The effect of annual report narratives on the cost of capital in the Middle East and North Africa: A machine learning approach

Mousa, Gehan ; Elamir, Elsayed; Hussainey, Khaled

Research in International Business and Finance

DOI:
10.1016/j.ribaf.2022.101675

Published: 01/12/2022

Peer reviewed version

The Effect of Annual Report Narratives on the Cost of Capital in the Middle East and North Africa: A Machine Learning Approach

Gehan A. Mousa
Faculty of Commerce, Benha University, Egypt

Elsayed A.H. Elamir
College of Business Administration, University of Bahrain, Kingdom of Bahrain

Khaled Hussainey
Faculty of Business and Law, University of Portsmouth, Portsmouth, UK

Abstract

This paper contributes to accounting literature by reexamining the impact of the quantity and readability of annual report narratives on cost of capital. This study employs a machine learning technique, namely, the model-based (MOB) recursive partitioning, while the least absolute shrinkage and selection operator is used to select variables from a sample of 720 bank–year observations from eight Middle Eastern and North African countries between 2008 and 2019. The model-based (MOB) recursive partitioning works with local and global models to explore hidden information in the data that leads to better results in both linear and nonlinear relationships. Our analysis shows that, on one hand, the readability of annual report narratives has an insignificant impact on cost of capital. On the other hand, it shows that the greater the amount of narrative disclosure, the lower the cost of capital, a result that varies between countries and according to corporate profitability.

Keywords: Readability, Corporate Disclosure, Annual Report Narratives, Machine Learning Techniques, Cost of Capital, Middle Eastern and North African (MENA) Countries.

1 Introduction

Corporate annual reports are one of the most effective communication channels between firms and stakeholders. It is the primary source of financial and non-financial information for stakeholders in the decision-making process (Penrose, 2008; Uyar, 2011; Moumen et al., 2015; Tran, 2022). Li (2010) demonstrated that the amount of narrative disclosure in corporate financial reports has increased over the years and that stakeholders, such as investors, rely on this information in making their investment decisions. If these disclosures are complex and difficult to read, then this weakens decision-makers' ability to make relevant decisions (Lim et al., 2018; Li, 2008). One of the most important characteristics for the quality of corporate annual reports is readability (Luo et al., 2018).

Prior research has examined the consequences of the readability of annual reports, in particular, how readability affects firm performance (Clatworthy and Jones, 2006; Loughran and McDonald, 2011), earnings persistence (Li, 2008), investment efficiency (Biddle et al., 2009), stock prices and market reaction (Tetlock, 2007; Lee, 2012), earnings management (Ajina et al., 2016), and the cost of capital (Ertugrul et al., 2017; Kepir, 2018; Hassan et al., 2018). Many studies have indicated that the readability of financial reports affects the quality of understanding the content of such reports. Therefore, the decisions of stakeholders such as investors, owners, creditors, and others can be affected. For example, financial reports that are difficult to read or that are not read may mislead decision-makers, as investors may not be able to infer information pertinent to earnings management. In addition, financial analysts may incorrectly predict corporate financial performance or share prices, and financial markets may therefore react negatively. Ultimately, financial transactions in these markets may be affected disastrously (Li, 2008; Lehavy et al., 2011; Loughran and McDonald, 2011; Lee, 2012; Ertugrul et al., 2017). Therefore, a low level of readability of financial reports reduces the likelihood of an in-depth understanding of the information therein and may weaken stakeholders' decision-making process.

The existing literature addresses various aspects of the financial report readability. Ahmed et al. (2013) examined the readability of 169 Islamic banks' mission statements for 2012 using several readability measures such as Flesch Reading Ease Score (FRES), the Flesch-Kincaid Grade Level score, Average Rate Index, Composite leading indicator, the Fog index, and the Simple Measure of Gobbledygook index. The authors found that these statements are difficult to understand, and it requires a person with at least 17 years of formal schooling to comprehend them. Using a large sample of U.S. firms covering the 2004–2016 period, Xu et al. (2020) determined that the greater the readability of corporate reports, the more commercial credit that companies were able to obtain from suppliers. Banks may ask companies with vague financial reports characterized by poor readability for more financial guarantees than others, and lenders may change the contract terms (Ertugrul et al., 2017). The literature also (Livingston and Zhou, 2010; Bonsall and Miller, 2017) shows that companies with unreadable or poorly readable financial reports receive a lower credit rating, an increased cost of capital, and tougher loan terms as well as experience fluctuations in the price of their shares.

In this study, we test the effect of annual report readability as a measure of disclosure quality and the effect of disclosure quantity on the cost of capital using a sample of listed commercial banks between 2008 and 2019 after controlling for several factors suggested by the accounting literature. Our study focuses on eight Middle Eastern and North African (MENA) countries for several reasons. First, as explained by Ghosh, "local banks are the mainstay of external finance for companies in the region, which typically have high family involvement.

The close lending relationship means that the focus on disclosure norms is not especially compelling" (2018, p. 79). Second, as discussed by Aljughaiman and Salama (2019), the banking sector in the MENA region is young (most banks having only been established in the 1970s or subsequently). These countries seek to develop their financial markets by encouraging listed companies to provide appropriate information to decision-makers. This study focuses on the banking sector because this is one of the most vital, leading sectors in this region. Our results indicate that annual report narratives' readability has an insignificant impact on the cost of capital in global and local models, except for the local model at Node 2 related to Egypt, which reports that readability has a significantly positive impact on the cost of capital. In addition, our findings show that disclosure quantity has a negative effect on the cost of capital. Such a negative impact varies among various countries and is affected by corporate profitability.

We offer a novel and important contribution to the accounting literature. Although many studies have been conducted (e.g., Lehavy et al., 2011; Loughran and McDonald, 2011; Lee, 2012; Ertugrul et al., 2017; Ginesti et al., 2017; Nazari et al., 2017; Garel et al., 2019) that examine the effect of readability and disclosure quantity on the cost of capital, they have relied primarily on traditional regression models. These models feature several problems. For example, traditional regression models often identify the significant variables in the model by determining the direction of the relationship (i.e., positive or negative), and some of them may ignore the nature of the relationship between the variables (i.e., linear or nonlinear). Consequently, these problems might weaken the findings of prior research. Our study contributes to the literature by using model-based (MOB) recursive partitioning, which is a unique method that can create a flexible model that works with both linear and nonlinear relationships between predication and interpretability of the model, showing the results as a learning regression tree. The model-based (MOB) recursive partitioning provides the strength of three things: regression trees (that relate outcome to their features by recursive binary partition), covariates variables, and parametric model to improve the model's performance. Furthermore, the model-based (MOB) recursive partitioning can automatically manage several types of data (categorical and numerical) and accommodate interaction effects among the features. Finally, it works with local and global models to explore hidden information in the data that leads to better results.

This paper proceeds as follows: Section 2 presents theoretical perspectives, a literature review, and hypothesis development. Section 3 discusses the sample, the measurement of variables, and the methodology. Section 4 provides data analysis and results, and Section 5 concludes.

2 Theoretical perspectives, literature review, and hypothesis development

Narrative disclosures incentivize companies achieving several benefits, such as reducing information asymmetry, agency costs and borrowing costs (Jensen and Meckling, 1976; Boubaker et al., 2015; Kothari et al., 2009; Ataullah et al., 2018; Athanasakou et al., 2020). In addition, they are an effective means of communication that can influence stakeholders. These disclosures may be used to send specific signals that may positively impact corporate reputation; consequently, they may yield several commercial benefits (Graham et al., 2005; Lundholm and van Winkel, 2006; Armitage and Marston, 2008). From a theoretical perspective, corporate narrative disclosures could have a positive, neutral, or even a negative impact on the cost of equity capital. The positive impact arises when narrative disclosures are

successful in reducing information asymmetry. Based on agency theory (Jensen and Meckling, 1976; Fama, 1980), a conflict of interests exists between managers, as the agents of the owners, and shareholders, as principals. Therefore, agency problems such as information asymmetry and agency costs arise. Consequently, managers may use narrative disclosure as a vital tool for reducing both information asymmetry and agency costs (Patelli and Prencipe, 2007; Kothari et al., 2009). These disclosures may also mitigate information risk, which will reduce the cost of capital.

The accounting literature argues that managers have a motivation to increase the quantity or quality of narrative disclosures. They use these disclosures to send a signal to various stakeholders. Consequently, they may derive a set of benefits from disclosure activities, including enhancing corporate reputation or image, attracting additional investment, and decreasing borrowing costs (Campbell et al., 2001; Bin-Abdullah and Ku Ismail, 2008; Bini and Dainelli, 2010; Basuony and Mohamed, 2014). Signaling theory suggests that companies with positive financial performance may send signals to stakeholders by increasing disclosure or financial report readability to differentiate themselves from competitors (Campbell et al., 2001; Connelly et al., 2011).

In contrast, managers have an incentive to manipulate the information disclosed in financial reports in many ways, such as choosing specific information to disclose, using a specific tone, or utilizing a complex linguistic writing style (Brennan and Merkl-Davies, 2013). According to the obfuscate information hypothesis, managers deliberately influence decision-makers by manipulating the level of narrative disclosure in financial reports as well as the readability of such reports.

Courtis argues that management may conduct obfuscation behavior, which means that management may use "a narrative writing technique that obscures the intended message, or confuses, distracts, or perplexes readers, leaving them bewildered or muddled" (2004, p. 294). For example, firms may obfuscate their annual reports by making the narrative sections of their reports more difficult to read when they have poor financial performance (Courtis, 2004; Li, 2008; Bloomfield, 2008; Ajina et al., 2016; Lo et al., 2017). Under the obfuscate information hypothesis, increasing the quantity of narrative disclosures or increasing their quality by enhancing their readability will not lead to a decrease in information risk. Therefore, the cost of capital will not decrease; on the contrary, it will increase, "especially if investors perceive overly lengthy narratives as more difficult to process and as increasing information risk" (Athanasakou et al., 2020, p. 31). Prior studies on the association between the quantity or quality of corporate disclosures and the cost of capital have yielded conflicting results.

