectors where each text from each year and each firm is taken as a particular document.

Table 3.4 Descriptive Statistics

| Variable | N | Mean | Min | Max | S.D. |
|---|---|---|---|---|---|
| Firm Value | 1352 | 1.35 | 0.25 | 9.01 | 0.83 |
| Similarity | 1189 | 0.16 | 0.01 | 0.84 | 0.11 |
| Positive | 1412 | 0.04 | 0.00 | 0.12 | 0.01 |
| Negative | 1412 | 0.02 | 0.00 | 0.07 | 0.01 |
| Readability | 1411 | 53.13 | 0.00 | 90.34 | 9.32 |
| Length | 1412 | 1241 | 64 | 4926 | 821 |
| ROA | 1264 | 0.07 | -0.72 | 1.19 | 0.09 |
| Size | 1379 | 9.68 | -3.22 | 18.07 | 3.60 |
| Growth | 1211 | 0.21 | -1.00 | 91.96 | 2.89 |
| Leverage | 1384 | 1.89 | 0.00 | 50.29 | 3.01 |

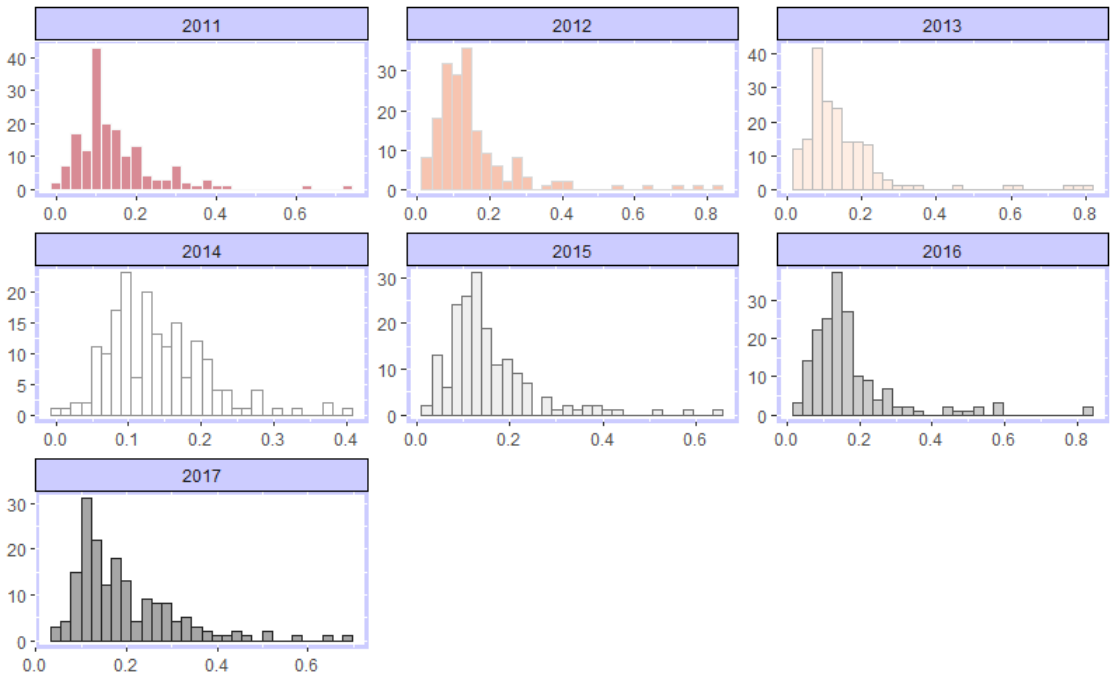
Regarding the financial variables, for firm value (Tobin's Q) we have a mean of 1.35 and a mean of 0.07 for ROA. On average, the firms from our sample have a leverage of 1.89 and a sales growth ratio (Growth) of 0.21. About the textual measures we have that the average similarity is 0.16, since Similarity is bounded between 0 and 1 , 0.16 means a relatively low repetition in annual reports from one year to another. For Positive we have low levels that are not beyond 12% with a mean of 4%. Also, for Negative we have low levels as values are not beyond 7%, with a lower mean of 2%. Readability has in average a value of 53.13, since this measure is bounded between 0 and 100, 53.13 means normal readability in the scale of this score, that is a text that is neither too easy nor too difficult to understand. Finally, for Length we have that in average the president's letter section have 1241 words, the smallest letter have 64 words and the largest 4926.

Figure 3.3: Similarity by Country




Note: Similarity is measured in pairs of years, indicating how similar an annual report is to the annual report of the immediately preceding year. i.e., 2011 similarity is the similarity between 2010 and 2011 annual report. Similarity vales are between 0 and 1.

