Does soft information determine credit risk? Text-based evidence from European banks*

Journal of International Financial Markets, Institutions & Money, 75, November 2021, 101303

https://doi.org/10.1016/j.intfin.2021.101303

Albert Acheampong

University of Bradford, UK

a.k.acheampong@bradford.ac.uk

Tamer Elshandidy**

Ajman University, UAE

Helwan University, Egypt

t.elshandidy@ajman.ac.ae

*We thank Jonathan A. Batten (the Editor) and the anonymous referee for constructive and helpful comments. This paper has benefited from comments and suggestions from participants at Bradford University Management School Conference (June 2019). We thank Mahdi Mousavi, Chengang Wang, and Steven Wu for their helpful suggestions.

**Correspondence should be addressed to Tamer Elshandidy, Ajman University, College of Business Administration, P. O. Box 346, Ajman, UAE. Email: t.elshandidy@ajman.ac.ae

Cite this article as:

Acheampong, A., Elshandidy, T. (2021). Does soft information determine credit risk? Text-based evidence from European banks. ***Journal of International Financial Markets, Institutions & Money, 75, 101303***

Does soft information determine credit risk? Text-based evidence from European banks

Abstract

This paper uses a supervised machine learning algorithm to extract relevant (soft) information from annual reports and examines whether such information determines credit risk (as measured by non-performing loans, Ohlson's O-score, Altman's Z-score, and credit rating downgrades). The paper also assesses how far both bank- and country-level characteristics influence variations in credit risks both within and between banks across 19 European countries between 2005 and 2017. Based on 1885 firm-year observations, we find that the text-based credit risk (soft) measure explains a substantial portion of the variation in NPLs, O-score, Z-score, and credit rating downgrades. We also find that bank-level characteristics and country-level characteristics are highly important for explaining variations in non-performing loans, O-score, and credit rating downgrades, as compared to Z-score. Overall, our results have implications for firms, regulators, and market participants who are seeking evidence on the credibility of annual reports in conveying relevant information that reflects actual credit risk.

JEL Classification: G21; G28; G32

Keywords: Credit risk, Machine learning, Repeated measures multilevel analysis, Corporate disclosure

1. Introduction

The current evidence of Bonsall et al. (2017) and Donovan et al. (2019) suggests that credit risk presents a signal to market participants by which they may assess the competitive position of a firm and its management capabilities. The European Banking Federation reports that credit risk, as measured by non-performing loans (NPLs) is a problem of the past due to the trajectories of NPLs of European Union (EU) banks reporting massive declines across communities, evidenced by the fall in NPLs over the years. For instance, in 2017 the NPLs of the EU (i.e., 3.7%) fell below the world average of 3.74%.[1] Prior research focuses largely on quantitative rather than qualitative (soft) information for estimating credit risk measures, using numbers from financial statements or historical stock price movements (Ali and Daly, 2010; Altman, 1968; Cucinelli et al., 2018; Ghosh, 2015; Ohlson, 1980). These credit risk measures are numerous and include multiple discriminant models, probability of default, logit models, and credit ratings. The heavy reliance on quantitative information over the years for estimating credit risk is perhaps because of the difficulties in capturing and quantifying qualitative (soft) information from corporate disclosures.

Together, prior research on textual analysis provides reliable measures for several features, such as readability (e.g., Li, 2008; Loughran and McDonald, 2014), tone (e.g., Feldman et al., 2010; Loughran and McDonald, 2011; Price et al., 2012), capacity for forward looking (e.g., Li, 2010; Muslu et al., 2015), and firm risk (e.g., Campbell et al., 2014; Elshandidy et al., 2013; Elshandidy et al., 2015). Another strand of the literature shows the application of supervised machine learning algorithms to quantify narrative sections of corporate reports (conference calls) that determine the credit risk (e.g., Donovan et al., 2019) and earnings surprise (e.g., Frankel et al., 2018) of firms. Hitherto, very little research outside the US context has combined quantitative with qualitative information in banks.

Our paper addresses this gap by employing supervised machine learning algorithms to model language in the narrative sections of annual reports of a large-scale sample of European banks in order to explain the observed variations in these banks' credit risks. Our paper complements that of Donovan et al. (2019), which focuses on conference calls from the USA context and employs credit default swap

[1] For more information see: https://eba.europa.eu/risk-analysis-and-data/risk-assessment-reports

spread as the credit risk proxy. They conclude that their text-based credit-risk measure explains a substantial portion of borrowers' credit risk. Our paper, however, differs significantly from Donovan et al. (2019) in terms of the approach and measures. For example, our paper employs data from the EU context, which institutionally yields different results from those that adopt the US regulatory setup which is linked with the Securities and Exchange Commission (SEC). Our paper also employs NPLs to measure credit risk because it has been widely used as a measure of credit risk (Zhang et al., 2016), and is linked to bank failure, presenting itself as a forerunner of bank crises (Ghosh, 2015). Furthermore, our paper models the variations in credit risk that can be explained by our text-based credit risk measure at two different levels, namely, the bank and country levels.

Therefore, we manually collected the annual reports of banks from 19 European countries for the period 2005 to 2017 to explore the usefulness of credit-risk information (as shown in Section 4). Our results, based on 1885 firm-year observations, show that the text-based credit risk measure is significant and exhibits positive association with NPLs, Ohlson's O-score, Altman's Z-score, and credit rating downgrades. This suggests that a significant variation in all credit risk measures across countries within and between banks over the chosen period can be explained by our proposed measure of soft information for credit risk. At country level, we find that financial stress over time have significantly higher explanatory power over all credit risk measures variations. We also find that corruption level explains significant variations in non-performing loans and credit rating downgrades. These results suggest that our text-based credit risk measure contains an economically significant approximation of actual credit risk, and, more importantly, it explains credit risk better than existing credit risk measures. Overall, our evidence supports the argument that banks' annual reports contain information which relates to credit risk and reflects a bank's risk exposure. Our findings are robust to a variety of sensitivity tests and alternative credit risk measures.

Our paper contributes to the finance and accounting literature in the following ways. First, we employ corporate disclosure from the EU context. Prior research (e.g., Donovan et al., 2019; Frankel et al., 2018) on the application of machine learning algorithms focuses largely on specific corporate disclosures from the US context (e.g., 10-Ks, MD&A, analyst reports, and conference calls). To our knowledge, our paper is the first to provide direct evidence on the value relevance of annual reports from the EU context on the application of supervised machine learning algorithms. Second, our paper provides guidance on determining credit risk by finding ways to map NPLs to the narrative sections of banks' annual reports with the application of machine learning algorithms. Our paper is the first to employ a machine learning algorithm which converts the soft content of corporate disclosures into hard content capable of explaining credit risk in a large-scale sample of EU banks. The previous literature focuses principally on quantitative measures to estimate credit risk (e.g., Ali and Daly, 2010; Cucinelli et al., 2018; Ghosh, 2015) However, to our knowledge, no studies have so far mapped language in annual reports to the NPLs of European banks.

Third, unlike earlier literature, we observe variations in credit risks over time, and then associate such variations with changes in text-based credit risk measure. Our paper models the language (soft) in banks' annual reports over a thirteen-year period to identify credit-risk relevant information more robustly than prior research, since we use Repeated Measures Multilevel Analysis (RMMA) to mitigate the problems caused by nested effects and to account for the variances at various levels over time that are normally ignored in traditional Ordinary Least Squares (OLS) (Robinson, 2003). Finally, the text-based (soft) credit risk measure explains an economically significant portion of the variations in NPLs, and more importantly, it indicates other credit risk measures. Given our results, we suggest that annual reports are useful for identifying the credit risk exposure of firms, thus complementing the work of prior writers (e.g., Campbell et al., 2014; Donovan et al., 2019). Our evidence, as a result, confirms the credibility of the narrative sections of corporate reports, which strengthens the reputation of existing credit risk measures.

Our paper provides important theoretical implications. First, our paper shows the importance of interacting between various levels in cross-country studies where data is nested in hierarchical structures. This answers Elshandidy et al.'s (2018a) call for more research papers to cover the financial firms in a cross-country setting utilizing computer-based assistance in capturing risk information. It further accommodates their suggestion that future research in cross-country settings should adopt multilevel techniques that can capture the hierarchical structure of cross-country data. Our evidence shows to cross-country banking research that it is essential to interact between country- and bank-level variables to understand the variations in credit risk and that this necessitates more future research to develop theoretical frameworks to better understand interactions between these different levels. Furthermore, our results suggest that more attention should be given to variables that may explain the variations in credit risk for banks over time. Our suggestion is consistent with the recent trend (Barth et al., 2013; Beck et al., 2003; Chan and Mohd, 2016) in the literature which explores widely varying governance indicators on bank characteristics (e.g., efficiency, loan quality). Finally, other related research on textual analysis can use our proposed method in quantifying the narrative sections of annual reports (or any similar corporate outlets) with regard to other attributes (e.g., operational risk disclosure). Specifically, our method adds significantly to machine learning algorithms in the financial literature by endorsing the current importance of widening this research scope to give more attention to the application of such algorithms in exploring various firm fundamentals. This is, of particular importance in view of the increased concern regarding whether accounting disclosure benefits market participants or not.