A growing body of literature has documented a negative linear association between the level of corporate disclosure and the cost of equity capital (Botosan, 1997, 2006; Francis et al., 2008; Kothari et al., 2009; Beyer et al., 2010). A group of studies has been conducted using a sample of U.S. companies to examine the relationship between narrative disclosure and the cost of capital, such as Botosan (1997), Francis et al. (2008) and Kothari et al. (2009). These studies have discovered a significantly negative relationship between the two variables. Using data from Swiss market, Petrova et al. (2012) found that the higher the level of voluntary disclosure, the lower the cost of capital. Baimukhamedova et al. (2017) discovered the same results using data from Kazakhstan. Using a sample of 190 firm–year observations from different four European markets—namely, Portugal, Ireland, Greece, and Spain—Martins (2014) reported a significantly negative association between International Financial

Reporting Standards (IFRS) 7 disclosure and the corporate cost of debt. Ataullah et al. (2018) showed a significantly negative association between disclosure quality (measured by the optimistic tone of corporate financial reports) and financial leverage. Recently, in the United Kingdom, Athanasakou et al. (2020) documented a nonlinear relationship between the quantity of corporate disclosure and the cost of equity capital; at a low level of disclosure, the relationship between the two variables is negative, while it becomes positive at a high level of disclosure.

Based on the above arguments, this study expects to encounter a negative association between corporate disclosure and the cost of capital. Consequently, we formulate the following hypothesis:

H1: The greater the quantity of corporate narrative disclosures, the lower the cost of capital.

Moreover, the accounting literature has documented a strong relationship between the quality of corporate disclosure and the cost of capital. Many studies have indicated that the greater the quality of disclosure, the greater the degree lenders' influence regarding loan terms and agreements (Hassan, 2009; Hassan et al., 2018). Farrelly et al. (1985) provided evidence on a significant relationship between the quality of corporate disclosure and the risk of debt financing. The authors found that the lower the quality of disclosure in the form of lower readability, the greater the financial leverage. Boubaker et al. (2019) determined that less readable filings decrease stock liquidity. In this line, Elamer et al. (2020, 2019) examined the association between risk disclosure and corporate governance (national–Islamic) in the banking sector. While Elamer et al. (2021) investigated the importance of risk disclosure for decision-makers in debt markets, Bourgain et al. (2012) evaluated the relationship between financial openness and risk disclosure in the banking industry. Contrary to previous studies, this study supplements the literature on MENA countries by examining the impact of corporate disclosure (i.e., quantity and quality) on the cost of capital. In recent studies conducted by Albarrak et al. (2019; 2020), the results have demonstrated that a company can achieve financial benefits, such as reducing its equity financing, if it increases available information about its environmental performance, such as carbon emissions, through Twitter.

Several studies have investigated the association between the readability of annual reports and the cost of capital and have reported conflicting results (Smith et al., 2006; Abu Bakar and Ameer, 2011; Eliwa et al., 2016; Guay et al., 2016; Ertugrul et al., 2017; Ginesti et al., 2017; Nazari et al., 2017; Garel et al., 2019). On one hand, many prior studies have found a negative association between the two variables. Guay et al. (2016) reported that firms use narrative disclosure to mitigate the negative effects of poorly readable company reports on the cost of capital. In this line, Ertugrul et al. (2017) investigated the impact of readability and tone of corporate annual reports on the cost of external financing using a sample of U.S. listed firms between 1995 and 2013. They determined that firms with poorly readable financial reports and an ambiguous tone had higher borrowing costs. In the United Kingdom, Eliwa et al. (2016) found a significantly negative association between the cost of equity and the quality of financial reporting by a sample of U.K. listed firms during the 2005–2011 period. Halim and Soenarno (2018) found a significantly negative association between annual report readability and the interest rate of loans. Such results indicate that the higher an annual report readability score, the lower the interest on loans, and the increase in the readability score indicates the difficult readability of annual reports. The authors explained in their findings that Indonesia is an inefficient market and that companies may have reported poor

financial performance. Thus, lenders may have wanted to obtain more guarantees by modifying borrowing conditions or raising the interest rate. Halim and Soenarno (2018) provided evidence that firms with poorly readable financial reports received lower agency costs; tax planning may be a reason for this. Rjiba et al. (2021) provided recent evidence that firms with greater textual complexity encountered a higher cost of equity capital.

Conversely, Kepir (2018) found no significant association between the readability of annual reports as measured by the BOG index and the cost of equity for a sample of U.S. firms. In Malaysia, Abu Bakar and Ameer (2011) examined the readability of corporate social responsibility (CRS) reports for a sample of 333 listed Malaysian companies in 2007. The authors reported an insignificantly negative association between the readability of CSR reports and corporate financial gearing. Comparable results have been reported by Smith et al. (2006), who analyzed the readability of the chairman's statement for a sample of Malaysian companies. In Qatar, Hassan et al. (2018) found an insignificantly negative association between the readability of narrative disclosures for a sample of 126 listed firms and financial leverage. In contrast, some studies have presented a positive association between the readability of financial reports and corporate financial leverage, such as Ginesti et al. (2017) in Italy and Nazari et al. (2017) in the United States.

According to the preceding discussion, this study expects to encounter a negative association between the readability of corporate disclosures and the cost of capital. Therefore, the following hypothesis is formed:

H2: The greater the readability of corporate annual reports, the lower the cost of capital.

3 Research methodology

3.1 Sample and data collection

Our sample consists of 720 bank–year observations for 60 commercial banks from eight MENA countries—namely, Egypt (EG); Jordan (JO), Bahrain (BAH), the United Arab Emirates (UAE), Saudi Arabia (SA), Kuwait (KU), Qatar (QA), and Oman (OM). Islamic banks have been excluded because they are managed and on different laws and regulations. Banks are selected in the current study since they have a unique importance to various countries' economies and financial markets. In addition, they represent a high percentage of the stock markets' trading volume in the selected countries. This study depends on a sample of eight emerging markets. Such markets have some advantages, such as a higher growth rate, and may therefore attract many investors. Therefore, our study may add value for several stakeholders in light of the paucity of research available on these markets. Regarding data collection, this study relies on various sources—namely, the Thomson Reuters database, the annual reports of the sample banks in the study, the websites of these banks, and, finally, data from the stock exchanges in the eight countries. Table 1 shows the distribution of our sample by country between 2008 and 2019.

[Insert Table 1 Here]

[Table 1: Frequency Distribution of the Sample of Banks by Country and Year**]**

3.2 Measurement of the study's variables

3.2.1 Cost of capital measure as a dependent variable

The accounting literature offers two approaches to calculate the cost of capital: the price-earnings valuation models (Easton, 2004; Ohlson and Juettner-Nauroth, 2005) or the residual income valuation models (Claus and Thomas, 2001). Kothari (2009) reported that price earnings models are better than the residual income valuation models in explaining the cross-sectional differences in stock prices. Clère (2019) suggested a different measure of the cost of debt based on default risk and systematic risk. Our study follows a stream of studies that use price earnings models. It employs the price earnings growth ratio (PEG) model (Easton, 2004). The PEG model has been used in several studies (Vena et al., 2019; Easton, 2004; Gray et al., 2009; Eliwa et al., 2016). The cost of capital variable is calculated as follows:

earnings per share in period ; price in period (Easton, 2004).

3.2.2 Independent variables (readability and disclosure quantity measures)

3.2.2.1 Readability measures (the Fog index and the FRES)

The current study has employed two measures of readability as follows. First, the Fog index was introduced by Robert Gunning in 1952 and is considered one of the most widely used indicators in the accounting literature to measure the readability of annual reports (Clatworthy and Jones, 2006; Biddle et al., 2009; Lawrence, 2013; Lehavy et al., 2011; Li, 2008). The Fog index reflects the number of official school years that a person requires to understand a document when reading it for the first time. Therefore, the higher the Fog index, the more difficult it is to read a document, indicating that a person with many years of academic experience, or a person with a high level of knowledge, is required to understand the document. The Fog index scores academic writing according to the following levels: 8–10 is very easy, 10–12 is acceptable, 12–14 is ideal, 14–18 is hard, and greater than 18 is unreadable (Biddle et al., 2009; Li, 2008). To measure readability, the Fog index depends on two main factors: the number of words in a sentence and the number of complex words in a sentence, which is measured by the number of syllables in the word. A word is complex or difficult to read if it contains more than three syllables (Ahmed et al., 2013; Guay et al., 2016). The Fog index formula (Gunning, 1952) is calculated as follows:

where average sentence length (ASL) = the number of words/the number of sentences; Nwmin3sy = the number of words with three syllables or more; and Nw = the number of words. Loughran and McDonald (2014) criticized the Fog index in business writings as several terms in the field of business such as "operations," "financial," and "customers" consist of more than three syllables and these terms are common and understandable in this field among specialists, while the Fog index classifies them as complex words. However, this index is still popular and used by many researchers in business (Biddle et al., 2009; Li, 2008; Loughran and McDonald, 2011; Ahmed et al., 2013; Xu et al., 2020).