Figure 3.4: Similarity by year

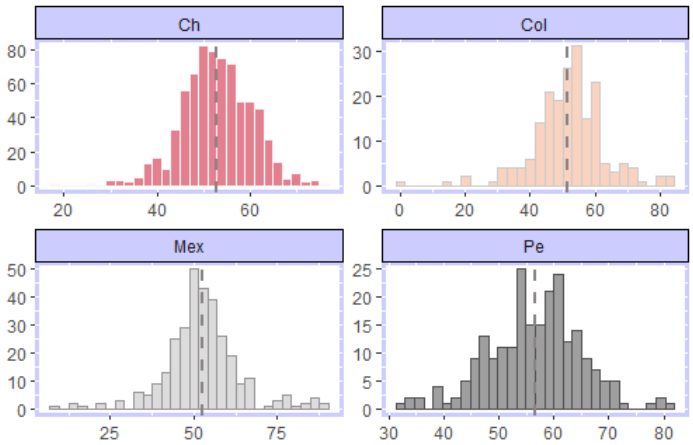



Note: Similarity is measured in pairs of years, indicating how similar an annual report is to the annual report of the immediately preceding year. i.e., 2011 similarity is the similarity between 2010 and 2011 annual report. Similarity vales are between 0 and 1.

In Figure 2.3 and 2.4 we have histograms of Similarity. By country and year respectively. Colombia and Mexico annual reports present highest values of similarity reaching 0.8 but this are not representative values. For the four countries we have a general concentration of similarity around values near 0.1 and 0.2. which represents a low level of similarity between two annual reports. Peru shows greater variation of similarity with values of similarity around 0.05

and 0.4. About the similarity by year, the highest values of similarity reaching 0.8 are present in years 2012, 2013 and 2016. For each year, representative values of similarity are around 0.1 and 0.2. This means that in general, a specific year annual report is as similar to its immediately last annual report in 10% and 20% of the content.

Figure 3.5: Readability by Country

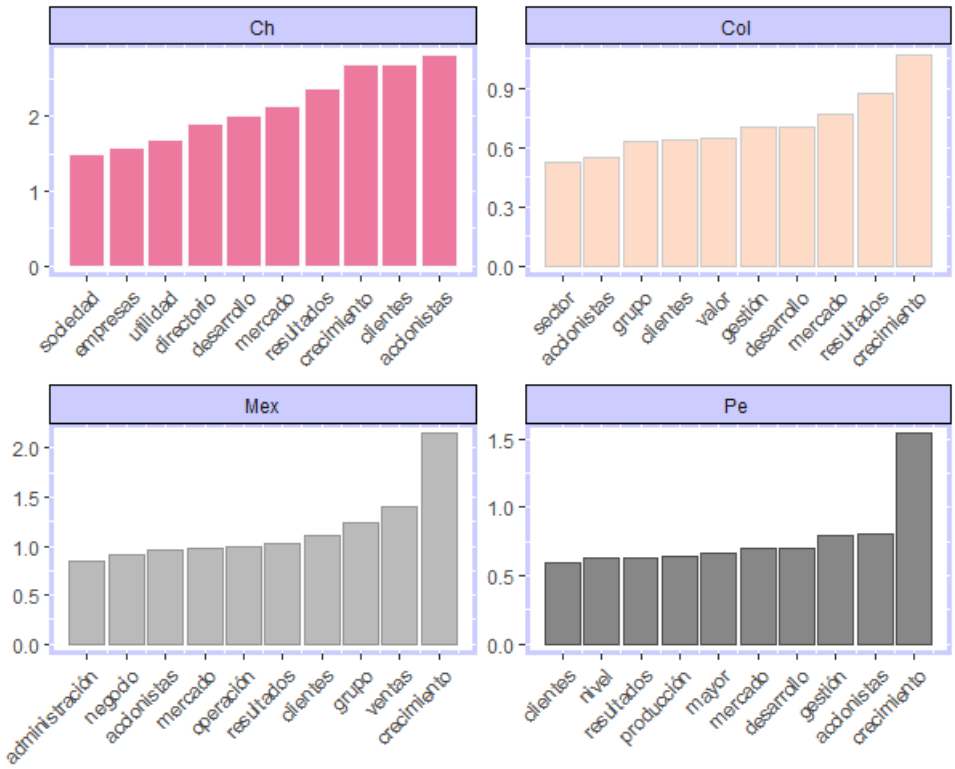



Note: Readability values ranging from 0 to 100.

In Figure 2.5 we have histograms of Readability by country. The average readability by country is 52.73, 51.65, 52.43, 56.52, for Chile, Colombia Mexico and Peru respectively. This values are placed in the normal readability of readability scale, which means that the presidents letters from the four countries are neither too easy nor too difficult to understand.