Our paper provides important practical implications to managers, regulators, and investors by contributing to the ongoing discussion on whether banks' annual reports have informative information content. Managers who know that supervised machine learning algorithms can be applied to their corporate disclosures to provide deeper insights may position themselves effectively to disclose more relevant information, thereby addressing the issue of information asymmetry. Regulators may encourage banking institutions to continue to disclose vital credit risk information. This becomes increasingly important, considering the great reliance of global economies on understanding the complexity and grey areas in the published reports of large firms. Investors are likely to reshape their trading strategies and, if they can, employ supervised algorithms to have more pointed information about firms. In theoretical terms, the high significance of corruption level and financial stress in explaining variations in credit risk across countries implies that any attempt to identify the factors that drive credit risk should take these two country factors into consideration.

The paper proceeds as follows: The next section discusses the institutional background in Europe. Section 3 reviews relevant literature and develops research hypotheses. Section 4 describes the sample selection, data collection, and empirical models. Section 5 presents the empirical results. Section 6 introduces further analysis and robustness checks of our results and finally, section 7 draws conclusions and suggests avenues for future research.

2. Institutional background

The European Banking Industry (EBI), over the years, has undergone rigorous rounds of reform through forces such as European integration, technology, and deregulation, to improve the soundness of banking practices and mitigate overall risk exposure (Goddard et al., 2007). Regarding risk, in the records of the EBI, credit risk continues to be ranked as the main risk for most banks. In order to enhance the stability and overall credit risk exposure of the industry, the EU over the years has worked together with global standard setters, supervisors, and the banking industry to ensure that banks are well capitalised, complemented by high liquidity levels, low risk balance sheets and stronger governance, as evidenced by the significant drop in NPLs. In addition, the EU has stiffened the requirements regarding transparency and centralised some activities to mitigate the effects of interconnection between banks, together with exercising strong supervision. This gave birth to the Single Supervisory Mechanism in the Euro area and the creation of three EU supervisory authorities for the banking insurance, pensions, and securities sector. To sustain these changes, the EU implemented some important banking reforms which were intended to alleviate any adverse impacts of bank failure. These reforms have been used to restore failing banks. They include the implementation of the Single Resolution Mechanism consisting of a Single Resolution Board, national resolution authorities and the inclusion of the banking union in the central resolution authority.

The genesis of these efforts can be directly traced to the regulatory bodies. For example, the Basel Committee's capital framework of Basel II has over the years introduced the concept of credit risk calibration, dependent on banks' internal risk models. This has allowed banks to estimate the probability of their own default, loss given default and exposure at default. The Basel Committee in 2006 issued guidance with regard to loan classification, which compelled banks to have a constructive credit classification system on the basis of credit risk.[2] This was geared to improving the stability of the banking system and bridging the harmonization gap created by the global financial crisis of 2008 (Bitar et al., 2018). Over the years, several standards have emerged, one of which is the often-invoked standard of the local Generally Accepted Accounting Principles (GAAP) in the area of accounting harmonization and integration of the financial market (Bitar et al., 2018).

The EU's massive step on 1 January 2005 is noticed as advancing publicly listed firms towards compliance with the International Accounting Financial Standards (IFRS). The adoption of the IFRS is one of the most substantial amendments in the financial reporting history of firms in Europe and has conferred benefits that include greater inter-company comparability, easier cross-border transactions, increased transparency, strong asset prices, statements of financial position that enable companies and lenders in Europe to conclude contracts more efficiently, and so forth (e.g., Agostino et al., 2011). However, it has raised matters of concern; specifically, the potential benefits from IFRS seem to be biased according to firm size. Firms with large market caps are seen to benefit more than others from adopting the IFRS (e.g., Horton et al., 2013).

The main aim of corporate reporting in Europe is to improve the quality of financial statements and to secure investors' interests globally (Fiordelisi et al., 2011). Since 2005, an increasing number of disclosure requirements and rules have been defined in Europe. The requirements for corporate risk disclosure have over the years become more detailed. For example, IFRS 7 requires firms to disclose both qualitative and quantitative information concerning three main types of risk exposures: credit risk, market risk, and liquidity risk. The information should detail to what extent firms are exposed to such risks, how the risks arise, and the measures taken to manage such risks. In 2008, the International Accounting Standards Board (IASB) revised the requirements of IFRS 7 to include the reclassification of

[2] In December 2014, another consultative paper was published regarding revisions to the credit risk standardised approach. For the first time, the Basel Commission on Banking Supervision (BCBS) suggested a definition of non-performing loans, whose threshold includes 90 days past due loans, and 30 days past due securities. The basis for this definition was to calculate the non-performing asset (NPA) ratio. This consultative paper was described by the Basel Committee "as at an early stage of development".

non-derivative financial assets and specified effective and transition dates of such requirements. In 2009, the IASB further revised the standard to require fuller disclosure about fair value measurements and liquidity risk that would give users more useful information. These regulations and their revisions in general reflect the intention of EU market regulators to strengthen the relevance and effectiveness of information disclosed by publicly listed firms (BCBS, 2015; Elamer et al., 2020).

To conclude, European banks are exposed to multiple regulatory requirements that pay great attention towards banks' transparency generally and risk disclosures particularly. Academic literature therefore investigates several aspects within the banking sector, as detailed in the following section.

3. Literature review and hypothesis development

3.1 Relevant literature on banks' credit risk and textual credit risk

In their recent review, Elshandidy et al. (2018a) survey 32 papers on risk disclosure that are synthesised based on their primary focus into those papers (16) interested in incentives for risk disclosure (what motivates firms to reveal risk information), papers (12) interested in the informativeness of risk information (whether risk information is value relevant or not), and papers (4) that are interested in studying both incentives and informativeness. After that they identify four areas of divergence among all reviewed papers, including type of disclosure (mandatory, voluntary, and aggregate), type of content analysis (manual and automated), type of analysis (within-country and cross-country), and type of sector (financial and non-financial). As our paper is interested in answering the question of whether textual credit risk is useful in identifying credit risk, the second theme of Elshandidy et al.'s (2018) reviewed papers is the most relevant to us. Among those 12 papers, there are just four papers that looked at financial firms, suggesting that risk disclosure research in the banking sector has not been widely explored compared to the non-financial sector (Maffei et al., 2014). However, these four papers neither study cross-country settings (mostly their context is the USA) nor apply automated content analysis in capturing risk (soft) information (the most used method is manual content analysis).

The above-mentioned empirical papers that investigate the incentives for and/or the informativeness of risk disclosure in the banking sector commonly analyse relevant regulatory requirements since such disclosure is subject to multiple regulations (as discussed in detail in Section 2). As a result, standard setters, regulators, practitioners, and academics over the years have shown concern about the quality and quantity of risk disclosure. There has, therefore, been call for higher transparency and quality in risk disclosure in the banking sector, as banks are very different from non-financial firms in that they follow distinctive accounting practices and regulations (Elshandidy et al., 2015), and also exhibit higher risk-taking behaviour that has been widely noted, especially during crisis periods, to be unstable and unprecedented (Erkens et al., 2012; Magnan and Markarian, 2011). Risk should therefore be managed constructively and disclosed in a way that reveals relevant information to investors and other stakeholders (Linsley and Shrives, 2005).

One of the main signs of stringent regulatory standards is increased corporate disclosure of firms' content, which stems from the motive of regulators to improve the effectiveness of textual (soft) information. Both hard and soft information in corporate disclosures are used by market participants to assess the risk position of firms. Within a firm's disclosure are various characteristics that can be extracted for assessing their credit risk (Bonsall et al., 2017). The overwhelming increase in the volume of soft information created a new avenue for researchers in the past to investigate the value-relevance of such information using machine-codes. Over the last two decades a significant development in the field of textual analysis has provided researchers with powerful tools to explore the in-depth meaning of corporate disclosures including such firm characteristics as readability (e.g., Li, 2008; Loughran and McDonald, 2014), tone (e.g., Feldman et al., 2010; Loughran and McDonald, 2011; Price et al., 2012), forward looking information (e.g., Li, 2010; Muslu et al., 2015), and firm risk (Campbell et al., 2014).

In analysing textual risk disclosure, prior studies employed various automated text mining techniques. For instance, Campbell et al. (2014) employed a predefined dictionary approach to classify risk factors into five types. In contrast to the predefined keywords approach in analysing textual risk disclosures, other studies employed the Latent Dirichlet Allocation (LDA) topic model to identify and quantify textual risk disclosure in corporate disclosures. LDA is an unsupervised Bayesian machine learning approach developed by Blei et al. (2003), which generates topics from a corpus of documents ideally by constructing features for classification that reduce data dimensionality (Blei and McAuliffe, 2007). Bao and Datta (2014) employed the LDA topic model to identify and quantify 30 types of risk disclosure in the narrative section of 10-Ks to investigate the impact of each type of identified risk on the volatility of the stock returns. Their work has been the most comprehensive classification of risk disclosure with 10-Ks (Elshandidy et al., 2018a).