Second, we use the (FRES), suggested by Rudolph Flesch in 1948, as a measure of estimating the readability of a document by school children or adults in the United States (Clatworthy and Jones, 2006; Smith et al., 2006; Abu Bakar and Ameer, 2011; Sattari et al., 2011). In contrast to the Fog index, the FRES focuses on the document, not the educational level of the reader. The higher the FRES, the more readable the document is. The FRES depends on two linguistic attributes: the average sentence length in words and the syllable count (expressed as the number of syllables per 100 words). The FRES (Flesch, 1948, p. 309) is calculated as follows:

where AWpS: the average words per sentence is measured by dividing the total number of words by the total sentences, and ASpW: the average syllables per word is calculated by dividing the total number of syllables by the total number of words (Bayerlein and Davidson, 2011). The FRES has a 100-point scale. Therefore, if the FRES of the document is 60 or greater, then the document is readable, while a score of less than 50 indicates that it is difficult to read, and a score of less than 30 indicates that the document is unreadable.

The Fog index and the FRES have been criticized as readability measures. For example, they ignore the characteristics of the document's reader such as their experience, culture, and knowledge, which impact the reader's ability to understand the content of the document. In addition, these measures were introduced a long time ago, and no significant improvements have been made to them despite the many advancements in English-language writing styles (Smith et al., 2006; Bayerlein and Davidson, 2011). This study uses these measures to determine readability for two reasons. The first is that these measures are still popular and supported by researchers due to their ease of calculation and their ability to measure readability without affecting the order of the sentences in the document. The second reason is that these measures have been used in several previous studies, which enables us to compare our results with those of other studies.

3.2.2.2 Disclosure quantity measure

The current study uses the number of sentences in the narrative sections of banks' annual reports as a proxy for disclosure quantity.

3.2.3 Control variables

Following prior research, we include nine control variables. Hail and Leuz (2006) included market beta, book-to-market ratio, firm size, and financial leverage as control variables. Other studies (Guedhami and Mishra 2009; Zéghal et al., 2011; Petrova et al., 2012; Martins, 2014; Ertugrul et al., 2017; Kepir, 2018; Hassan et al., 2018; Halim and Soenarno, 2018; Garel et al., 2019; Athanasakou et al., 2020) also used these variables as well as others such as market to book ratio, the age of the bank, the time period (year dummies), and country. Table 2 presents details regarding the measurement of the variables.

[Insert Table 2 Here]

[Table 2: Definition of the Variables**]**

3.3 Analytical methodology

The current study has employed two methods, namely, the least absolute shrinkage and selection operator (LASSO) and model-based recursive partitioning using the model-based (MOB) recursive partitioning.

3.3.1 Least absolute shrinkage and selection operator

The multiple linear regression model can be expressed as follows:

the dependent variable, independent variables, parameters and error.
The ordinary least squares (OLS) method estimates the parameters by minimizing the sum of squares of errors or the square the loss function as

which reflects the estimated values (Hastie et al., 2009). In theory, the estimators of OLS are unbiased and may have a large variance in case of (a) there is multicollinearity among independent variates, and (b) there are many independent variables. The collinearity causes coefficients for correlated variables to become inflated and fluctuate significantly. One consequence of the fluctuations is overfitting for coefficients, i.e., high variance in the bias-variance space (Kuhn and Johnson, 2013; Hastie et al., 2009; Kutner et al., 2005).

The regularization method (LASSO) is used to make variance reduction in the cost of the bias as one of the leading techniques in this direction. LASSO adds a term of non-zero coefficients as

the parameter numbers (Tibshirani, 1996; Hastie et al., 2007). LASSO model consists of the square errors and -penalty term. penalty has the smoothing parameter and the total coefficient absolute. In general, the LASSO has three advantages in comparison with OLS regression. It makes a reduction in model variance, automatic selection for feature variables, and it gives zero as a coefficient of non-important features (Joliffe et al., 2003; Wu and Lange, 2008; Shrivastava et al., 2020).

Model-Based (MOB) Recursive Partitioning

Model-based (MOB) recursive partitioning (Zeileis et al., 2008) is a general approach to account for heterogeneity in the data in which data are divided into a response variable (Y), feature variables that are separated into the main variable(s) (variables of interest, , and control or partition variables (Z) as follows:

,

Where there are variables of interest and partitions or control variables. The fundamental thought is that every node is related to one model. To evaluate whether a partition of the node is essential, a fluctuation test for unstable parameters is implemented. If there is significant instability regarding any partition variables,, divide the node into K locally optimal portions

and iterate the procedure. If no additional significant instabilities exist, the recursive procedures stop and a tree is returned with every leaf related to a model of type , and is the parameter.

In some cases, it is unsuitable to use a single global model to fit all of the data. Still, it may be more convenient to divide the data into known covariates or control variables, such that a well-fitting model can be obtained locally in every part of the partition. In these cases, a recursive partitioning approach using partitioning features () can be used to adaptively find a good fit for this partition (Hothorn and Zeileis, 2020).

The steps for algorithm (Zeileis et al., 2008) are as follows:

1. Use all observations to fit the model in the current node by estimating and by minimizing the objective function and is given observations.
2. Evaluate if the parameter estimates are steady for each ordering from .
3. In case there any of the variables are unstable, choose related to the maximum parameter instability; otherwise, stop.
4. Compute the split point(s) that locally optimize the objective function .
5. Split the note into child nodes and repeat the procedure.

It may be noted that the splits are deleted from the tree when they do not improve the fitting of the model to the Akaike information criterion (Akaike, 1974) or the Bayesian information criterion (Schwarz, 1978). The advantages of using this technique are (a) the objective function is used for both parameter estimation and partition, (b) the recursive splits permit automated exposure for interactions between the feature variables and for modeling nonlinear relationships, and (c) yields a segmented model that simplifies and clarifies the interpretation of the model (Strobl et al., 2011; Hothorn and Zeileis, 2020).

4 Data analysis and results

4.1 Descriptive analysis

The descriptive statistics (i.e., mean, median, standard deviation, first quartile, third quartile, skewness, and kurtosis) for this study's variables are provided in Table 3. It may be observed that the means and medians are near each other. The skewness measure has a symmetrical distribution when its value is zero. In addition, variables such as *LEVRA*, *FOGME*, and *FLEME* have large amount of skewness while others such as *LFIRM*, *LLOAN*, *BETAC*, *AGEOB*, and *CCOST* have small amount of skewness close to zero. The kurtosis measure may be used to characterize a normal distribution among a set of symmetrical distributions when its value is 3. The variables nearest to 3 are *CCOST* and *LSENN*, while those that are the furthest from 3 are *FOGME* and *FLEME*. The standard deviations for *FLEME* and *AGEOB* are greater with respect to the others, indicating a high level of fluctuation within these variables.

[Insert Table 3 Here]

[Table 3: Descriptive Statistics and Shape Measures (Skewness and Kurtosis) of the Variables]

Figure 1 depicts the *CCOST* variable across countries and years. In terms of country, there are considerable fluctuations in *CCOST* among the countries; the highest value of *CCOST* is in EG, followed by UAE and OM, and the most similar countries are JO and SA. In terms of year, fluctuations in the *CCOST* variable can be divided into three intervals—years (08, 09, 10, 11, 12, 13, and 14), years (15, 16, 17, and 18), and, finally, year (19).

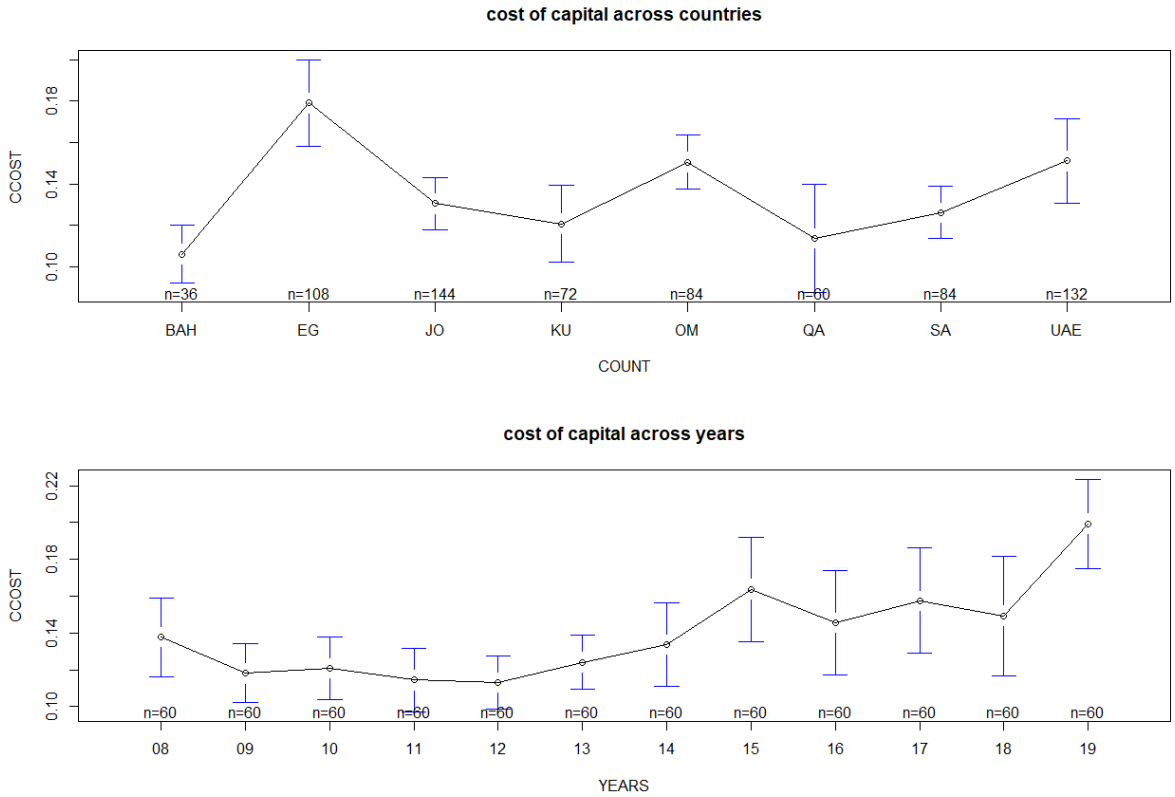

Figure 1. *CCOST* Variable across Countries (First Panel) and Years (Second Panel) in a Sample of Banks

4.2 The correlation matrix

Figure 2 displays the correlation matrix with a histogram and scatter plot of the study's variables. There is a significant linear correlation between *CCOST* and *LEVRA*, *MBOOK*, and *LSENN*. In addition, there is a significantly high linear correlation between *LFIRM* and *LLOAN* (0.94), as well as *FOGME* and *FLEME* (–0.98). These high correlations may cause multicollinearity, which affects the ability of a classical regression to recognize which features are important in the model (Kutner et al., 2005). One measure to detect multicollinearity is the variance inflation factor (VIF) (VIF > 10).