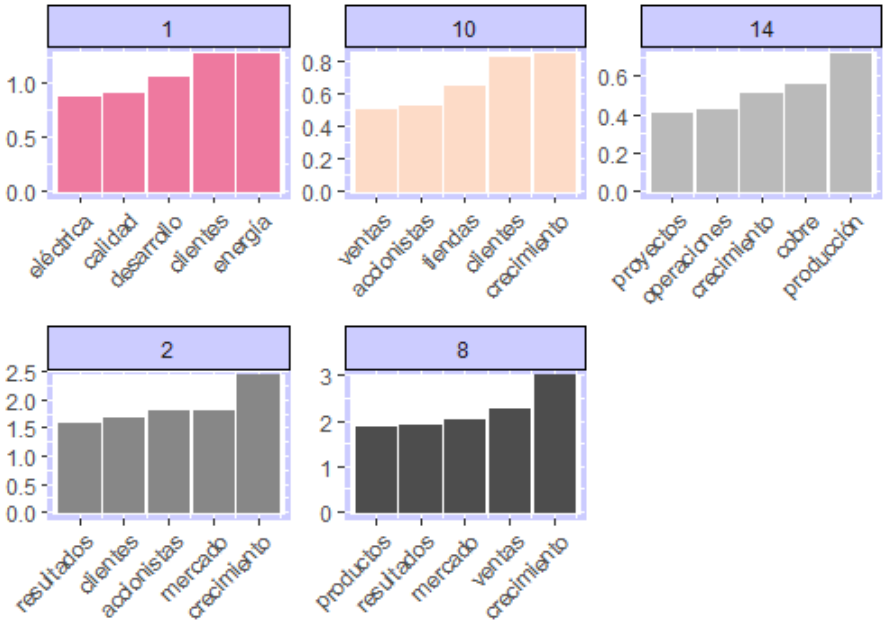
In Figure 2.6 and 2.7 we have words frequency by country and sector respectively. Regarding the words frequency by country, we have the ten most frequents words in president's letter for each country. There are several words that are common in the four countries, the words "accionistas", "crecimiento" and "resultados" are common in Chile, Colombia, Mexico and Peru president's letters. About the words frequency by sector, we have the five most representative sectors in our sample, this is, Electricity, Gas and Water Companies (1), Financial and Insurance Services (2), Manufacturing Industries (8),Mining, Quarrying and Oil and Gas Extraction (14), and Retail Trade (10). The word "crecimiento" is common to president's letter in sectors 2, 8, 10 and 14. Also, the word "clientes" is common to president's letter in sectors 1, 2 and 10.

Figure 3.6: Words Frequency by Country




Note: Ten most common words in president's letter from Chile, Colombia, Mexico and Peru.

Figure 3.7: Words Frequency by Sector




Note: Five most common words in president's letter by sector. Electricity, Gas and Water Companies(1), Financial and Insurance Services(2), Manufacturing Industries(8), Mining, Quarrying and Oil and Gas Extraction(14), and Retail Trade(10).

## 3.5 Structural Equations Model

As stated earlier, given the unobservable nature of our variable of interest, "Disclosure Quality", we use Structural Equations Modelling (SEM) to estimate the relationship between "Disclosure Quality" and firm value. Structural equations models are a family of multivariate statistical models that allow us to estimate the effect and relationships between multiple variables, including latent constructs or not directly observable variables. Structural equations modelling accounts for the relationships between observed and latent variables in different types of theoretical models, which provide a quantitative test of a hypothesis established (Lomax and Schumacker, 2004). SEM models combines regression, factorial analysis, and analysis of variance for simultaneously estimating interrelated dependency relationships.

Three features that characterize Structural Equation Models (Hair et al., 2006) are:

- Estimation of multiple and interrelated dependence relationships
- The ability to represent unobserved concepts in these relationships and correct the measurement errors in the estimation process.
- Defining model to explain the entire set of relationships

Thus, this methodology allows us to formulate hypotheses about how sets of variables defining constructs and how they relate to each other and explicitly take measurement error into account (Lomax and Schumacker, 2004). In this sense, we are able to analyze latent and observed variables with their associated measurement error terms, so we can test whether our disclosure hypothesised model is supported by sample data. Relying on this properties, to address both, the simultaneity of regression equations and the estimation issues of the unobserved latent factor "Disclosure Quality" we follow the Anderson and Gerbing (1988) general model of two-step modelling. This SEM approach consists of a measurement and a structural model.

## 3.6 Measurement Model

The measurement model is a confirmatory factor model which results from Confirmatory Factor Analysis (CFA). This analysis requires an specific theoretical background for our latent variable "Disclosure Quality", but in the absence of this, we use the confirmatory factor model in an exploratory manner (Long, 1983). In this model we specify how the latent factors or constructs are related to the observed variables. Factor analysis attempts to explain the variation and covariation in a set of observed variables in terms of a set of unobserved factors (Long, 1983). The factors are of two types, common factors that may directly affect more than one of the observed variables and unique or residual factors that may directly affect only one observed variable (Long, 1983), the error term.

In our measurement model we establish one latent factor "Disclosure Quality" common to five observed variables or indicators , Similarity, Readability, Positive and Negative tones and Length which are defined as we mentioned before. Mathematically, for each indicator we have the following equation indicating its relationship with the factor:

$$y_i = \lambda_{ik}\ \eta_k + \varepsilon_i \tag{3.5}$$

Where,

$y_i$ is the observed indicator, $\lambda_{ik}$ is the factor loading for the i measured variable predicted by the factor k, $\eta_k$ is the latent variable k and $\varepsilon_i$ is the residual.