Unlike the unsupervised topic model, other studies employed supervised machine algorithms to analyse textual risk disclosures. Huang and Li (2011) proposed a supervised text mining technique called multi-label categorical K-nearest neighbour (ML-CKNN) to classify twenty-five risk factors. However, the application of a dictionary-based technique and ML-CKNN requires a pre-defined list of risk factors. Typically, risk factors are likely to be difficult to pre-define (Wei et al., 2019). One machine learning algorithm that addresses this issue is the supervised Latent Dirichlet Allocation (sLDA) topic model. sLDA is useful for prediction because it parses text documents by fitting a response variable (Blei and McAuliffe, 2007). sLDA adds an additional layer to unsupervised LDA by selecting topics highly correlated with the response variable. Donovan et al. (2019) employed sLDA to extract topics in conference calls to explain and predict credit risk as measured by credit default swaps (CDS) spread. Thus, the sLDA outperforms the dictionary and unsupervised text mining techniques. Hence, in this current paper, we adopt supervised LDA to model the narrative section (soft) of annual reports, which is useful in explaining the credit risk of EU banks, as discussed in the following sub-sections.

3.2 Hypothesis development

3.2.1 Textual credit risk information and credit risk measures

Credit risk is evident from a firm's inability to cover future financial commitments, as subsequently revealed by the large increase in NPLs in firms' published financial statements. To mitigate this problem and ensure the efficient management of credit risk, firms should comply with existing standards and supervisory regulations (Butaru et al., 2016). These regulations require banks to provide information to investors in their corporate reports, which serves as a basis for estimating future share prices, related cash flows, and future credit risk (e.g., Elshandidy, 2011). Risk disclosure is therefore an important aspect of credit risk management (Allini et al., 2016; Jones et al., 2018) as it helps market participants to make comparison of their anticipated returns with associated risks, thus increasing the satisfactoriness of their decisions regarding portfolio investment (e.g., Abdullah et al., 2015; Linsley and Shrives, 2005). However, corporate disclosure may present difficulties to investors and may be of less significance because the qualitative information in corporate disclosure far outweighs that in the quantitative section, which could make evaluation and assessment difficult (e.g., Gomes et al., 2007; Linsley and Shrives, 2005).

Most of the literature (e.g., Elshandidy and Shrives, 2016; Jones et al., 2018; Miihkinen, 2013) that focuses on the economic implications of financial and non-financial information (i.e., aggregate risk, credit risk, financial risk, and operational risk disclosures) relies on the fact that such information provides market participants with useful information to make informed decisions about issuers. Empirical findings, however, are mixed. For example, Miihkinen (2013) finds that quality of mandatory risk disclosure improves market liquidity and thus reduces information asymmetry between market participants, and that this effect is conditional on some firm-specific characteristics such as size and analysts coverage. As a support to the relevance of risk disclosure, Elshandidy and Shrives (2016) reveal in their study that disclosed risk-related relevant information improves market liquidity and reduces investors' perceived risks. Oliveira et al. (2011) report that some of the key drivers for voluntary risk disclosure by banks are corporate reputation and stakeholder monitoring. They also find disclosure on credit risk to be optimal levels of mandatory compliance. However, Maffei et al. (2014) conclude that the risk disclosure is less useful for decision making since such disclosure is generic. This conclusion is consistent with early evidence on risk disclosure, as most studies identify risk disclosure to be more qualitative, non-comparable, not sufficiently enough to be forward-looking, unclear, and not salient for evaluating risk exposure on an on-going basis (Linsley and Shrives, 2005; Oliveira et al., 2011).

In assessing firm credit risk using qualitative information, prior literature develops word-lists that capture credit risk and finds that such soft information of MD&A in US 10-K filings (Mayew et al., 2015) or in entire narrative sections of annual reports as in the UK (Elsayed and Elshandidy, 2020) has predictive power three years or two years prior to actual bankruptcy, respectively. In considering the impact of risk information on credit risk in some European banks, Jones et al. (2018) investigate the usage and the accuracy of graphical credit risk as a response to levels of credit risks for 235 firm-year observations across five countries (Germany, France, Italy, Spain, and UK) from 2006 to 2010. They find evidence that banks use graphics to reveal incremental credit risk information. They further find that such usage, however, is subject to the underlying levels of banks' credit risks. Banks with higher credit risk are likely to use fewer informative graphs, supporting the impression management explanation that managers might bias disclosed information in their favour. By extracting relevant information from conference calls and MD&A of 10-K filings, Donovan et al. (2019) find evidence that extracted soft information has incremental power in capturing and estimating credit risk above the traditional or quantitative credit risk measures such as O-score and Z-score. They find also that the relative relevance of extracted information is dependent on disclosure outlet. Particularly, while information extracted from conference calls has a significant power in explaining variations in credit risk within and between firms, information extracted from MD&A has a power in explaining variations on credit risk across firms rather than within firms.

We argue that banks tend to disclose more credit risk information geared towards meeting the needs of their investors in their quest to respond to regulatory and standard practices. This prompts banks to adopt the approach of others who disclose high levels of risk information, therefore raising their status by sending a signal to their investors about their competence and commitment in managing risks effectively (Elshandidy et al., 2015). In such a setting, banks seeking to be perceived as legitimate are likely to be representational and standardised in their narrative disclosure. This results in formulating the following hypothesis:

Hypothesis 1: European banks' narrative sections of annual reports contain credit risk information which is likely to reflect the actual credit risk.

3.2.2 Textual credit risk information and credit rating changes

Credit rating agencies (CRAs) consult information from corporate disclosures as part of their independent risk assessment (e.g., Altman et al., 2002; Shorter and Seitzinger, 2012). The statistical methodologies developed by CRAs and researchers identify two main sources of information for their risk modelling (Das et al., 2009). These are soft or qualitative information released in corporate reporting and hard or quantitative information such as stock returns and volatilities (Bozanic and Kraft, 2018). Considering both public and private information released by rated firms, Bonsall et al. (2014) find evidence that managers release more bad news to ratings agencies than they reveal publicly to investors. They also find that while CRAs incorporate such information to assess a firm's credit risk, changes in credit ratings are not fully reflected in stock prices. Looking at the readability attributes of narrative sections of 10-K filings, Bonsall and Miller (2017) find that less readable information is linked to less favourable credit rating and more disagreement. Looking at a broader aspect of transparency, DeBoskey and Gillett (2013) study whether public disclosure information together with earnings quality information, intermediary information, and insider information (as a proxy for corporate transparency) provide incremental explanatory power for external credit ratings. They conclude that both the aggregate and individual aspects of corporate transparency have a significant impact on credit ratings and explain an incremental amount of its cross-sectional variations. Other evidence (Kuang and Qin, 2013) suggests that CRAs, in their credit risk assessment, incorporate management risk taking incentives. They conclude that CRAs incorporate forward-looking information in their independent risk assessment of firms.

Recent studies of credit ratings reveal that the determination of creditworthiness of an entity is improved by its qualitative information to the public and such information tends to impact analysts' judgements of credit risk of firms (Bozanic and Kraft, 2018). Specifically, Bozanic and Kraft (2018) find that the narrative section of corporate disclosure contains credit risk relevant information that explains credit rating adjustments. Also rating agencies rely on soft information from corporate disclosure to aid in their independent risk assessment (i.e., credit rating adjustments). Furthermore, they find that higher informativeness for credit rating downgrade is associated with a greater amount of corporate disclosure suggesting that market participants perceive information disclosed by firms to be credible, especially when interpreted in line with analysts' private information. Finally, the informativeness for rating downgrades associated with a greater amount of disclosure declined following the introduction of specific regulations. Considering risk information disclosed in narrative sections of annual reports of banks in 12 emerging countries from 2006 to 2013, Elamer et al. (2020) find positive associations between risk disclosure index (developed based on IFRS 7 and 9 to capture six principal risk areas) and both credit ratings and the changes in those ratings. Their evidence suggests that banks provide relevant risk information that impacts long-term issuer default ratings. These results are consistent with Grassa et al. (2020) who find that credit ratings are significantly and positively associated with banks' risk disclosures, and that the association is more prominent in conventional banks than in Islamic banks. Bozanic et al. (2018) report that, when more credit risk information is disclosed by firms, credit rating analysts tend to make soft adjustments that pertain to increase in credit risk. In addition, they indicate that in instances where managers provide forward-looking and earning-based information, CRAs are more likely to make adjustments that reflect reduction in credit risk.

Thus, there is a general consensus among the above-mentioned empirical papers that corporate disclosures influence credit ratings, suggesting that managers use credit risk disclosure to signal market participants of their credit risk, and most importantly, this information can be incorporated in credit ratings. This leads to the following hypothesis:

Hypothesis 2: The narrative sections of European banks' annual reports contain credit risk relevant information that is likely to be associated with future credit rating downgrades.