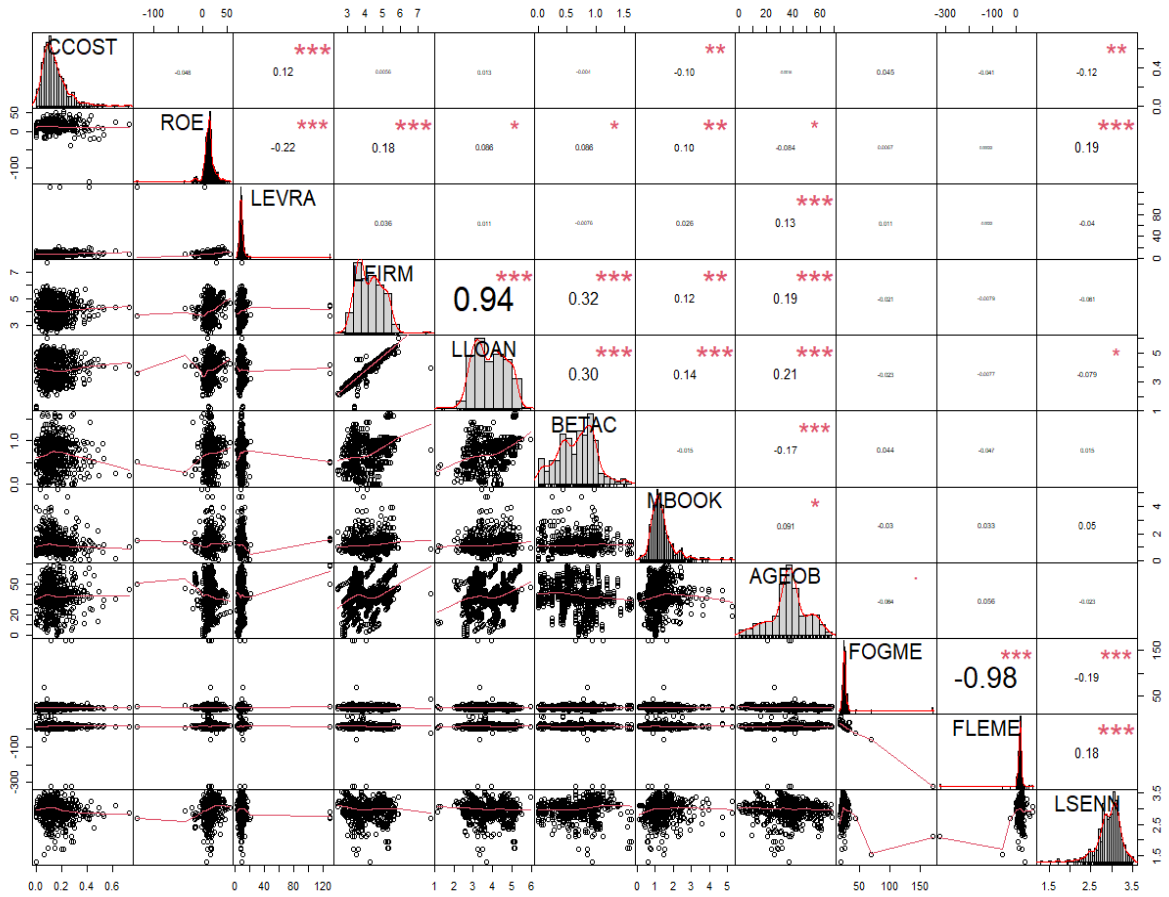

Figure 2: The Correlation Matrix of this Study's Variables

Note: (***) significant at the level of 0.001, (**) significant at the level of 0.01, (*) significant at the level of 0.05 and (.) significant at the level of 0.10.

It can be noted that Figure 3 displays a bar chart for the VIF in which *FOGME* and *FLEME* are shown to have VIFs greater than 25, while those of the other variables are less than 10. Consequently, the LASSO model is used to select variables that will be used in machine learning models (i.e., the model-based (MOB) recursive partitioning). To build these models, the data were divided into 540 observations; training data comprised 75% and 180 of the observations were used as testing data (i.e., 25%).

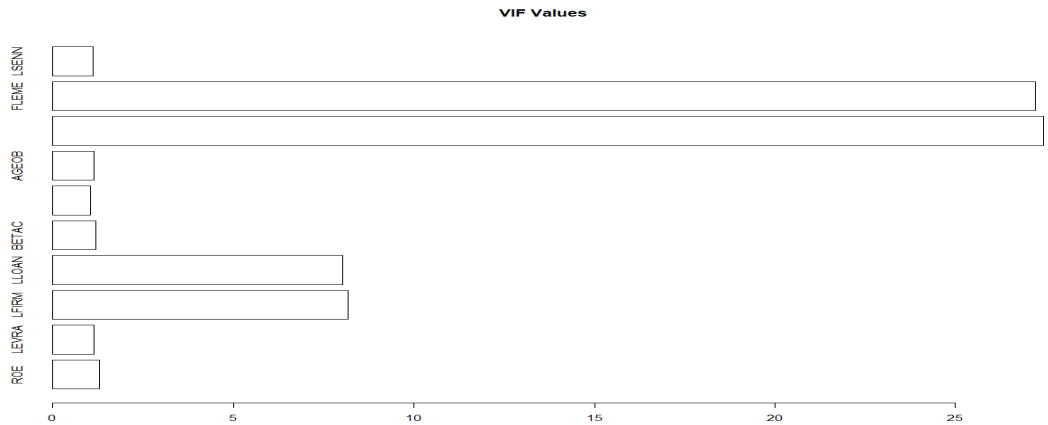

Figure 3: Variance Inflation Factor (VIF) of this Study's Variables

4.3 The LASSO method

The LASSO method is used to select relevant features in the training data and estimate the parameters of cost equity models in case of existing multicollinearity between features variables.

The selecting procedure for the LASSO method is employed to increase the estimates' prediction accuracy by removing non-important features and shrinking the estimates which can reduce variance without increasing the bias.

The general model for *CCOST* using LASSO method can be written as

The full linear version of this model can be written as

The coefficients , are the model parameters, FE stands for fixed effect, and is the model error.

Table 4 shows the regression coefficients of the LASSO model. It should be noted that collinear features are pushed toward zero, for example, the FOG index measure. The LASSO model keeps all of the variables except for *FOGME* because it has a high correlation (−0.98) with the *FLEME* variable. Except for the control variables, *FLEME* and *LSENN* have negative coefficients, which means that when *FLEME* and *LSENN* are increased by one unit, *CCOST* will decrease by –0.001 and −0.034 units, respectively.

[Insert Table 4 Here]

[Table 4: LASSO-Estimated Coefficients for the *CCOST* Model]

Figure 4 shows the optimal value of using a 10-fold cross-validation based on mean square error. It can be noted that is the tuning parameter for the LASSO regression that is used to control the overfitting of the training data.

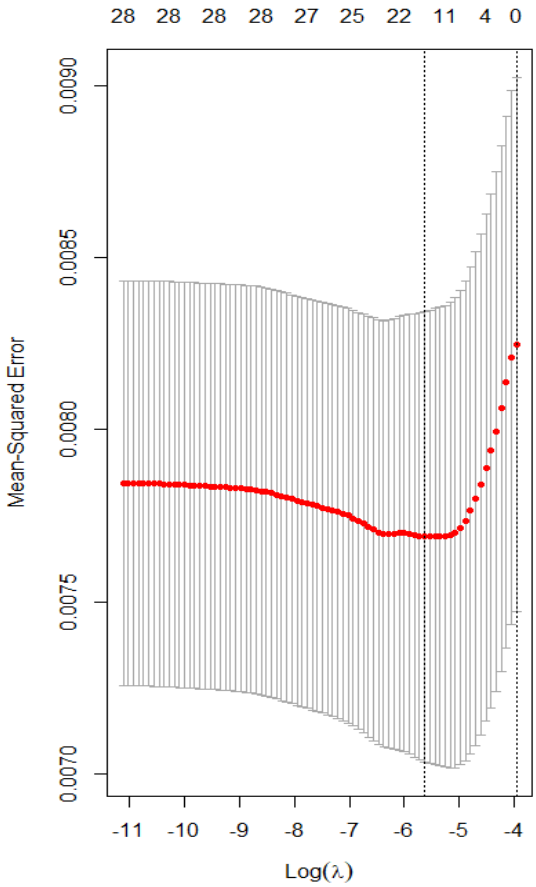

Figure 4: The Optimal Value of Using 10-Fold Cross-Validation Based on Mean Square Error Using LASSO Approach