The error term contains information about systematic variance unique to each indicator and random measurement error (Newsom, 2015). The latent variable variance is derived from the variances and covariances of the observed variables. The covariance matrix is descomposed into the matrices of loadings, factor variances and measurement residuals (Newsom, 2015).

$$Var(y_i) = \lambda_{ik}^2 \, Var(\eta_k) + Var(\varepsilon_i) \tag{3.6}$$

$$Cov(y_i, y_j) = \lambda_{ik} \, \lambda_{jk} Var(\eta_k) \tag{3.7}$$

$$Var(\eta_k) = \frac{Var(y_i) - Var(\varepsilon_i)}{\lambda_{ik}^2} \tag{3.8}$$

For each equation, $Var(\varepsilon_i) = \theta_{ii}$ and $Var(\eta_k) = \Psi_{kk}$

## 3.7 Structural Model

The structural model is the general model where we get together the measurement and the regression model. The structural model we propose to estimate in Figure 2.8. contains one measurement model for "Disclosure Quality" with indicators Similarity, Readability, Positive and Negative tones, and Length. Also, we have one regression model to infer the relationship between Firm Value and the latent construct "Disclosure Quality", including the control variables, Size, Leverage and Growth. In SEM convention is usual to use circles to represent "Latent" variables and squares to represent "Observed".

For the estimation of the model factor loadings, path coefficients and residual variances, we need maximum likelihood estimates. In the estimation of the SEM model the purpose is to minimize the difference between the observed and estimated population covariance matrices.

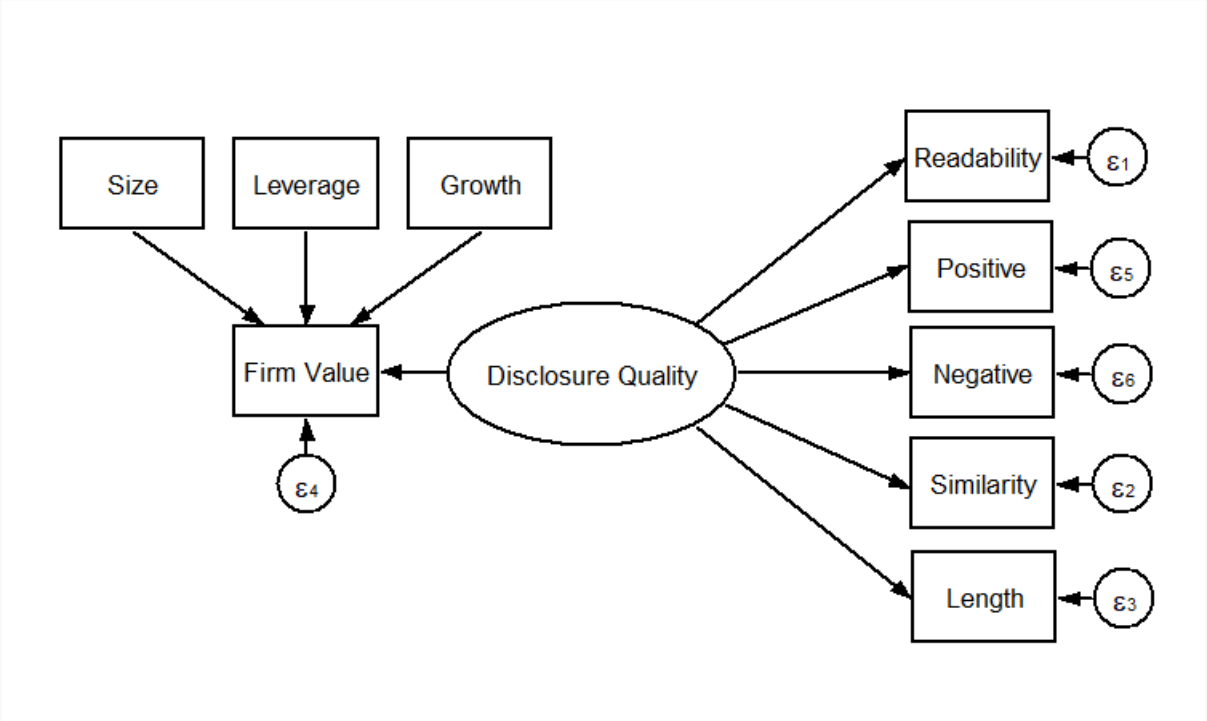
$$F = \sum_{i=1}^{N} log|\Sigma_i| + \sum_{i=1}^{N} log(\mathbf{x}_i - \mu_i)' \, \Sigma_i^{-1} \, (\mathbf{x}_i - \mu_i) \tag{3.9}$$

Where,

The subscript i refers to the observed cases, $\mathbf{x}_i$ to the variables observed for case i and $\mu_i$ and $\Sigma_i$ contain population means and covariances of the variables observed for case i.

The fit function in Equation 2.8 applies with clusters as units of observation, and individual within clusters as variables (Hox et al., 2017). Also, allowing for incomplete data we estimate using Full Maximum Likelihood asumming missing at random (MLR), which allows for robust chi-squares and standard errors to non normality. SEM estimation provides standardized

Figure 3.8: Hypothesised Model of Disclosure Quality and Firm Value




estimates to consider variables that are measured in a different metric.

Regarding the goodness of fit, to asses how well the hypothetical model approximates the sample data, we carefully have to consider jointly different goodness of fit measures available in SEM, as Chi-square test, Akaike's Information Criterion, the Comparative fit index (CFI), Tucker Lewis index (TLI) and RMSEA.