3.2.3 The interaction between bank-level and country-level factors in explaining variations on credit risk[3]

Elshandidy et al. (2018a) call for more research papers to cover financial firms in a cross-country setting and utilizing computer-based assistance in capturing risk information. They suggest that future research in cross-country settings should adopt multilevel techniques that can capture the hierarchical structure of cross-country data, so that researchers can incorporate country level factors with those factors at bank-level. Because our data has a nested effect (i.e., different levels of banks across countries over years), in order to explain the expected variations in credit risk, we distinguish between factors at the bank level (with particular attention to textual (soft) credit risk information, as discussed in this section) and factors at the country level such as financial stress (e.g., Cardarelli et al., 2011), corruption (e.g., Park, 2012), governance (e.g., Barth et al., 2013), GDP growth rate (e.g., Louzis et al., 2012), and inflation level (e.g., Ali and Daly, 2010). All these factors are discussed in detail in the following section.

[3] We thank the anonymous referee for suggesting this point to us.

Given our expectation of significant variations in credit risk (i.e., slope variance), the investigation of which country-level characteristics are associated with such variation is of much importance. Our nested data makes interactions among predictor factors, particularly, cross-level interactions, very important in the application of multilevel models (e.g., Elshandidy et al., 2015; Hox, 2010). Cross-level interaction occurs when the impact of lower-level (bank level) variables on the response variable varies based on the changes in higher-level (country level) variables (Bliesse and Hange, 2004).

Financial stress leads to high uncertainty and/or risk, abrupt liquidity drought, and concerns regarding the health of the banking system (Balakrishnan et al., 2011). Countries which are financially stressed exhibit contraction in their economic activities as a result of high procyclicality of leverage in the banking system (Cardarelli et al., 2011). This is especially the case for countries with arm's length financial system. The effect could be seen on increased uncertainty about underlying asset value, which increases irregularities in investor behaviour, evident through high unwillingness to hold risky assets (Hakkio and Keeton, 2009). This, therefore, triggers corruption among high officials, which is then associated with problematic loans and sluggish economic growth (Chen et al., 2015; Park, 2012). Strong economic governance creates a very optimistic economic environment, which boosts banking operations (Chan and Mohd, 2016; Barth et al., 2013). The improvement in the operations of banks ensures bank efficiency because of high institutional quality, which leads to low cost of operations geared toward lower levels of bureaucracy. A fall in GDP growth rate retards the improvement in debt servicing and concurrently increases NPLs. The reverse is true when an economy experiences economic boom (Ali and Daly, 2010; Salas and Saurina, 2002) whereas inflation disrupts the price system of societies through savings cuts, loss of investor confidence, and ideally retards economic growth and puts financial health at risk. A country with high inflation is exposed to high social and economic costs at high rates. It reduces the net income and debt servicing capabilities of borrowers, thus resulting in increased risk of NPLs (Jankowitsch et al., 2007). Put together, a country with high financial stress, corruption level, inflation and with low governance, and GDP will experience slow economic growth which eventually affects the financial health, whereas countries with low financial stress, corruption, inflation and with high governance and GDP growth rate experience high economic growth coupled with strong financial health.

Countries with strong governance create a conducive environment for firms to operate because such countries commit to compliance with laws, financial stability, efficient and effective control of corruption, which provide the required structures for ensuring economic performance (Çule and Fulton, 2013; Price et al., 2011). Bianco et al. (2012) posit that financial stress serves as a tool for investors to intelligently monitor the systemic risk and financial system. Thus, any bad impression from their assessment might raise concerns regarding their investment. A country with high (low) financial stress, corruption, inflation (governance, GDP growth rate) puts its banks at the risk of failure because such effects widen credit spread arising from investor unwillingness to hold additional debt which increases the cost of borrowing to get funding (Bharath et al., 2008). The impact of this is that banks may freeze their operations, when participants notice greater counterparty credit risk (Bonaccors di Patti and Sette, 2016). With greater credit risk, we argue that banks, in turn, might disclose relevant credit risk information in their annual reports in response, to reduce investors' perceived risk and improve market liquidity (Elshandidy and Shrives, 2016), geared towards shareholder monitoring and corporate reputation (Oliverira et al., 2011). In terms of credit risk disclosure, investors will be less concerned if banks operate in countries with a sound economic and governance system coupled with a stable financial environment. In contrast, they would be very concerned when banks operate in countries with very unstable economic and governance structures risk (e.g., Bianco et al.., 2012). We therefore assess whether the association between credit risk and text-based credit risk within countries varies as a function of country level characteristics. We formulate the following hypothesis:

Hypothesis 3: The relationship between text-based credit risk and credit risk is likely to be relatively stronger in countries categorized by high financial stress, corruption, inflation and low governance and GDP growth rate.

4. Data and empirical method

4.1 Sample selection and data sources

The criteria and process for our sample selection are described in Table 1 (Panel A). The sample covers the thirteen-year period from 2005 to 2017. We used 2005 as the starting point to take account of the mandatory adoption of IFRS in Europe to ensure the comparability of accounting standards across countries. We obtained annual reports from Thomson One or, when this was not available, from a bank's website. We focused on annual reports because they remain a major source of information for investors (Elshandidy et al., 2013; Elshandidy et al., 2015). The initial sample contained 2782 observations, but 897 firm year observations were excluded due to lack of data. We ended with a final sample of 1885 observations, representing 145 banks from 19 countries. Table 1 (Panel B) reports the distribution of the sample of banks across countries.

[Insert Table 1]

4.2 Measuring textual credit risk

Our paper adopts supervised LDA to model the narrative section (soft) of annual reports which is useful in explaining the credit risk of EU banks. A key advantage of LDA, compared to other approaches to identifying topics within a corpus, is that no prior knowledge of the topics is required (Appendix C provides a detailed description of LDA).

For pre-processing annual reports, we used the *tm* package in R statistical software. We convert all annual reports from PDF to text files using R, after which we created a corpus for all text files.[4] We converted all texts into lowercase and remove all numbers, punctuation signs, special characters, and stop-words using both the "English" and "smart" stop-words in the *tm* package. Next, we replaced all the words with similar meanings in the remaining corpus by a single word. For example, words such as *regulate, regulates, regulation, regulations, regulated, regulator* were all replaced by one word - *regulate*. Finally, we performed a Document Term Matrix (DTM) on the remaining texts, which helped to identify each word with its frequency in each document. From the DTM, all the words which appear fewer than three times in any annual report were excluded.

For sLDA, we use the *lda* package in R with inferences from the variation expectation-maximization algorithm, following the original process by Blei et al. (2003). We determined the optimal number of topics (k) using the perplexity function in R. For the best performance of topic generation, we estimated the topics for different models ranging from 10 to 200 topics. The harmonic mean reports

[4] It is worth mentioning that from our total sample of 1885 bank-year observations, there were 52 non-English annual reports (4 Norwegian banks over our sample period of 13 years). We translated these 52 non-English reports to English using the translateR package in R, which can be accessed via https://cran.r-project.org/web/packages/translateR/translateR.pdf. We rerun all of our main analyses (discussed in Section 5) after excluding these 52 translated annual reports. Our unreported results (available upon request from the corresponding author) were similar to our main results that include the translated reports.

the optimal number of topics (k) to be 19. To achieve better results, we converted the corpus into vocab, after which we map each term in the vocab to its original document. We then divided our data equally into a training set and a testing set. The training set is a random sample of half the total banks that we used to identify the credit risk estimates by training the narrative section of the annual reports. The testing set represented the other half of the random sample, which we used to assess whether the out-of-sample credit risk estimate reflected the credit risk as measured by the NPLs.

We used sLDA to map the language in the annual reports (vocab) to the NPLs (response variable) to identify the latent topics that best predicted NPLs for out-of-sample documents based on the topic frequencies of the words in the annual reports (Blei and McAuliffe, 2007). Using the response variable for bank *i* in year *t*, the frequency of all terms in the annual reports in the training set was used to identify 19 topics based on the optimal score. The weights of the topics for all annual reports in year *t* were used to estimate the out-of-sample prediction for NPLs, which we labelled NPL_TXT. This was our text-based credit risk measure, which is used as our main independent variable. Figure 1 shows the NPL_TXT predicted for Barclays Bank Plc in comparison with NPLs. This shows a moving trend for both variables in the same direction, except for the years 2007 to 2009, when the two measures moved in opposite directions. This change can be ascribed to the outbreak of financial crises in these years.