4.4 The Model-Based (MOB) Recursive Partitioning

Fitting one global model to data could hide important details and information, especially if the data include complex relationships and strong variations. Consequently, dividing data into homogeneous groups using a set of control or partition variables can explore this hidden information by fitting a local model to each group, which often leads to better results. Since the LASSO method excluded the *FOGME* variable, the model-based (MOB) recursive partitioning can be written as follows:

The full model can be written as follows:

where , and are the model estimated coefficients

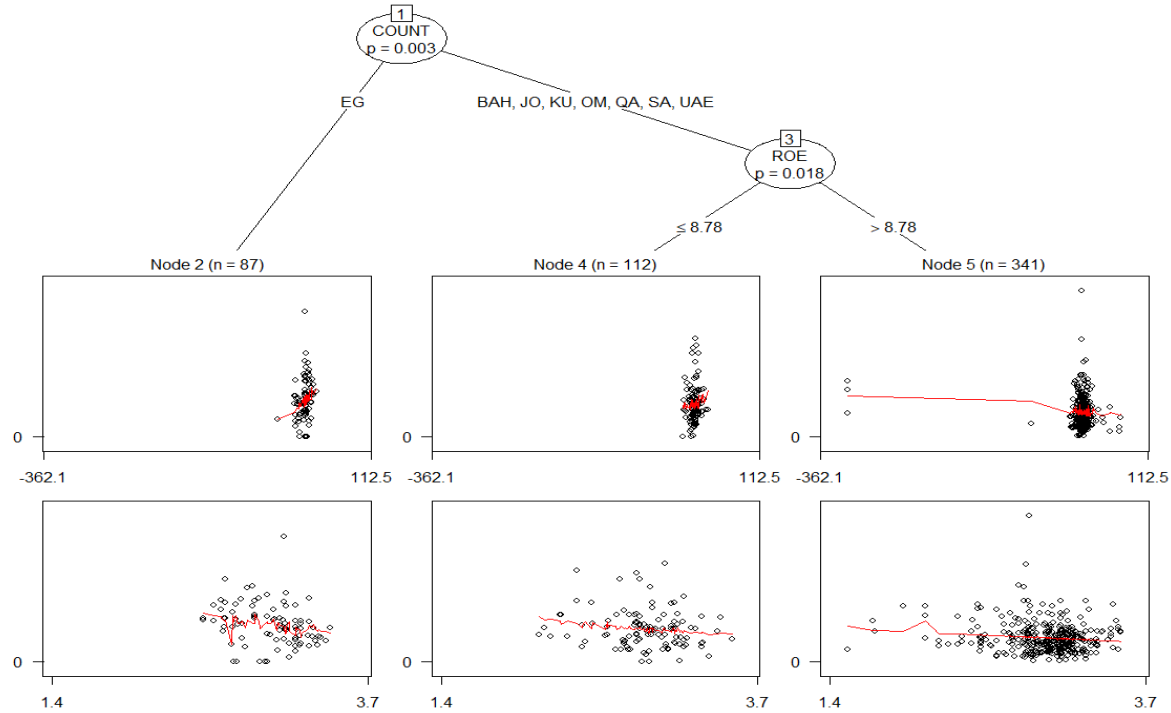

Figure 5: Model-Based Recursive Partition Tree with Partial Scatter Plot for *FLEME* versus *CCOST* (Upper Panel) and *LSENN* versus *CCOST* (Lower Panel) at the Leaves

The resulting model-based recursive partition tree is shown in Figure 5, which includes a scatter plot with a fitted linear regression in the terminals. More details for each node are provided in Table 5. In the estimating process, the global model using all of the data (720 bank–year observations) is fitted in Node 1, resulting in a country slope of about −0.001 (*FLEME*) with a p-value of 0.84 (not significant). In contrast, *LSENN* has a coefficient of −0.05 with a p-value of less than 0.01 (significant at the 1% level). The stability of the global model is evaluated with respect to all (partition variables). The instability tests are given with Bonferroni-adjusted values. It can be noted that significant instability is obtained at country, *ROE* and *LEVRA*, which is used for splitting at the lowest significant value, 0.00 (country). This gives the best split at two groups namely, EG and the remaining countries (i.e., BAH, JO, KU, OM, QA, SA, and UAE). For the country, EG, the local model at Node 2 with 87 bank–year observations, there is no more parameter instability in terms of partitioning variables that can be detected in which all values are greater than 5%. Therefore, the linear regression is fitted for this node. For Node 3, another local model, significant instability is obtained for *ROE*, which gives the best split at 10.087. For *ROE*, more than and less than 8.78, no instability parameter is detected; therefore, the linear regression is fit at Node 4 (a local model with 112 bank–year observations) and Node 5, (a local model with 341 bank–year observations).

[Insert Table 5 Here]

[Table 5: The Estimated Coefficients for the Model-Based Recursive Partition Tree (the model-based (MOB) recursive partitioning)]

Table 5 presents the global model (Node 1) and four local models (Nodes 2, 3, 4, and 5). In Table 5, *FLEME* has a significantly positive effect on *CCOST* in the local model at Node 2 (EG) with a p-value of 0.04 (significant at the 5% level). In contrast, there is no significant

effect on *CCOST* at Node 4, which includes all countries except for EG and *ROE* that are less than or equal to 8.78. Based on the results shown in Table 5, H1 is rejected in the global model since *FLEME* has an insignificant association with a p-value of 0.84 and with local models at Nodes 3, 4, and 5 (since the p-values are insignificant) (i.e., 0.51, 0.36, and 0.19, respectively) except for in local model at Node 2 (EG). Consequently, it can be concluded that the country variable is one of the determinants for *CCOST*. *FLEME* has a significant positive effect on *CCOST*. Such finding is contrasting with Eliwa et al. (2016), Guay et al. (2016) and Ertugrul et al. (2017) who find a negative association between readability measures and *CCOST*, in addition, Kepir (2018) finds an insignificant association between the two variables. While it is consistent with Ginesti et al. (2017) and Nazari et al. (2017) who show a positive association between corporate financial leverage and readability measures. Furthermore, Halim and Soenarno (2018) report that the poorer the readability of the financial reports, the lower the interest rates on loans. The authors explain that companies may have reported poor financial performance, and thus lenders modify borrowing terms or raise the interest rate to obtain more guarantees. Our finding is not in line with disclosure theories, such as agency theory and impression management theory which expect a negative association between *FLEME* and *CCOST*. Such a result can be explained by EG country is an inefficient market where theoretical assumptions and accepted standards do not apply, and it is characterized by imbalance and the absence of perfect competition.

Regarding H2, there are mixed results. For example, *LSENN* has a significant negative effect on *CCOST* in the global model (Node 1) by using nine control variables. More information can be obtained as control variables are divided. Local model at Node 3 which considers all countries except for EG country, *LSENN* has a significant negative effect on *CCOST*. Also, in the local model at Node 5 that includes all countries except for EG and *ROE* more than 8.78, *LSENN* has the same effect on *CCOST*.

On the other hand, since the p-values at Node 2 and Node 4 are 0.15 and 0.13, respectively, *LSENN* is not significant at the 5% level. Therefore, H2 is supported globally and partially in all countries, except for EG and *ROE*, when it is more than 8.78, it is rejected.

In summary, it can be concluded that, in general, *LSENN* has a significantly negative association with *CCOST*. Such a result is consistent with several previous studies (Botosan, 1997; Botosan, 2006; Francis et al., 2008; Kothari et al., 2009; Beyer et al., 2010) that have reported the same relationship between the two variables.

Furthermore, our findings refer to the country variable and *ROE* are determinants of *CCOST*. Our results can be explained by the fact that the greater the amount of corporate disclosure, the lower the cost of capital, and this result varies among countries and according to the financial performance of companies in these countries. The finding for H2 is consistent with agency theory and impression management theory, which suggest a negative association between the level of narrative disclosures and *CCOST*. Companies manage the level of narrative disclosure to achieve several benefits, and one of these benefits is reducing *CCOST*.

The greater the disclosure, the more reassurance is given to lending bodies about the financial position of the company wishing to borrow. Lending bodies may therefore provide the company with better terms for lending or reduce the interest rate on its loans.

There is a robustness test performed on the previous results. In addition to , the modified price earning measure (, Easton, 2004) is used for the cost of equity. This measure can be written as follows:

where and dividends per share at the period , is the rate of growth in abnormal earnings post forecast horizon.

This proxy is used as an alternative to *CCOST*. Equation 7 is retested after using as a realization of *CCOST*. Similar results to the non-modified measure () are found using model-based (MOB) recursive partitioning for modified measure , which in line with our main results shown in Figure 5 and Table 5.

6. Conclusion

The current study examines the effect of readability and disclosure quantity on the cost of capital through eight MENA countries using a sample of 720 bank–year observations. Our results are not consistent with the findings of previous studies that find a negative association between readability and the cost of capital (Eliwa et al., 2016; Guay et al., 2016; Ertugrul et al., 2017). It reports an insignificant association between the two variables except for the local model at Node 2 (Egypt). It finds a significant positive association between readability and the cost of capital. Also, we find, in general, a significant negative association between disclosure quantity and the cost of capital, which is consistent with several studies (Petrova et al., 2012; Martins, 2014; Baimukhamedova et al., 2017; Ataullah et al., 2018; Athanasakou et al., 2020). Such negative associations can vary among countries and are affected by financial performance.

Our study provides new insights overlooked by previous studies by using the model-based (MOB) recursive partitioning as a machine learning technique. The model-based (MOB) recursive partitioning considers both linear and nonlinear relationships and can work with local and global models to explore hidden information in the data, which ultimately leads to better results. This is the first study—to the best of our knowledge—that addresses the effect of readability and quantity of corporate disclosure on the cost of capital by means of a global and local model. Such insights provide new considerations for regulators, investors, creditors, and other stakeholders when considering the consequences of the quantity and quality of corporate disclosure.