# 4 Results and Discussion

Variables correlations are given in Table 3.1. First, we describe the correlations among the textual variables Similarity, Positive and Negative tones, Readability and Length with Firm Value. For Similarity, Negative tone and Length we have a negative relationship with Firm Value. Since high values of these indicators are a sign of poor quality disclosure, it is intuitively reasonable for them to be negatively related to Firm Value, as explained next. About Similarity, this is a measure of the similarity between two president letters of different but contiguous years. For example, the similarity is basically the comparison between the textual contents of 2010 vs 2011 or 2011 vs 2012 president's letters. A high value of this indicator means that we have a low level of new information disclosed from one year to another and so, the two letters are very similar which is a negative signal about the quality of information. On the Negative tone, it's not very clear, but as the percentage of Negative tone increase in a president letter we have a signal of pessimism that can be misinterpreted by stakeholders, in this sense this might be a negative feature of information but is not conclusive. About Length, this is a measure of text size. As Length increase this could be a sign that the information in the president letter is more filling information than relevant information about the firm. Also, very long texts can be tedious for the reader and thus can affect the reader's perception of relevant information.

Following the same intuition, for Readability and Positive tone, we have a positive relationship with Firm Value. Since these measures could be indicators of the good quality of the text, this relates positively to firm value. About Readability, this measure allows us to examine the complexity of a text. Increases in Readability could be interpreted as information easier to understand given the scale of measurement. But from a different point of view, as texts are easier to read, the level of education required is lower which may be a sign that the text is not aimed at a target audience like high education stakeholders and so, it is possible that information contained in the president's letter is not very informative. On the Positive tone, as the percentage of Positive tone increase this could be interpreted as a signal of optimism that can be received by stakeholders as a good signal although this might be a positive feature of information this is not conclusive.

For Leverage and Growth, we have a negative association with firm value. Leverage is expected to have a negative effect on firm value due to underinvestment problems (Fosu et al., 2016; Myers, 1977) and Growth is expected to have a negative effect as firms with high growth opportunities have greater levels of information asymmetry that affects negatively the firm value Myers and Majluf (1984). Size is positively related to firm value, this is a size effect by which larger firms are high valued (Mak and Kusnadi, 2005).

Table 4.1 Variables Correlation

| Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1.Firm Value | - | | | | | | | | | |
| 2.Similarity | -0.037 | - | | | | | | | | |
| 3.Positive | 0.024 | -0.177* | - | | | | | | | |
| 4.Negative | -0.193* | 0.028 | -0.145* | - | | | | | | |
| 5.Readability | 0.019 | 0.216* | -0.262* | 0.251* | - | | | | | |
| 6.Lenght | -0.078* | 0.299* | -0.193* | 0.004 | 0.091* | - | | | | |
| 7.ROA | 0.701* | -0.009 | 0.076* | -0.137* | 0.050 | -0.061* | - | | | |
| 8.Size | 0.096* | -0.038 | 0.107* | -0.132* | -0.002 | 0.154* | 0.127* | - | | |
| 9.Growth | -0.013 | 0.011 | -0.016 | 0.050 | 0.072* | -0.006 | -0.023 | -0.065* | - | |
| 10.Leverage | -0.081* | 0.103* | 0.048 | 0.097* | 0.055* | 0.051 | -0.109* | 0.241* | -0.020 | - |

$*p < .05$

Without a specific theoretical basis for our latent variable "Disclosure Quality", as we stated before, we propose five indicators to approximate the quality of information disclosed and thus, our latent variable. In the first step, the measurement model, we estimated the confirmatory factor analysis model to determine how the Similarity, Positive, Negative, Readability and Length indicators are empirically related to "Disclosure Quality". Then, in the second stage, we estimate the relationship between Dislocusre Quality and Firm Value in the SEM model. We estimated CFA and SEM models using R program and (FIML) estimation which is a robust estimate in the presence of missing values.

Regarding the first step estimations, the factor analysis (CFA), In Table 3.2 we have the results for the measurement model. All of the textual variables are significant in explaining the variance of the Disclosure Quality factor. Similarity, Negative, Readability and Length variables are negatively related to Disclosure Quality. The negative effect of Similarity suggests that

higher values of similarity are signal of poor quality of the president's letter as the similarity between two different years letters means that a firm is not disclosing new information. The negative effect of Readability may be read from the fact that higher levels of Readability show that the level of education required to understand the information disclosed is not high. Thus, the meaning of words used could not be targeted to specific public or shareholders which makes this information less useful for them. The negative effect of Length may be capturing the fact that the longer the texts, the more irrelevant information is likely to be included. Thus, a higher extent of the president's letter may be undervalued by shareholders. The effects of Positive and Negative tones are not conclusive as this are textual features that cannot explicitly mean good or bad quality, but as we stated before, them can serve as a sign of optimism and pessimism for stakeholders.