[Insert Figure 1]

4.3 Variables - measurement and description

4.3.1 Credit risk measures

We employed credit risk measures widely justified by prior literature: NPLs, Ohlson's (1980) O-score (OSCORE), Altman et al.'s (1998) Z-score (ZSCORE) and S&P credit rating downgrades (DOWN). In theory and practice, NPLs remain among the key credit risk variables and are closely correlated with firm distress and default (Ghosh, 2015). NPLs are highly ranked in ascertaining the credit risk of financial institutions (Zhang et al., 2016). A report by European Banking Authority suggests that NPLs are the main credit risk faced by European banks. A sample of NPLs for selected countries in our sample is shown in Figure 2, which reveals that over our chosen period, the EU recorded high levels of problematic loans. The NPL ratio was stable between 2005 and 2007 and experienced an increasing trend from mid-2007. The increasing trend continued (e.g., Croatia, Greece, Italy, and Portugal) until 2015 and then began to stabilise. Ireland, in contrast, saw an improvement in its NPL portfolio from 2013 after the global financial crisis of 2008. Several studies employ the NPL ratio as a credit risk proxy (e.g., Berger and DeYoung, 1997; Cucinelli et al., 2018; Fiordelisi et al., 2011; Ghosh, 2015; Louzis et al., 2012; Williams, 2004). We followed prior literature and measured the NPLs as the total loan default divided by total assets. It is measured annually. Accordingly, we used NPLs as our main credit risk proxy.[5]

[Insert Figure 2]

We used OSCORE and ZSCORE to analyse the explanatory power of our text-based credit risk measure beyond NPLs. We measured OSCORE using Ohlson's (1980) model, whereas ZSCORE was measured according to Altman et al. (1998), focusing on manufacturing firms and therefore applicable to banks.[6] DOWN is a dummy variable which takes the value of 1 when a bank is downgraded by S&P in year *t* and 0 otherwise. Donovan et al. (2019) employ S&P rating downgrades to investigate its relationship with text-based CDS spread and find that text-based CDS explains much more significant information in relation to downgraded ratings from S&P.

4.3.2 Bank-level variables

Based on the literature, we employed the following bank-level variables: leverage (LEV), diversification income (DIV_INC), bank size (SIZE), and return on equity (ROE). Leverage was measured as total debt scaled by total assets. Investors are very concerned about the way capital structure changes with respect to risk exposure. Specifically, they are sensitive to changes in leverage upon any credit risk exposure announcement (Kalemli-Ozcan et al., 2012). Therefore, we examined how leverage explains variations in credit risk. Stern and Feldman (2004) postulate that on the grounds of the "too-big-to-fail" presumption that in response to the likelihood of protection from the government, creditors reduce discipline and banks pursue risks. Concurrently, large banks increase their leverage excessively, grant loans to risky borrowers and therefore experience deterioration in loan quality (Salas and Saurina, 2002). Therefore, we anticipate a positive association between leverage and credit risk.

[5] As a robustness check in Section 6.2, we employed provision for loan loss as an alternative proxy for NPLs.

[6] The "Z" score is applicable to non-manufacturing and emerging markets. It is given as 6.56A + 3.26B + 6.72C + 1.05D where A=working capital/total assets, B=retained earnings/total assets, C=earnings before interest and taxes/total assets, D=market value of equity/total liabilities.

Diversification income (DIV_INC) was measured as the ratio of non-interest income to total income. Due to the increase in non-trading activities (e.g., insurance underwriting, trading, and derivatives) among banks, their non-interest income has become very important. Banks that rely heavily on non-trading activities tend to report higher credit risk because of lack of expertise in non-core activities (Čihák and Hesse, 2010). Stiroh and Metli (2003) also report a positive association between diversification income and credit risk. In contrast, Hu et al. (2004) find that diversification income has a negative relationship with credit risk of banks in that diversification helps banks to gather enough information from various services, which helps lower credit risk. We therefore expect either a positive or negative relationship between DIV_INC and credit risk.

Bank size (SIZE) was measured as the natural logarithm of total assets. There is no consensus reached by prior literature on the association between bank size and credit risk. Large banks are in the best position to manage exposure to credit risk efficiently because of the diversification opportunities through their branch networks and expertise hence, reducing exposure to credit risk (Čihák and Hesse, 2010; Bourkhis and Nabi, 2013). Stern and Feldman (2004) postulate that large banks often overlook discipline and pursue risks. Concurrently, large banks increase their leverage excessively, grant loans to risky borrowers and therefore experience deterioration in loan quality (Salas and Saurina, 2002) Louzis et al. (2012) also find that bank size has a positive effect on credit risk. Based on this discussion, we therefore anticipate either a positive or negative association between bank size and credit risk.

Bank profitability was measured using return on equity (ROE), which captures the profit banks make given their total equity capital. Higher profits indicate lower credit risk because of the high projection for growth and resilience to adverse shocks. Ghosh (2015) argues that banks which are more profitable have little incentive to engage in more risky activities. This therefore reflects the "bad management hypothesis" of Berger and DeYoung (1997), which states that non-profitable banks record high levels of NPLs and are open to default. In contrast, Garcia and Robles (2008) report that highly profitable banks are exposed to high NPLs because they have an incentive to engage in risky behaviour. In addition, banks may increase their current earnings by conforming to liberal policies and cover the degree of exposure to problematic loans, with the motivation to convince market participants of their competent credit evaluation expertise (Rajan, 1994). For this reason, we anticipate either a positive or negative association between ROE and credit risk.

4.3.3 Country-level variables

Based on relevant literature, we employed the following country variables: financial stress (FIN_STR), corruption index (CI), government effectiveness (GOV_EFF), GDP growth rate, and inflation level. First, the financial stress index reports the exposure of a country's difficulties in cash flows. The index consists mainly of market-based and financial-based measures that account for segments of the financial market – the equity market, bond market, and foreign exchange market. It also considers the co-movement across market segments.[7] Cardarelli et al. (2011) argue that countries which are financially stressed experience contraction in their activities due to high procyclicality of leverage in their banking systems especially with countries with more arm's length financial system. This suggests the adverse impact on banks within such countries. Thus, we expect a positive relationship between financial stress and credit risk.

Second, Park (2012) finds that corruption causes problematic loans in the banking sector. Corruption was measured by means of the Corruption Perception Index (CPI) published by Transparency International. We employed the CPI because of the degree of corruption it covers, its assessment process, and its relatively long-term experience, among other things. The CPI ranges from 0 to 10 with higher scores reflecting lower corruption. Following Park (2012) and to ease interpretation, CI of a country is calculated as the residual of the maximum score of CPI (10) and a country's actual CPI, indicating that a lower CI would reflect lower corruption levels. Lower levels of corruption improve loan quality and are associated with more moderate growth (Chen et al., 2015; Park, 2012). We therefore anticipate a positive association between corruption level and credit risk.

Third, we also controlled for the quality of a country's governance structure. Prior research finds that strong country governance creates a highly optimistic economic environment which boosts banking operations (Chan and Mohd, 2016; Barth et al., 2013; Beck et al., 2003). To control for the country

[7]For more information about financial stress index methodology, See: https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp1873.en.PDF.

governance effect, we used one of the World Bank's Worldwide Governance indicators: government effectiveness, which reflects how far governments are actively committed to their aligned policies for economic development. This index has a scale of -2.5 (weak) to +2.5 (strong). An economy with an effective government improves its banks' efficiency because it boosts institutional quality. This implies that effective governance cuts down the cost of banking operations by reducing the level of bureaucracy that banks face. Consequently, as bank efficiency is associated with lower credit risk, we anticipate a negative association between government effectiveness and credit risk.

Fourth, a fall in GDP growth rate retards the improvement in debt servicing and concurrently increases NPLs. The reverse is true when an economy experiences economic boom (Ali and Daly, 2010; Salas and Saurina, 2002). On this basis, we anticipate a negative association between GDP and credit risk. Finally, inflation rate is measured as the percentage change in the consumer price index. The real value of loans is reduced upon a rise in inflation, whereas the real income of borrowers falls, which causes problems in loan repayments. Ideally, a rise in inflation facilitates repayment of loans by eroding the real value of loans outstanding. In contrast, Jankowitsch et al. (2007) report that a rise in inflation reduces the net income and debt servicing capabilities of borrowers, thus leading to an increased risk of NPLs. Therefore, we expect either a positive or negative association between inflation and credit risk.

4.4 The empirical model

We simultaneously analyse two different levels of credit risk by adopting a method that accounts for nested effects between levels. Put differently, characteristics at a higher level are likely to influence characteristics at a lower level (e.g., country level characteristics may influence bank level characteristics). Traditional single level modelling such as OLS regression fails on data which are hierarchically structured because its assumption of independence is violated. Hierarchical data cause dependencies between variables nested at different levels, called "intra-class dependency", hereafter called intra-class correlation (ICC).[8] In the present case, having banks nested in the same country could result in correlated errors. To

[8] ICC is the measure of a proportion of variation in the outcome variable that occurs between groups compared with the total variation present. ICC is calculated differently at each level by dividing the variance at the current level by the total variance. The total variance is the sum of variance at levels 1 and 2. For instance, level 1 ICC is calculated as level 1 σ^2 /total σ^2.

remedy this, we applied appropriate RMMA with a Restricted Maximum-Likelihood (REML) estimation approach to account for all levels of analysis. We employed RMMA for the following reasons: it provides a very flexible and powerful set of tools to handle hierarchical data, with respect to both model formulation and hypothesis testing. In addition, the use of a multilevel approach from a statistical point of view is shown to be more precise (Robinson, 2003) and conceptually enriching. Relying on individual level data as the traditional OLS does may miss important group-level effects, a fallacy commonly termed the atomistic myth**.** This may impact on the estimated variance and the available covariant effects, leading to the increase of Type I error. It can also result in substantive errors in interpreting the results of significant statistical tests.