Our findings offer interesting policy implications. First, more effort is needed to encourage bank managers in the MENA region to increase both the quality and quantity of their annual report narratives. This could be accomplished through regulators, who should find mechanisms to improve the quality of banks' corporate governance systems in the MENA region, which could lead to an improvement in banks' narrative reports (Grassa and Chakroun, 2016; Grassa, Moumen, and Hussainey, 2020; González et al. 2021). Second, our findings inform regulators about the benefits of improving the quantity and quality of banks' narrative reports to investors and disclosing firms. Since narrative sections of annual reports are still voluntary, additional effort is required to provide to improve the credibility of these narratives. Finally, our findings indicate that improving the quantity and readability of banks' narrative disclosures in the MENA region may promote more efficient capital markets.

Similar to prior studies, our research features some limitations. First, the present study considers several variables such as readability, disclosure quantity, bank size, bank age, and others. Other factors such as the economic and political systems of the country under analysis have been ignored. Future studies may examine the impact of such factors on the cost of capital. Second, this study depends on a relatively small sample from only one sector, the banks in eight countries in the MENA region. There may be a need to increase the sample size and use various sectors from several countries in future research. The findings of this study can provide a number of implications for different stakeholders such as investors, shareholders, and creditors. For example, an increase in the readability of financial reports does not necessarily lead to a decrease in the cost of capital. The matters may be related, as are the results of our study, to the country variable. Moreover, examining the relationship between disclosure levels and the cost of capital must be considered from several angles, as the direction of this relationship may change due to the influence of other factors such as the nature of the country under study and the financial performance of companies in this country. In addition, such relationships may change over the course of the study.

Conflicts of Interest: The authors do not have any conflicts of interest to declare. This research did not receive any funding from any institution or organization.

References

Abu Bakar, A. S., Ameer, R., 2011. Readability of corporate social responsibility communication in Malaysia. Corp. Soc. Responsib. Environ. Manag. 18 (1), 50–60. https://doi.org/10.1002/csr.240

Ahmed, I, Shaukat, M, Z., Islam, T., 2013. Mission statements readability: an insight into Islamic banks. J. Islamic Account. Bus. Res. 4 (2), 132–150. https://doi.org/10.1108/JIABR-04-2012-0019

Ajina, A., Laouiti, M., Msolli, B., 2016. Guiding through the fog: Does annual report readability reveal earnings management? Res. Int. Bus. Financ. 38, 509–516. https://doi.org/10.1016/j.ribaf.2016.07.021

Akaike H., 1974. A new look at the statistical model identification. IEEE Trans. Automat. Contr. 19, 716–723. https://doi.org/10.1109/TAC.1974.1100705

Albarrak, M., Elnahas, M., Papagiannidis, S., Salama, A., 2020. The effect of twitter dissemination on cost of equity: A big data approach. Int. J. Inf. Manage. 50, 1–16. https://doi.org/10.1016/j.ijinfomgt.2019.04.014

Albarrak, M., Elnahas, M., Salama, A., 2019. The effect of carbon dissemination on cost of equity. Bus. Strategy Environ. 28(6), 1179–1198. https://doi.org/10.1002/bse.2310

Aljughaiman, A. A., & Salama, A. 2019. Do banks effectively manage their risks? The role of risk governance in the MENA region. *Journal of Accounting and Public Policy*, *38*(5), 106680. 10.1016/j.jaccpubpol.2019.106680

Armitage, S., Marston, C., 2008. Corporate disclosure, cost of capital and reputation: evidence from finance directors. Br. Account. Rev. 40(4), 314–336. https://doi.org/10.1016/j.bar.2008.06.003

Ataullah, A., Vivian, A., Xu, B., 2018. Optimistic Disclosure Tone and Conservative Debt Policy. Abacus. 54(4), 445–484. https://doi.org/10.1111/abac.12140

Athanasakou, V., Eugster, F., Schleicher, T., Walker, M., 2020. Annual Report Narratives and the Cost of Equity Capital: U.K. Evidence of a U-shaped Relation. Eur. Account. Rev. 29(1), 27–54. https://doi.org/10.1080/09638180.2019.1707102

Baimukhamedova, A., Baimukhamedova, G., Luchaninova, A., 2017. Financial disclosure and the cost of equity capital: the empirical test of the largest listed companies of Kazakhstan, J Asian Finance. Econ. Bus. 4 (3), 5–17. https://doi.org/10.13106/jafeb.2017.vol4.no3.5

Basuony, M. A. K., Mohamed, E. K. A., 2014. Determinants of Internet Financial Disclosure in GCC Countries, Asian J. Financ. Account. 6(1), 70, 2014. http://dx.doi.org/10.5296/ajfa.v6i1.5085

Bayerlein, L., Davidson, P., 2011. The influence of connotation on readability and obfuscation in Australian chairman addresses, Manag. Audit. J. 27 (2), 175–198. https://doi.org/10.1108/02686901211189853

Beyer, A., Cohen, D. A., Lys, T. Z., Walther, B. R., 2010. The financial reporting environment: Review of the recent literature, J. Account. Econ. 50 (2–3), 296–343. https://doi.org/10.1016/j.jacceco.2010.10.003

Biddle, G. C., Hilary, G., Verdi, R. S., 2009. How does financial reporting quality relate to investment efficiency? J. Account. Econ., 48(2), 112–131. https://doi.org/10.1016/j.jacceco.2009.09.001

Bin-Abdullah, A., Ku Ismail, K. I., 2008. Disclosure of Voluntary Accounting Ratios by Malaysian Listed Companies, J. Fin. Rep. Account. 1–20. http://dx.doi.org/10.1108/19852510880000632

Bini, L., Dainelli F., G. F., 2010. Signalling theory and voluntary disclosure to the financial market. Evidence from the profitability indicators published in the annual report. SSRN Electronic Journal. Available at SSRN: https://ssrn.com/abstract=1930177

Bloomfield, R., 2008. Discussion of annual report readability, current earnings, and earnings persistence, J. Account. Econ. 45, 248–252. http://dx.doi.org/10.1016/j.jacceco.2008.04.002

Bonsall, S., Miller, B., 2017. The impact of narrative disclosure readability on bond ratings and the cost of debt, Rev. Account. Stud. 22 (2), 608–643. https://doi.org/10.1007/s11142-017-9388-0

Botosan, C. A., 1997. Disclosure level and the cost of equity capital, Account. Rev. 72(3), 323–349.

Botosan, C. A., 2006. Disclosure and the cost of capital: What do we know? Account. Bus. Res. 36 (Supplement 1), 31–40.

Boubaker, S., Gounopoulos, D., Rjiba, H., 2019. Annual report readability and stock liquidity. Financial Mark. Instit. Instrum. 28(2), 159–186.

Boubaker, S., Hamrouni, A., Liang, Q. B., 2015. Corporate governance, voluntary disclosure, and firm information environment. J. Appl. Bus. Res. 31(1), 89–102.

Bourgain, A., Pieretti, P., Zanaj, S., 2012. Financial openness, disclosure and bank risk-taking in MENA countries. Emerg. Mark. Rev. 13(3), 283–300.

Brennan, N., Merkl-Davies, D. M., 2013. Accounting Narratives and Impression Management Publication Date Publication information Lisa Jack, Jane Davison, Russell Craig (Eds.). The Routledge Companion to Accounting Communication P. In The Routledge Companion to Accounting Communication, 8, 109–132.

Campbell, D., Shrives, P., Bohmbach-Saager, H., 2001. Voluntary Disclosure of Mission Statements in Corporate Annual Reports: Signaling What and To Whom? Bus. Soc. Rev. 106(1), 65–87. http://dx.doi.org/10.1111/0045-3609.00102

Clatworthy, M., Jones, M, J., 2006. Differential patterns of textual characteristics and company performance in the chairman's statement, Account. Audit. Account. J. 19 (4), 493–511. https://doi.org/10.1108/09513570610679100

Claus, J., Thomas, J., 2001. Equity premia as low as three percent? Evidence from analysts' earnings forecasts for domestic and international stock markets. J. Financ. 56, 1629–1666. https://www.jstor.org/stable/2697733

Clère, R., 2019. After Modigliani, Miller, and Hamada: A new way to estimate cost of capital. J. Int. Financ. Manag. Account. 30(3), 223–249. https://doi.org/10.1111/jifm.12109

Connelly, B. L., Certo, S. T., Ireland, R. D., Reutzel, C. R., 2011. Signaling theory: A review and assessment, J. Manage. 37(1), 39–67. https://doi.org/10.1177%2F0149206310388419

Courtis, J. K., 2004. Corporate report obfuscation: artefact or phenomenon? Br. Account. Rev. 36, 291–312. http://dx.doi.org/10.1016/j.bar.2004.03.005

Easton, P. D., 2004. PE ratios, PEG ratios, and estimating the implied expected rate of return on equity capital. Account. Rev. 79(1), pp 73–95. https://doi.org/10.2308/ACCR.2004.79.1.73

Elamer, A. A., Ntim, C. G., Abdou, H. A., 2020. Islamic governance, national governance, and bank risk management and disclosure in MENA countries. Bus. Soc. 59(5), 914–955.

Elamer, A. A., Ntim, C. G., Abdou, H. A., Owusu, A., Elmagrhi, M., Ibrahim, A. E. A., 2021. Are bank risk disclosures informative? Evidence from debt markets. Int. J. Financ. Econ. 26(1), 1270–1298.

Elamer, A. A., Ntim, C. G., Abdou, H., Zalata, A., Elmgrahi, M., 2019. The impact of multi-layer governance on risk disclosure: Evidence from MENA banks. Account. Forum. 42 (2), 246–281.