Table 4.2 Indicators Effect on Disclosure Factor

| Indicators | Effect |
|---|---|
| Similarity | -0.432*** |
| Positive | 0.493*** |
| Negative | -0.303*** |
| Readability | -0.542*** |
| Length | -0.321*** |

$*p < .1 **p < .05 ***p < .01$

Table 4.3 Direct, Indirect, and Total Effects of Variables on Firm Value

| Variable | Direct Effect | Indirect Effect | Total Effect |
|---|---|---|---|
| Disclosure | 0.080** | | 0.080** |
| Similarity | | -0,035 | -0,035 |
| Positive | | 0.039 | 0.039 |
| Negative | | -0.024 | -0.024 |
| Readability | | -0,043 | -0,043 |
| Length | | -0,026 | -0,026 |
| Size | 0.042*** | | 0.042*** |
| Growth | -0.001 | | -0.001 |
| Leverage | -0.043*** | | -0.043*** |

$*p < .1 \;\; **p < .05 ***p < .01$

In Table 3.3 we have SEM estimations. The main interest of this research is the path from Disclosure Quality to Firm Value. This path gives a positive relationship as we expected. A coefficient of 0.080 suggests that for each standard deviation increase in disclosure quality, the firm value increase by 0.080. This effect evidences that different features of information disclosed to shareholders are important for the firm valuation. Regarding control variables, Size is positively related to firm value and significant as we expected. Growth and Leverage are negatively related to firm value, but only Leverage is significant. In SEM methodology is possible

to obtain indirect effects of indicators, Similarity, Positive, Negative, Readability and Length, on Firm Value. Similarity, Negative, Readability and Length have a negative effect on Firm Value and Positive is positively related to Firm Value.

Model fit indices are set in Table 3.4. For RMSEA and SRMR smaller values are better, and for Chi-squared, CFI and TLI higher values are better.

Table 4.4 Goodness of Fit

| Indicator | Value |
|---|---|
| Chi-squared | 283.609 |
| RMSEA | 0.095 |
| SRMR | 0.060 |
| CFI | 0.612 |
| TLI | 0.466 |

# 5 Robustness

As a test of robustness, we proposed to measure the firm value as the Roa of the company. In the first step, the measurement model, we found that textual indicators have no significant change, remain close to the initial estimates and maintain their signs and significance. In the structural model, the control variables remain almost the same, but the effect of sales growth is positive and significantly different from the initial estimates using Tobin's Q as a measure of the firm value. In this exercise, the goodness of fit measures improve.

Table 5.1 Indicators Effect on Disclosure Factor

| Indicators | Effect |
|---|---|
| Similarity | -0.430*** |
| Positive | 0.497*** |
| Negative | -0.297*** |
| Readability | -0.543*** |
| Length | -0.322*** |

$*p < .1. **p < .05. ***p < .01$

# 6 Conclusions

Information disclosure is an important aspect of a company's governance policy that affects the investor's decision-making and also the firm value. In this research, we estimate the effect of disclosure on firm value for firms in MILA countries over the period 2011-2017. We use Structural Equations Modelling (SEM), where our latent variable "Disclosure Quality" is measured

Table 5.2 Direct, Indirect, and Total Effects of Variables on Firm Value

| Variable | Direct Effect | Indirect Effect | Total Effect |
|---|---|---|---|
| Disclosure | 0.078** | | 0.078** |
| Similarity | | -0,034 | -0,034 |
| Positive | | 0.039 | 0.039 |
| Negative | | -0.023 | -0.023 |
| Readability | | -0,042 | -0,042 |
| Length | | -0,025 | -0,025 |
| Size | 0.037*** | | 0.037*** |
| Growth | 0.024** | | 0.024** |
| Leverage | -0.086*** | | -0.086*** |

$*p < .1. **p < .05. ***p < .01$

Table 5.3 Goodness of Fit

| Indicator | Value |
|---|---|
| Chi-squared | 254.915 |
| RMSEA | 0.089 |
| SRMR | 0.057 |
| CFI | 0.662 |
| TLI | 0.536 |

using five textual analysis variables as indicators. We find a positive and statistically significant effect of the latent factor Disclosure Quality on Firm Value.

The indirect effects of textual variables on Firm Value are also captured. Similarity, Negative, Readability and Length are negatively related to Firm Value and Positive is positively related. In the exploratory analysis, we have significant effects of textual measures on Disclosure Quality. This evidence shows the importance of features related to information disclosed in the quality of the information perceived by the shareholders and gives some light for policy issues about information requirements that can enhance the firm's perception.

The major limitation of this research is the exploratory fact of the indicators of Disclosure Quality. Since there is no theoretical basis to support the choice of variables that signal the quality of information disclosed we use factor analysis in an exploratory manner which leads us to explore the relationship between the textual variables and Disclosure Quality but this can represent specification issues. However, the relationships found in this research may be the path to future research and provide a different perspective on the effects of the information disclosed on the firm value.

# References


Abrahamson, E. and Amir, E. (1996). The information content of the president's letter to shareholders. *Journal of Business Finance and Accounting*, 23(8):306–686.

Ammann, M., Oesch, D., and Schmid, M. M. (2011). Corporate governance and firm value: International evidence. *Journal of Empirical Finance*, 18(1):36–55.

Anderson, J. C. and Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological bulletin*, 103(3):411–423.

Antweiler, W. and Frank, M. Z. (2004). Is all that talk just noise? the information content of internet stock message boards. *The Journal of finance*, 59(3):1259–1294.