In terms of estimation approach, we used REML estimation, for the following reasons. First, its variance estimation, specifically its degrees of freedom, yields unbiased estimates. It accounts for the sum of squared differences between individual values and mean values. It also accounts for the number of parameters being estimated when determining the degrees of freedom for random component estimation. REML is generally preferred in multilevel modelling, although for testing variance decomposition, Maximum Likelihood Estimation is necessary.

To examine the factors that capture variations in credit risks, our analyses were conducted at two different levels (i.e., bank and country levels). By applying RMMA, we assumed that variables are repeated over time and this presents additional analytical issues. That is, observations across time are correlated with the group that they belong to, and therefore they create a strong within-cluster correlation. Similarly, it may be assumed that banks which are nested in the same country have similar credit risk exposure because they have the same macroeconomic influences. To examine the factors that capture variations in credit risks, we first develop a null model, containing no predictors. This was done to obviate the influence of any fixed effect but focus instead on random effects, which, in turn, provided information relevant to the variance decomposition of the dependent variable. From this point, the null model estimated the relative importance of each level in the variance of the dependent variable. To identify the amount of proportion accounted for in the total variances in the credit risk measures accounted by the explanatory variables, we estimated adjusted R^2 at both levels.[9] Also, to assess the accuracy of each model in comparison with the null model, we included change in -2 Log Likelihood and Chi-square test. A decrease in the difference suggests improvement in the model. The dependent variables that we employed were NPL, OSCORE, ZSCORE, and DOWN, all of which are widely justified in the prior literature as credit risk measures.

Our model for RMMA at two different levels is shown below:

$$CR_{tik} = \beta_{0ik} + \beta_1 T_{tik} + \sum_{n=1}^{Nr} \beta_{rn} Xbl_{nik} + \sum_{n=1}^{Nc} \beta_{cn} Xcl_{nik} + \varepsilon_{tik} + r_{ik} \tag{1}$$

where CR_{tik} represents the credit risk (i.e., NPL, OSCORE, ZSCORE, or DOWN) of bank i in country k in year t. β_{0ik} is the intercept of bank i in country j. β_1 is the slope of the time-varying variables in relation to bank i in country j. T_{tik} is the linear component of time for bank i in country k at time t and is the main component at level (1), as shown in Model 1, the null model. B_{rn} denotes the effect of Xbl_{nik} (function of bank-level variables) on the linear component of time of the credit risk measure. B_{cn} is the effect of Xcl_{nik} function (of country-level characteristics) on the linear component of time in the credit risk measure. ε_{tik} and r_{ik} are the errors at level 1 and level 2, respectively. Appendix A describes these variables and their sources.

5. Results

5.1 Descriptive Statistics

Table 2 reports the summary statistics of the variables employed in our analysis. We observe high sample variability in NPLs and bank size. In our sample, the NPL ratio has a mean and standard deviation of 5.72% and 4.78%, respectively. A closer look at specific countries from our sample reveals that the highest NPL ratio (i.e., mean) is in Greece (i.e., 7.71%) and the lowest was in Ireland (0.67%). We observe from Appendix B that the trend in NPLs was lower and stable before the outbreak of the global financial crisis of 2008. Afterwards, NPLs are seen to rise, which signifies deterioration in asset quality and hence

[9] Adjusted R^2 is calculated as $1 – [(1 - R^2) * n-1/ (n – k - l)]$ where R^2 is computed as $1 – [(\sigma^2 m1 + \tau^2 m0)/ (\sigma^2 null + \tau^2 m0)]$. Hence, $m1$ is the current model's variance component, whereas $m0$ is the null model's variance component. k is total number of parameters; n is total sample size.

high exposure to credit risk. The high variation in this variable suggests that the bank-level characteristics that exist in the banking industry of Europe vary greatly across banks. The high variability in NPLs and size is consistent with the view of Cucinelli et al. (2018).

NPL_TXT has a mean and standard deviation of 0.07 and 1.60, respectively. From Figure 1, we observe a similar trend in NPL_TXT with NPLs for Barclays Bank Plc except for the years 2007 to 2009, when the two measures moved in opposite directions. This similar trend signifies the effectiveness of sLDA in capturing information relevant to credit risk from annual reports that reflect credit risk, as measured by NPLs. On average, the banks in our sample appear to be sound, as indicated by OSCORE and ZSCORE. The mean and standard deviation of OSCORE is reported as 0.35 and 0.15, respectively. The banks in our sample tend to be financially sound and not likely to be heading for bankruptcy (i.e., of high credit risk). This is supported by the mean of ZSCORE.

The mean of ROE is 0.10 which reflect that, on average, the banks in our sample were profitable. The diversification income with average standard deviation of 0.39 (0.27) indicates how European banks operate with different streams of income, which enables them to withstand any unplanned shocks. This can be directly traced to the openness of banks to non-trading activities. It also indicates the impact of the stringent regulation reflected in the credit risk management practices adopted by European banks over the years to reduce the risk faced through loan defaults. Most countries in our sample tend to have a corrupt economic system – the average corruption index is high at 0.55. Meanwhile, the governance of countries tends to be very efficient with a mean of 0.15.

[**Insert Table 2]**

As can be seen in Table 3, we report the Pearson and Spearman correlation coefficients for the variables employed in the analysis. The Pearson (Spearman) correlations are above (below) the diagonal. As anticipated, NPL_TXT is positively correlated with all the credit risk measures in our sample except DOWN. This result embodies preliminary evidence that NPL_TXT captures significant information from annual reports, which accounts for the levels of credit risk. NPL_TXT is negatively correlated with leverage*,* size, and inflation but has a positive association with diversification income, corruption index and GDP. Because correlation captures exclusively the covariance of two variables, it cannot be employed directly to establish causality, as discussed in the following sub-sections.[10]

[Insert Table 3]

5.2 Credit risk determination - impact of bank and country characteristics on NPL variations

Variance across time plays an important role in the exposure of banks' credit risk. It mirrors, for example, the significant impact of macroeconomic shocks that banks may encounter in a specific year. A report by the European Commission in 2018 reveals that the variations of NPLs in the EU resulted from the persistently asymmetric impact of the financial crises.[11] Consequently, the EC observes that these variations happen from time to time, despite the immense improvement in NPLs over the years. This significance indicates the existence of possible variations that could be explored further by adding bank variables (Models 2 - 3) and country variables (Model 4).

To examine the amount of NPL variation captured by our text-based (soft) credit risk measure, we include NPL_TXT in Model 2. The result shows that NPL_TXT is positive and significantly associated with NPLs, at a p-value of 0.000. The economic impact of this finding indicates that an increase by one standard deviation (1.60) in NPL_TXT would lead to an approximate increase of 5.425 (1.135 * 4.78) in NPLs. This result indicates that credit risk exposure as measured by NPLs is significantly influenced by NPL_TXT. It further indicates an accurate credit risk representation of NPL_TXT in explaining actual credit risk. This finding suggests that the narrative section of annual reports do contain information that is relevant to credit risk, a suggestion which complements prior studies (Campbell et al., 2014; Donovan et al., 2019). As regulators continue to strengthen their standards towards effective risk management, banks are compelled to disclose highly relevant information which serves as a basis for estimating future credit risk (Elshandidy, 2011). From Model 3, we find that all variables exhibit statistical significance with credit risk. Specifically, NPL_TXT continues to hold its positive significance at the p-value of 0.000. The significance of NPL_TXT suggests that its incremental information is economically meaningful and that banks disclose a high level of risk information, which raises their status by sending

[10] We addressed the possibility of multicollinearity by examining the variance inflation factors (VIFs) for each explanatory variable employed. The VIFs indicated that the degree of multicollinearity in this study was not severe. Additionally, from the correlation matrix, we observed the correlation coefficients among the independent variables to be less than 0.7, which dispelled the concern over highly correlated variables.

[11] For more information see: https://ec.europa.eu/info/publications/181128-non-performing-loans-progress-report_en

a signal to their investors about their competence and commitment in managing risk effectively (Elshandidy et al., 2015). We note that banks which are highly leveraged, diversified, and large exhibit increase in credit risk, whereas profitable banks experience lower credit risk which is consistent with prior studies (e.g., Salas and Saurina, 2002; Louzis et al., 2012). In essence, our evidence suggests that the existence of regulatory reporting standards and public scrutiny has compelled banks not only to increase transparency about their credit risk position but also to manage disclosure to create a useful impression of their credit risk (Jones et al., 2018). Thus, our findings support H1.

Under Model 4, we find financial stress and corruption level to be significant in explaining some of the variations in NPLs. The negative association of financial stress indicates that as countries experience high levels of stress in their financial structures, credit risk level begins to improve which is in contrast with our expectation. The positive significance of corruption indicate that the quality of loans is lower among banks in countries where the economic system is corrupt. The economic impact of the corruption level on credit risk is that a standard deviation increase (0.34) in corruption could result in a 5.731 (1.199 * 4.78) rise in credit risk between banks. This empirical finding complements the work of Chen et al. (2015), who find that lower levels of corruption are positively associated with loan quality and more moderate growth. In contrast with our expectation, government effectiveness showed no significant association with NPLs, indicating that it is not an important driver in explaining variations in NPLs. GDP shows a negative and significant relationship, while inflation shows positive statistical significance with NPLs both at p-values of 0.000. The impact of GDP indicates how banks from countries whose economic growth is impaired are likely to experience increase in credit risk which is consistent with the findings of Ali and Daly (2010) and Jiménez and Saurina (2006). Thus, after considering both bank and country-level variables, our findings support H1.