Eliwa, Y., Haslam, J., Abraham, S., 2016. The association between earnings quality and the cost of equity capital: Evidence from the UK, Int. Rev. Financial. Anal. 48, 125–139. https://doi.org/10.1016/j.irfa.2016.09.012

Ertugrul, M., Lei, J., Qiu, J., Wan, C., 2017, Annual report readability, tone ambiguity, and the cost of borrowing, J. Financial Quant. Anal. 52 (2), 811–836. https://doi.org/10.1017/S0022109017000187

Fama, E. F., 1980. Agency Problems and the Theory of the Firm, J. Polit. Econ. 88(2), 288–307. https://www.jstor.org/stable/1837292

Farrelly, G., Ferris, K., Reichenstein, W., 1985. Perceived risk, market risk, and accounting determined risk measures, Account. Rev. 60 (2), 278–288.

Flesch, R., 1948. A new readability yardstick', J. Appl. Psychol. 32, 221–33.

Francis, J., Nanda, D., Olsson, P., 2008. Voluntary disclosure, earnings quality, and cost of capital, J. Account. Res. 46(1), 53–99.

Garel, A., Gilbert, A. B., Scott, A., 2019. Linguistic Complexity and Cost of Equity Capital (March 12, 2019). Available at https://ssrn.com/abstract=3240292

Ghosh, S., 2018. Governance reforms and performance of MENA banks: Are disclosures effective? Glob. Finance. J. 36, 78–95.

Ginesti, G., Sannino, G., Drago, C., 2017. Board connections and management commentary readability: the role of information sharing in Italy, Corp. Governance: Int. J. Bus. Soc. 17(1), 30–47. http://dx.doi.org/10.1108/CG-01-2016-0015

González, M., Guzmán, A., Tellez-Falla, D. F., Trujillo, M. A., 2021, Determinants of corporate tone in an initial public offering: Powerful CEOs versus well-functioning boards. Research in International Business and Finance 58 101481.

Graham, J. R., Harvey, C. R., Rajgopal, S., 2005. The economic implications of corporate financial reporting, J. Account. Econ. 40, 3–73.

Grassa, R., Chakroun, R., 2016. Ownership structure, board's characteristics and corporate governance disclosure in GCC banks: what about Islamic banks? Int. J. Account. Audit. Perform. Eval. 12(4), 360–395.

Grassa, R., Moumen, N., Hussainey, K., 2020. Is bank creditworthiness associated with risk disclosure behavior? Evidence from Islamic and conventional banks in emerging countries. Pacific Basin Finance J. 61, 101327.

Gray, P., Koh, P. S., Tong, Y. H., 2009. Accruals quality, information risk and cost of capital: Evidence from Australia. J. Bus. Financ. Account. 36(1–2), pp 51–72. https://doi.org/10.1111/j.1468-5957.2008.02118.x

Guay, W., Samuels, D., Taylor, D., 2016. Guiding through the fog: financial statement complexity and voluntary disclosure, J. Account. Econ. 62 (2/3), 234–269. https://doi.org/10.1016/j.jacceco.2016.09.001

Guedhami, O., Mishra, D., 2009. Excess control, corporate governance, and implied cost of equity: International evidence, Fin. Rev. 44, 489–524. http://dx.doi.org/10.1111/j.1540-6288.2009.00227.x
Gunning, R., 1952. The Technique of Clear Writing. New York: McGraw-Hill.
Hail, L., Leuz, C., 2006. International differences in the cost of equity capital: Do legal institutions and securities regulation matter? J. Account. Res. 44, 485–531. http://dx.doi.org/10.21315/aamjaf2016.12.S1.2
Halim, S., Soenarno, Y. N., 2018. The Costs and Benefits of Reducing Complex Disclosures, J. Financ. Account. 6 (1), 9–14. DOI: 10.12691/jfa-6-1-2
Hassan, M. K., Abu Abbas, B., Garas, S. N., 2018. Readability, governance, and performance: a test of the obfuscation hypothesis in Qatari listed firms, Corp. Governance: Int. J Bus. Soc. 19(6). http://dx.doi.org/10.1108/CG-05-2018-0182
Hassan, M. K., 2009. UAE corporation-specific characteristics and level of risk disclosure, Manag. Audit. J. 24 (7), 668–687. https://doi.org/10.1108/02686900910975378
Hastie, T., Taylor, J., Tibshirani, R., Walther, G., 2007. Forward stagewise regression and the monotone lasso, Electron. J. Stat. 1, 1–29. http://dx.doi.org/10.1214/07-EJS004
Hastie, T., Tibshirani, R., Friedman, J. H., 2009. The Elements of Statistical Learning (2nd Ed.). New York: Springer, 337–384.
Hothorn, T., Zeileis, A., 2020. party kit: A Toolkit for Recursive Partitioning. R package version 1.2-13, URL http://CRAN.R-project.org/package=partykit
Jensen, M. C., Meckling, W. H., 1976. Theory of the firm: Managerial behavior, agency costs and ownership structure, J. Financ. Econ. 3, 305–360. https://doi.org/10.1016/0304-405X (76)90026-X.
Joliffe, I. T., Trendafilov, N. T., Uddin, M., 2003. A modified principal component technique based on the lasso, J. Comput. Graph. Stat. 12, 531–547. https://doi.org/10.1198/1061860032148.
Kepir, R., 2018. The influence of readability on cost of equity. Erasmus University Rotterdam, Erasmus School of Economics, Master's thesis in Accounting and Auditing.
Kothari, S. P., Li, X., Short, J., 2009. The effect of disclosures by management, analysts, and financial press on cost of capital, return volatility, and analyst forecasts: A study using content analysis, Account. Rev. 84(5), 1639–1670. http://www.jstor.org/stable/27784235
Kuhn, M., Johnson, K., 2013. Applied predictive modelling. New York: Springer.
Kutner, M. H., Nachtsheim, C., Neter, J., Li, W., 2005. Applied linear statistical models. 5th ed. McGraw Hill Erwin.
Lawrence, A., 2013. Individual investors and financial disclosure, J. Account. Econ. 56, 130–147.
Lee, Y. J., 2012. The effect of quarterly report readability on information efficiency of stock prices, Cont. Account. Res. 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, Account. Rev. 86 (3), 1087–1115.
Li, F., 2010. Managers' self-serving attribution bias and corporate financial policies, Working paper, University of Michigan.
Li, F., 2008. Annual report readability, current earnings, and earnings persistence, J. Account. Econ. 45 (2/3), 221–247.
Lim, E. K., Chalmers, K, Hanlon, D., 2018. The influence of business strategy on annual report readability, J. Account. Public. Policy. 37, 65–81.
Livingston, M., Zhou, L., 2010. Split bond ratings and information opacity premiums, Financ. Manag. 39 (2), 515–532.
Lo, K., Ramos, F., Rogo, R., 2017. Earnings management and annual report readability, J. Account. Econ. 63 (1), 1–25.
Loughran, T., McDonald, B., 2011. When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks, J. Financ. 66, 35–65.
Loughran, T., McDonald, B., 2014. Measuring readability in financial disclosures, J. Financ. 69(4), 1643–1671.
Lundholm, R., van Winkle, M., 2006. Motives for disclosure and non-disclosure: a framework and review of the evidence, Account. Bus. Res, International Accounting Policy Forum, 43–8.
Luo, J., Li, X., Chen, H., 2018. Annual report readability and corporate agency costs, China. J. Account. Res. 11 (3), 187–212.
Martins, M. J. T., 2014. Disclosure Index and the Cost of Debt on Portugal, Ireland, Greece, and Spain: An exploratory research regarding IFRS 7 disclosures, cost of debt and its impact on 2011 and 2012. Master's thesis in Accounting. Available at: (https://repositorio.iscte-iul.pt/bitstream/10071/10291/1/Dissertation _Martins_Disclosure_Index_IFRS7.pdf). (Access March 2021).
Moumen, N., Othman, H. B., Hussainey, K., 2015. The value relevance of risk disclosure in annual reports: Evidence from MENA emerging markets, Res. Int. Bus. Financ. 34, 177–204.