Armstrong, C. S., Core, J. E., Taylor, D. J., and Verrecchia, R. E. (2011). When does information asymmetry affect the cost of capital? *Journal of Accounting Research*, 49(1):1–40.

Banghoj, J. and Plenborg, T. (2008). Value relevance of voluntary disclosure in the annual report. *Accounting Finance*, 48:159–180.

Barari, M. (2004). Equity market integration in latin america: A time-varying integration score analysis. *International Review of Financial Analysis*, 13(5):649–668.

Botosan, C. A. (1997). Disclosure level and the cost of equity capital: South African evidence. *The Accounting Review*, 72(3):323–349.

Botosan, C. A. and Plumlee, M. A. (2002). A re-examination of disclosure level and the expected cost of equity capital. *Journal of Accounting Research*, 40(1):21–40.

Brown, S. and Hillegeist, S. A. (2007). How disclosure quality affects the level of information asymmetry. *Review of Accounting Studies*, 12(2-3):443–477.

Brown, S. and Tucker, J. W. (2011). Large-Sample Evidence on Firms' Year-over-Year MD&A Modifications. *Journal of Accounting Research*, 49(2):309–346.

Chauhan, Y. and Kumar, S. B. (2018). Do investors value the nonfinancial disclosure in emerging markets? *Emerging Markets Review*, 37:32–46.

Damodaran, A. (2012). *Investment valuation: Tools and techniques for determining the value of any asset*, volume 666. John Wiley & Sons.

Diamond, D. W. (1985). Optimal release of information by firms. *The Journal of Finance*, 40(4):1071–1094.

Diamond, D. W. and Verrecchia, R. E. (1991). Disclosure , Liquidity , and the Cost of Capital. *Journal of Finance*, 46(4):1325–1359.

Dutta, S. and Nezlobin, A. (2017). Information disclosure, firm growth, and the cost of capital. *Journal of Financial Economics.*, 123(2):415–431.

Easley, D. and O'hara, M. (2004). Information and the cost of capital. *The Journal of Finance*, 59(4):1553–1583.

Espinosa-Méndez, C., Gorigoitía, J., and Vieito, J. (2017). Is the virtual integration of financial markets beneficial in emerging markets? evidence from mila. *Emerging Markets Finance and Trade*, 53(10):2279–2302.

Fosu, S., Danso, A., Ahmad, W., and Coffie, W. (2016). Information asymmetry, leverage and firm value: Do crisis and growth matter? *International Review of Financial Analysis*, 46:140–150.

Frazier, K. B., Ingram, R. W., and Tennyson, B. M. (1984). A methodology for the analysis of narrative accounting disclosures. *Journal of Accounting Research*, pages 318–331.

Garay, U. and González, M. (2008). Corporate governance and firm value: The case of venezuela. *Corporate Governance: An International Review*, 16(3):194–209.

Garay, U., González, M., Guzmán, A., and Trujillo, M. A. (2013). Internet-based corporate disclosure and market value: Evidence from Latin America. *Emerging Markets Review*,

17(1):150–168.

González, M., Téllez, D., Trujillo, M. A., et al. (2019). Governance, sentiment analysis, and initial public offering underpricing. *Corporate Governance: An International Review*, 27(3):226–244.

Greenstein, M. M. and Sami, H. (1994). The impact of the SEC's segment disclosure requirement on bid-ask spreads. *The Accounting Review*, pages 179–199.

Haggard, K. S., Martin, X., and Pereira, R. (2008). Does voluntary disclosure improve stock price informativeness? *Financial Management*, 37(4):747–768.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., Tatham, R. L., et al. (2006). Multivariate data analysis (vol. 6).

Hassan, O. A. (2018). The impact of voluntary environmental disclosure on firm value: Does organizational visibility play a mediation role? *Business Strategy and the Environment*, 27(8):1569–1582.

Hox, J. J., Moerbeek, M., and Van de Schoot, R. (2017). *Multilevel analysis: Techniques and applications*. Routledge.

Jensen, M. C. and Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of financial economics*, 3(4):305–360.

Jo, H. and Harjoto, M. A. (2011). Corporate governance and firm value: The impact of corporate social responsibility. *Journal of business ethics*, 103(3):351–383.

La Porta, R., Lopez-de Silanes, F., Shleifer, A., and Vishny, R. (2002). Investor protection and corporate valuation. *The Journal of Finance*, 57(3):1147–1170.

Lambert, R., Leuz, C., and Verrecchia, R. E. (2007). Accounting information, disclosure, and the cost of capital. *Journal of Accounting Research*, 45(2):385–420.

Lang, M. H., Lins, K. V., and Miller, D. P. (2003). Adrs, analysts, and accuracy: Does cross listing in the united states improve a firm's information environment and increase market value? *Journal of Accounting Research*, 41(2):317–345.

Leuz, C. and Verrecchia, R. E. (1999). The Economic Consequences of Increased Disclosure. *Ssrn*, 38(2000):91–124.

Lomax, R. G. and Schumacker, R. E. (2004). *A beginner's guide to structural equation modeling*. psychology press.