Concerning the between-level analysis or the variations at both levels, we observe that 20% of the variation in NPL (18% in the intercept and 2% in time) is between the banks (level 2); these results are significant at a p-values of 0.000 each. The remaining 80% is within banks over time (level 1) which is significant at a p-value of 0.000. We observe that bank-level variables captured 29% (adjusted-R^2) of all NPL variations between banks across the countries under consideration. In comparison with the null model, these variations decreased by 12%, to 6% at the p-value of 0.000. The variations in NPL within banks over the period under study increased by 13% to 93% compared to the null model at p-value of 0.000. These results in general indicate that simply considering bank variables improves the model's ability to explain variations in NPLs between banks, and, more importantly, reduces the unexplained variations between banks across countries, rather than the unexplained variations within banks over the thirteen years under consideration. Adding country variables improves the model's ability in explaining variations in NPLs at 35% (adjusted-R^2). In addition, the variation between banks decreased to 7% compared to the null model, at a p-value of 0.017 whereas the variations within banks over 2005 to 2017 increased to 92% at a p-value of 0.000. Country variables increased the model ability to explain NPL variations, as compared to the model that includes bank variables only (Models 2 - 3) as indicated by the change in -2 Log likelihood and the chi-square test. With hindsight, this is to be expected, because these factors might have a direct impact on each country's regulations, which in turn can be regarded as a core driver in credit risk narrations within each country, and thus a significant driver of NPL variations. Considering this result, it is essential to consider a country's financial stress index and corruption, owing to the significant and direct impact on credit risk disclosure practices. Consequently, a major aim of the IASB, that is, achieving international convergence, should be expanded from just adopting single sets of high-quality accounting standards, to considering the corruption levels of countries across the globe (Elshandidy et al., 2015).

[Insert Table 4]

5.3 Credit risk determination - impact of bank and country characteristics on OSCORE variations

From Model 1 (Table 5), we find that 91% of the total variation in OSCORE is accounted for within banks (level 1) while 9% is captured between banks (level 2), across time in the period 2005 to 2017. The variations exhibit statistical significance at the p-values of 0.000 each, which suggests that significant variation could be further explained at both levels.

Model 2 (Table 5) shows a positive and significant association of NPL_TXT with OSCORE at the p-value of 0.000. The economic impact of this finding suggests that a standard deviation increase (1.60) in NPL_TXT could lead to a 0.007 (0.044 * 0.150) increase in OSCORE, which is very small but may be bigger for some banks. This complements our earlier finding and therefore indicates that our text-based credit risk measure is an important driver in explaining credit risk and highly reflects that bank disclosure contains various characteristics such as firm, industry, and country-specific information that can be employed by participants of the debt market to assess credit risk of firms (Bonsall and Miller, 2017; Donovan et al., 2019). NPL_TXT therefore represents a sound measure of credit risk. This supports H1. We note from Model 3 that banks which are highly diversified and profitable, experience lower credit risk, probably due to lack of expertise to hedge some of their risk exposures. This is in contrast with the work of Čihák and Hesse (2010) and Stiroh and Metli (2003). In contrast with our expectation, leverage, bank size, and return on equity show no statistical significance in explaining some of the variations in OSCORE.

From Model 4 of Table 5 where we augment country variables, NPL_TXT is still positive and significant at the p-value of 0.000. We observe that banks which are highly leveraged and large experience increase in credit risk levels whereas banks which are highly diversified and profitable tend to exhibit improvement in adverse credit. About the country level variables, high degree of financial instability is significant and positively associated with credit risk. This complements Cardarelli et al. (2011) who find that increase in financial stress results in economic contraction due to the high procyclicality of banking systems, especially in the case of more arm's length financial systems. In contrast with our expectation, corruption and government effectiveness are not significant drivers in explaining variations in OSCORE. We further find that banks from countries with high GDP tend to exhibit low credit risk exposure at a p-value of 0.000, which complements prior studies (e.g., Jiménez and Saurina, 2006). Consistent with Jankowitsch et al. (2007) **i**nflation exhibits a positive association with OSCORE at a p-value of 0.000. These results support H1.

Adding NPL_TXT increased the model fit significantly. The variations between banks over time increased by 24% to 33% compared to the null model which is significant at a p-value 0.000 whereas the variations within banks significantly decreased from 91% to 67% across the countries under consideration. After augmenting other bank-level variables, the ICC reports that approximately 6% of the total variation between banks across countries over the thirteen-year period is captured, which is a 3% decrease when compared to the null model whereas the variations within banks increased by 3%, to

94% at p-value of 0.000. This indicated the existence of unexplained variations. After augmenting country variables, we note significant improvement in the model-fit of all OSCORE variations between banks across countries. Also, we observe an increase in the total variations within banks (level 1), significant at the p-value of 0.000, and a decrease in total variation between banks (level 2) across countries at a p-value of 0.046. This suggests that a greater amount of variance is explained between banks more than within banks over time across countries.

[Insert Table 5]

5.4 Credit risk determination - impact of bank and country characteristics on ZSCORE variations

The null model in Model 1 (Table 6) indicates that most of the variation (i.e., 84%) is accounted within banks (level 1) whereas the remaining variance is captured between banks (level 2). These variations exhibit significance at p-values of 0.000 each, which suggest that there are unexplained variations that could be accounted for at both levels. We therefore augment bank variables in Models 2 - 3 and country variables in Model 4.

In Model 2 (Table 6), NPL_TXT reveals a positive and significant association with ZSCORE at the p-value of 0.000. This implies that NPL_TXT captures significant information in explaining ZSCORE. We observe from Model 3 that banks which are large and profitable experience decline in their credit risk, whereas highly leveraged banks experience deterioration in credit risk. With regard to leverage, on the basis of the too-big-to-fail presumption, in response to the likelihood of protection from government, banks reduce their disciplinary measures and therefore pursue more risk (Stern and Feldman, 2004). We explain the effect of profitability to mean that as banks perform well, their credit risk falls, which confirmed the efficient hypothesis of Berger and DeYoung (1997). The negative association of profitability with credit risk also complements prior studies (Ali and Daly, 2010; Ghosh, 2015). Based on our findings, with reference to the psychological perspective, banks avoid practices of wrong commission by disclosing relevant credit risk information in the narrative sections of annual reports this is especially for banks with high credit risk exposure, knowing they could be subject to public scrutiny and omission could tarnish their reputations (Jones et al., 2018). The total variation in ZSCORE captured by Model 3 is seen at 11% (adjusted-R^2). This significant variation demonstrates the effectiveness of our text-based credit risk measure in reflecting actual credit risk, which is in agreement with prior studies (e.g., Campbell et al., 2014; Donovan et al., 2019). This therefore supports H1.

Under Model 4 (Table 6) NPL_TXT maintains its significant association at the p-value of 0.000. Banks that are highly leveraged report increase in credit risk whereas profitable and large banks report lower levels of credit risk. This supports our previous findings. In contrast with our expectation bank size is not significant in explaining some of the variations in credit risk. Regarding country variables, financial stress exhibit positive and significant association with credit risk. Corruption and government effectiveness showed no statistical significance in explaining credit risk. We further find GDP and inflation to be significant at p-values of 0.000 and 0.078, respectively. A rise in GDP retards debt servicing and therefore increases credit risk (Ali and Daly, 2010; Salas and Saurina, 2002), while increase in inflation reduces net income and debt servicing, which concurrently increases credit risk (Jankowitsch et al., 2007).

The bank variables captured 11% (adjusted R^2) of all ZSCORE variations between banks across the countries under consideration. When compared to the null model, we observe a 4% decrease in the variations captured between banks significant at a p-value of 0.000 and an increase of 5% in the variations within banks significant at a p-value of 0.000. This indicates that considering bank variables alone improves the model-fit in capturing variations in ZSCORE between banks, and more importantly, reduces the unexplained variations between banks across countries, rather than the unexplained variations within banks over the thirteen years under consideration. We observed that adding country variables improved the model-fit slightly. The variations captured between banks further decreased by a percentage whereas the variations within banks across the countries under consideration increased significantly when compared to the null model. Total ZSCORE variation explained after adding country variables is seen at 13% as indicated by the adjusted-R^2.

[Insert Table 6]

5.5 Credit rating downgrades and annual report predictions

We employed logistic multilevel analysis to examine whether NPL_TXT predicted future rating downgrades. CRAs engage both the soft and hard content of annual reports in their independent assessment. We therefore examined whether NPL_TXT is associated with future credit rating downgrades. We report the results in Table 7.