http://dx.doi.org/10.1016/j.ribaf.2015.02.004
Nazari, J. A., Hrazdil, K., Mahmoudian, F., 2017. Assessing social and environmental performance through narrative complexity in CSR reports, J. Contemp. Account. Econ 13 (2), 105–178. doi: http://dx.doi.org/10.1016/j.jcae.2017.05.002
Ohlson, J. A., Juettner-Nauroth, B. E., 2005. Expected EPS and EPS growth as determinants of value, Rev. Account. Stud. 10 (2–3), 349–365.
Patelli, L., Prencipe, A., 2007. The relationship between voluntary disclosure and independent directors in the presence of a dominant shareholder, Eur. Account. Rev. 16 (1) 5–33.
Penrose, J. M., 2008. Annual report graphic use: A review of the literature, J. Bus. Commun., 45(2), 158–180.
Petrova, E., Sotiropoulos, I., Vasileiou, K, Z., 2012. Relationship between cost of equity capital and voluntary corporate disclosures, Int. J. Econ. Financ. 4 (3), 83–96.
Rjiba, H., Saadi, S., Boubaker, S., Ding, X. S., 2021. Annual report readability and the cost of equity capital. J. Corp. Finance. 67, 101902.
Sattari, S., Pitt, L. F., Caruana, A., 2011. How readable are mission statements? An exploratory study, Corp. Comm. Int. J. 16 (4), 282–292.
Schwarz, G. E., 1978. Estimating the dimension of a model, Ann. Stat. 6, 461–464.
Shrivastava, S., Jeyanthi, M., Singh, S., 2020. Failure prediction of Indian Banks using SMOTE, Lasso regression, bagging and boosting, Cogent. Econ. Financ. 8 (1), 1729569.
Smith, M., Jamil, A., Johari, Y. C., Ahmad, S. S., 2006. The chairman's statement in Malaysian companies: a test of obfuscation hypothesis, Asian Rev. Account. 14 (1/2), 49–65.
Strobl, C., Wickelmaier, F., Zeileis, A., 2011. Accounting for Individual Differences in Bradley-Terry Models by Means of Recursive Partitioning, J. Educ. Behav. Stat. 36(2), 135–153.
Tetlock, P. C., 2007. Giving Content to investor sentiment: The role of media in the stock market, J. Financ. 62(3), 1139–1167.
Tibshirani, R., 1996. Regression shrinkage and selection via the lasso, J. R. Stat. Soc. Series B 58, 267–288.
Tran, L. T. H. 2022. Reporting quality and financial leverage: Are qualitative characteristics or earnings quality more important? Evidence from an emerging bank-based economy. *Research in International Business and Finance*, *60*, 101578.
Uyar, A., 2011. Firm characteristics and voluntary disclosure of graphs in annual reports of Turkish listed companies, Afr. J. Bus. Manag. 5(17), 7651–7657.
Vena, L., Sciascia, S., Cortesi, A., 2019. Integrated reporting and cost of capital: The moderating role of cultural dimensions. J. Int. Financ. Manag. Account. 31(2), 191–214. https://doi.org/10.1111/jifm.12113
Wu, T., Lange, K., 2008. Coordinate descent procedures for lasso penalized regression, Ann. Appl. Stat. 2(1), 224–244.
Xu, H., Pham, T. H., Dao, M., 2020. Annual report readability and trade credit," Rev. Account. Financ. 19 (3), 363–385.
Zéghal, D., Chtourou, S., Sellami, Y. M., 2011. An analysis of the effect of mandatory adoption of IAS/IFRS on earnings management, J. Int. Account. Audit. Tax. 20, 61–72.
Zeileis, A., Hothorn, T., Hornik, K., 2008. Model-Based Recursive Partitioning, J. Comput. Graph. Stat. 17(2), pp. 492–514.

Tables

Table 1: Frequency Distribution of the Sample of Banks by Country and Year

Year	08	09	10	11	12	13	14	15	16	17	18	19	Sum
BAH	3	3	3	3	3	3	3	3	3	3	3	3	36
EG	9	9	9	9	9	9	9	9	9	9	9	9	108
JO	12	12	12	12	12	12	12	12	12	12	12	12	144
KU	6	6	6	6	6	6	6	6	6	6	6	6	72
OM	7	7	7	7	7	7	7	7	7	7	7	7	84
QA	5	5	5	5	5	5	5	5	5	5	5	5	60
SA	7	7	7	7	7	7	7	7	7	7	7	7	84
UAE	11	11	11	11	11	11	11	11	11	11	11	11	132
Totals	60	60	60	60	60	60	60	60	60	60	60	60	720

(*) BAH: Bahrain. EG: Egypt. JO: Jordan. KU: Kuwait. OM: Oman.QA: Qatar.SA: Saudi Arabia. UAE: United Arab Emirates.

Table 2: Definition of the Variables

Variable	Code	Measure
Response variable		
Cost of capital	*CCOST*	earning per share in period . price of the share in period . The study employs the price earnings growth (PEG) ratio model (Easton, 2004).
Feature variables		
Fog measure	*FOGME*	Fog index (Gunning, 1952). 0.4 * (ASL + 100 * (Nwmin3sy / Nw)) where average sentence length (ASL) = the number of words/number of sentences; Nwmin3sy = the number of words with three syllables or more; and Nw = the number of words.
Flesch measure	*FLEME*	Flesch reading ease score (Flesch, 1948, p. 309) is computed as follows: Flesch Score = 206.835–(1.015 X Average Words per Sentence)–(84.6 X Average Syllables per Word). where average words per sentence is measured by dividing the total number of words by the total sentences, and "average syllables per word is calculated by dividing the total number of syllables by the total number of words (Bayerlein and Davidson, 2011.
Disclosure quantity	*LSENN*	Logarithm of the number of sentences
Partition variables		
Return on equity	*ROE*	Net income/shareholder equity
Leverage	*LEVRA*	Total debts/total assets
Firm size	*LFIRM*	Logarithm of total assets
Loans	*LLOAN*	Logarithm of total loans
Beta coefficient	*BETAC*	A measure of corporate stock's market volatility
Market to book ratio	*MBOOK*	The ratio of corporate market value to book value
Age of bank	*AGEOB*	The number of years from the date of establishment of a company
Years	*YEARS*	The data were collected for 2008 through 2019. We control for year effects by using year dummies.
Country	*COUNT*	The data were collected from eight countries (namely, Egypt (EG), Jordan (JO), Bahrain (BAH), United Arab Emirates (UAE), Saudi Arabia (SA), Kuwait (KU), Qatar (QA), and Oman (OM)). We control for year effects by including year dummies in our model.

Table 3: Descriptive Statistics and Shape Measures (Skewness and Kurtosis) of the Variables

Variables	First quartile	Mean	Standard deviation	Median	Third quartile	Skewness	Kurtosis
CCOST	0.078	0.140	0.092	0.120	0.182	1.603	4.659
ROE	8.706	13.382	11.171	13.464	17.116	-3.250	44.988
LEVRA	6.548	8.500	8.212	7.827	9.206	13.598	198.544
LFIRM	3.409	4.079	0.797	4.050	4.722	0.315	-0.532
LLOAN	3.104	3.800	0.829	3.824	4.507	0.020	-0.835
BETAC	0.449	0.698	0.334	0.741	0.938	-0.026	-0.246
MBOOK	0.897	1.287	0.640	1.154	1.474	2.077	7.433
AGEOB	32	37.772	13.439	38.000	45	-0.304	0.176
FOGME	23.61	25.857	9.867	25.123	26.62	13.274	190.374
FLEME	13.78	16.561	23.550	18.019	21.82	-12.406	176.752
LSENN	2.825	2.948	0.290	2.996	3.134	-1.341	3.677

(*) Standard deviation; First quartile; Third quartile; Fog index: *FOGME*; Flesch measure: *FLEME*; Disclosure quantity: *LSENN*; Return on equity: *ROE*; Leverage: *LEVRA*; Firm size: *LFIRM*; Loans: *LLOAN*; Beta coefficient: *BETAC*; Market to book ratio: *MBOOK*; Age of bank: *AGEOB*.

Table 4: LASSO-Estimated Coefficients for the *CCOST* Model

Feature	Coefficient	Feature	Coefficient
Constant		Country QA	0.001
FOGME	.	Country SA	0.024
FLEME	-0.001	Country UAE	0.026
LSENN	-0.034	Year 09	-0.02
ROE	-0.001	Year 10	-0.017
LEVRA	0.001	Year 11	-0.034
LFIRM	-0.029	Year 12	-0.029
LLOAN	0.014	Year 13	-0.011
BETAC	-0.008	Year 14	-0.008
MBOOK	-0.012	Year 15	0.016
AGEOB	0.001	Year 16	0.005
Country EG	0.067	Year 17	0.010
Country JO	0.0001	Year 18	-0.001
Country KU	-0.016	Year 19	0.049
Country OM	0.021		

(*) Fog index: *FOGME*; Flesch measure: *FLEME*; Disclosure quantity: *LSENN*; Return on equity: *ROE*; Leverage: *LEVRA*; Firm size: *LFIRM*; Loans: *LLOAN*; Beta coefficient: *BETAC*; Market to book ratio: *MBOOK*; Age of bank: *AGEOB*; Egypt: EG; Jordan: JO; Kuwait: KU; Oman: OM.

Table 5: Estimated Coefficients for the Model-Based Recursive Partition Tree (the model-based (MOB) recursive partitioning)

		Regressor				Partition variable								
D	Co	Constant	*FLEME*	*LSENN*		*ROE*	*LEVRA*	*LFIRM*	*LLOAN*	*BETAC*	*MBOOK*	*AGEOB*	YEARS	COUNT
1		0.286	-.001	-0.05		24.66	19.97	5.69	8.97	7.66	14.63	9.08	49.1	50.3
		<0.01	0.84	<0.01		0.01	0.04	1.00	0.98	1.00	0.35	0.98	0.27	0.00
2		0.344	0.002	-0.07		10.55	10.26	13.87	11.86	12.25	7.63	5.09	35.6	5.11
		0.03	0.04	0.15		0.51	0.56	0.17	0.35	0.31	0.90	1.00	0.97	1
3		0.265	-.001	-0.04		21.87	22.03	9.46	7.53	8.32	9.05	13.5	11.2	40.8
		<0.01	0.51	<0.01		0.02	0.02	0.96	1.00	0.99	0.98	0.48	0.81	0.80
4		0.283	0.001	-0.05		2.57	17.42	9.52	3.85	3.13	6.20	4.37	22.4	19.5
		<0.01	0.36	0.13		1.00	0.07	0.82	1.00	1.00	1.00	1.00	1.00	0.85
5		0.213	-.001	-0.03		13.73	5.80	4.87	4.13	10.37	12.80	11.8	28.1	20.8
		<0.01	0.19	0.04		0.46	1.00	1.00	1.00	0.89	0.59	0.72	1.00	0.95

(*) Stand for node: D; Coefficient: Co; P-value: p; Flesch measure: *FLEME*; Disclosure quantity: *LSENN*; Return on equity: *ROE*; Leverage: *LEVRA*; Firm size: *LFIRM*; Loans: *LLOAN*; Beta coefficient: *BETAC*; Market to book ratio: *MBOOK*; Age of bank: *AGEOB*; Years: *YEARS*; Country: *COUNT*.