Long, J. S. (1983). *Confirmatory factor analysis: A preface to LISREL*, volume 33. Sage Publications.

Lopes, A. B. and de Alencar, R. C. (2010). Disclosure and cost of equity capital in emerging markets: the brazilian case. *The International Journal of Accounting*, 45(4):443–464.

Loughran, T. and Mcdonald, B. (2014). Measuring readability in financial disclosures. *J. Finance*, 69(4):1643–1671.

Loughran, T. and Mcdonald, B. (2016). Textual Analysis in Accounting and Finance: A Survey. *Journal of Accounting Research*, 54(4):1187–1230.

Mak, Y. T. and Kusnadi, Y. (2005). Size really matters: Further evidence on the negative relationship between board size and firm value. *Pacific-Basin finance journal*, 13(3):301–318.

Mangena, M., Li, J., and Tauringana, V. (2016). Disentangling the effects of corporate disclosure on the cost of equity capital: a study of the role of intellectual capital disclosure. *Journal of Accounting, Auditing & Finance*, 31(1):3–27.

Mellado, C. and Escobari, D. (2015). Virtual integration of financial markets: A dynamic correlation analysis of the creation of the latin american integrated market. *Applied Economics*, 47(19):1956–1971.

MILA (2017). Mila, el mercado mas grande de latinoamerica. *MILA*.

Myers, S. C. (1977). Determinants of corporate borrowing. *Journal of financial economics*, 5(2):147–175.

Myers, S. C. and Majluf, N. S. (1984). Corporate financing and investment decisions when firms have information that investors do not have. *Journal of financial economics*, 13(2):187–221.

Newsom, J. T. (2015). *Longitudinal structural equation modeling: A comprehensive introduction*. Routledge.

Patel, S. A., Balic, A., and Bwakira, L. (2002). Measuring transparency and disclosure at firm-level in emerging markets. *Emerging Markets Review*, 3(4):325–337.

Schreder, M. (2018). Idiosyncratic information and the cost of equity capital: A meta-analytic review of the literature. *Journal of Accounting Literature*, 41:142–172.

Szigriszt, F. (1993). *Sistemas predictivos de legilibilidad del mensaje escrito: fórmula de perspicuidad*. PhD thesis, Universidad Complutense de Madrid, Servicio de Publicaciones.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of finance*, 62(3):1139–1168.

# A Appendix

## A.1 variables Description

- Dependent Variable

  Tobin's Q is measured as the market value of assets divided by the book value of assets.

$$Tobin'sQ = \frac{Total\ Market\ Value\ of\ Firm}{Total\ Asset\ Value\ of\ Firm} \tag{A.1}$$

  ROA is measured as the ratio of firm's Net Income to Total Assets.

$$ROA = \frac{Net\ Income}{Total\ Assets} \tag{A.2}$$

- Control Variables

  Firm Size is measured as the natural logarithm of total revenue.

$$Size = Ln(Revenue) \tag{A.3}$$

  Leverage is measured as the liabilities to equity ratio.

$$Leverage = \frac{Total\ Liabilities}{Total\ Equity} \tag{A.4}$$

  Growth Opportunities is measured as the growth rate of revenue.

$$Growth = \frac{Revenue_t - Revenue_{t-1}}{Revenue_t} * 100 \tag{A.5}$$

- Independent Variables

  **Similarity:**

  We use Similarity also known as Cosine Similarity. This is a measure for narrative disclosure introduced by Brown and Tucker (2011).

$$Similarity = cos(\theta) = \frac{v_1}{\|v_1\|} \cdot \frac{v_2}{\|v_2\|} = \frac{v_1 . v_2}{\|v_1\| \|v_2\|} \tag{A.6}$$

  Where,

  $v_1$ is the vector for document 1 and $v_2$ is the vector for document 2.

  $\theta$ is the angle between $v_1$ and $v_2$

  $\|v_1\|$ and $\|v_2\|$ are the vector length of $v_1$ and $v_2$ respectively

Similarity is bounded between 0 and 1, with a higher score indicating more similarity.

**Readablity:**

We use the FLesch-Szigriszt Readability Index (IFSZ) for Spanish texts proposed by Szigriszt (1993).

$$IFSZ = 206.835 - 62.3\,\frac{S}{P} - \frac{P}{F} \tag{A.7}$$

The score is bounded between 0 and 100, where a value between 0-50 indicates that the text is difficult; between 55-65 is normal and a score higher than 70 indicates that the text is easy.

$S$ it the total number of syllables, $P$ the total number of words and $F$ the total number of phrases.

**Positive and Negative Tones:**

We use the positive and negative dictionaries for Spanish proposed by González et al. (2019). The procedure is to search inside the document for a list of words "bags of words" that are likely to be associated with positive and negative language.

$$score_i^{tf.idf} = \frac{1}{(1 + log\ a_i)} \sum_{j=1}^{J} w_{ij}^{tf.idf} \tag{A.8}$$

Where, $a$ is the total number of words in document i, $w_{ij}^{tf.idf}$ is the weight of each term in the document and $J$ is the total number of words.

**Length:**

This is a textual measure that is built from the word count of a text.

$$Length = Length(Text) \tag{A.9}$$