From Model 1 (Table 7), consistent with our expectations, it appears that NPL_TXT shows positive statistical significance at the 1% significance level, indicating that the text-based credit risk measure contains economically significant insights in predicting rating downgrades. This is consistent with the work of Donovan et al. (2019). The Pseudo R^2 was 5.9%. This significant relationship was further examined by augmenting the bank variables in Model 2. The positive and significant relationship existing between NPL_TXT and rating downgrade still holds after accounting for bank variables. All the other bank variables are significant. This leads us to accept H2.

We further examined this significant relationship by adding country variables as shown in Model 3. As reported, the significance of NPL_TXT is maintained after accounting for the country variables. Its positive significance is seen at the 1% significance level. In order to assess the economic worth of the model in Model 3, we observe from the NPL_TXT that issuers that fall in the bottom quartile have a 4.70% chance of being downgraded, whiles issuers in the top quartile have a 6.10% chance of being downgraded with regard to the marginal effect. Moreover, it could be observed that the Pseudo R^2 improved marginally from Model 2 to Model 3 with an incremental change of approximately 1.1% (0.076 - 0.065). In Model 3, the coefficient of NPL_TXT indicates that a one percent increase in standard deviation results in an approximately 9.9% chance of a future credit rating downgrade by S&P ratings, relative to the overall likelihood of credit ratings reported in Model 2. The economic significance was reasonably large, therefore, suggesting that NPL_TXT was a strong measure for capturing credit risk information relevant to explanations of the issuer credit risk, as evident from rating downgrades from S&P. This complements prior studies which suggest that the narrative section of corporate disclosure does contain risk relevant information that aids CRAs in their independent rating assessments (Boumparis et al., 2019; Bozanic and Kraft, 2018). Thus, annual reports contain incremental information (Debrosky et al., 2013) which serves as a forward-looking source for CRAs (Bozanic et al., 2018; Kuang and Qin, 2013). Regarding the controls, apart from government effectiveness; all other variables are significant in rating downgrades. This leads us to accept H2. We observe that the addition of the bank-level variables improved the model fit to 6.5% compared to model 1 as indicated by the Pseudo R^2. This suggests that the bank variables are significant drivers in explaining variations in credit rating downgrade. In addition, country variables improved the model fit to 7.6% when compared to Model 1.

[Insert Table 7]

5.6 The interaction between bank-level and country-level factors in explaining variations on credit risk Cross-level interaction

To test our H3, we follow the approach by Marcato et al. (2018). The interaction term between text-based credit risk and each of the five country level variables is employed to capture that effect. As suggested by Baron and Kenny (1986), the variables that form that interaction must also be included in the specification. Because of the interaction, a general concern of multicollinearity arises, which might influence the accuracy of our estimate and interpretation. To test for potential multicollinearity, we follow Aiken et al. (1991) and Jaccard et al. (2003) and mean centre the variables for interaction. We report the results in Table 8 (Models 2-4).

From Table 8 (Model 2), we examine the possible existence of a moderating role of our country level variables, interacting each country variable with the text-based credit risk measure. We find that NPL_TXT is positively associated with NPL at a p-value of 0.000, which confirms our previous findings. In terms of the country level variables, financial stress and inflation are the only moderators that enhance the effect of text-based credit risk on credit risk as measured by NPL. Specifically, banks operating in countries with high financial stress, exhibit a significant marginal high credit risk (coefficient = 1.278, p-value = 0.000). At the same time, banks in countries with high inflation report significant marginal high credit risk levels (coefficient = 1.831, p-value = 0.036). When we observe NPL_TXT, financial stress and inflation increase the credit risk level. Specifically, when there are no moderator values on the existing relationship between NPL_TXT and NPL, the effect of a one standard deviation increase in NPL_TXT results in a 13.72 (4.78 * 2.87) increase in NPL. However, if financial stress moderates the relationship between NPL_TXT and NPL, the marginal increase in credit risk is 9.77 (1.278 * 4.78 * 1.60). This indicates that a bank is more likely to report increase in credit risk when it operates in a country where financial stress is high. For inflation, the significant and positive interaction term (INF * NPL_TXT) reflect a marginal increase of 14.00 (1.831 * 4.78 * 1.60) in credit risk. This shows a relative increase by 0.213 (1.831/8.601). The ICC reports that approximately 9% (8% in the intercept and 1% in time) of the total variation between banks across countries over the thirteen-year period is captured. The variations within banks increased by 11%, to 91% compared to the null model at a p-value of 0.000, whereas we observe a decrease in the total variations within banks (level 1), significant at the p-value of 0.000. The improvement in model-fit is observed at 37% (adjusted-R^2) of all NPL variations between banks across countries higher than models without interaction terms (Baron and Kenny, 1986). This suggests that some of the variance is explained owing to the interaction terms. In line with our expectation, these results confirm the notion that a country's financial stress and inflation level increase the interpretation of the relationship existing between text-based credit risk and credit risk, and thus support H3.

We also retest our third hypothesis (H3), by employing principal component factor analysis (PCFA) to assess the aggregate effect of country level characteristics (financial stress, corruption, governance, GDP, and inflation). With PCFA, we create one hybrid variable that captures common variation of the five country variables, hence mitigating multicollinearity (Hair et al., 1998).[12] Model 3 of Table 8 reveals that the hybrid variable (FACTOR) is significant and negatively associated with credit risk at the p-value of 0.027. The same direction of significance is observed for the cross-level interaction FACTOR * NPL_TXT at the p-value of 0.000. Thus, country characteristics reduces, the effect of credit risk disclosure in reflecting credit risk. That is, there is a significant relative reduction of 1.54 (0.447/0.290) in credit risk levels as measured by NPL. The economic impact of this finding reveals that a one standard deviation increase in country characteristics results in 2.137 (4.78 * 0.447) decrease in credit risk. This suggests that the relationship between credit risk and credit risk disclosure gets weaker with decrease in country characteristics, which supports H3.

To further assess the cross-level interaction of hypothesis (H3), we aggregate all our five country-level variables (financial stress, governance, corruption, GDP, and inflation level) into one aggregate score (Country-Aggregate) to capture the overall country effect. We follow the following steps; first we

[12] Our unreported results show that over 61% of the variance is captured by FACTOR, which has also an eigen value of 1.9, suggesting that this factor has the highest loading. We validate the FACTOR by using both the Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy (0.64 against 0.50 as a common value for KMO) and Bartlett's test of sphericity (Chi-square 50.022 at a p-value of 0.000).

transform all country variables (i.e., continuous variables) into dichotomous variables based on the median of each variable, then we take the sum of the five dichotomous variables which we label as AGGREGATE. To this end, our aggregated country variables is formulated based on the median of the equally weighted summed values of the five country level variables. Our results reported in Model 4 of Table 8 show that Country-Aggregate (AGGREGATE) is significant and negatively associated with credit risk. The effect of AGGREGATE on credit risk disclosure becomes positive and significant. With a p-value of 0.008, the interaction term AGGREGATE* NPL_TXT reflects that the strength of country characteristics increases the effects of credit risk disclosure in driving low credit risk. Therefore, banks with high credit risk disclosure operating in countries with high Country-Aggregate have lower levels of credit risk as measured by NPL when compared to banks operating in low economic institution economies. The economic impact of this finding indicates that a one percentage standard deviation increase results in a 0.038 (4.78 * 0.008) decrease in credit risk. We also observe that adding the interaction term improves the model fit in explaining variations in credit risk. The total variation captured is seen at 41% (adjusted R^2) and this therefore supports H3.

[Insert Table 8]

6. Robustness checks

6.1 Alternative econometric model (predictive power of NPL_TXT using OLS regression)

We employed traditional OLS regression using our full sample to analyse the impact of NPL_TXT on the credit risk variables without nested effects, as specified in Equation (3). The results are reported in Table 9. Serial correlation was addressed by clustering standard errors at bank level.

$$CR_{ik} = \beta_{0ik} + \beta_1 Xbl_{ik} + \beta_2 Xcl_k + \varepsilon_{ik} \quad (2)$$

CR_{ik} represents the credit risk measures of bank *i* in country *k*. Credit risk measures were the same as employed above. β_{0ik} is the intercept, Xbl_{ik} represents bank-level variables for bank *i* in country *k, Xcl_k* represents the country variables, and ε_{ik} is the error term. All bank- and country-level variables are the same as discussed earlier.

From Panel A of Table 9, we report the association between NPL_TXT and NPLs without any bank and/or country variables for Model 1. It shows a positive and significant association at the p-value of 0.000. This suggests that our text-based credit risk measure captures useful language in annual reports for explaining NPLs. This finding complements prior studies which suggest that annual reports are informative and present a good reflection of banks' credit risk (e.g., Campbell et al., 2014; Donovan et al., 2019). The variation in NPLs captured by NPL_TXT is observed to be 15%. Next, we examined the incremental power of NPL_TXT in explaining credit risk as measured by NPLs. We included bank variables, as reported in Model 2 and country variables, as reported in Model 3. All continuous variables were winsorised at the 1st and 99th percentiles to control for the influence of outliers. From Model 2, it appears that NPL_TXT is significant and positively associated with NPLs. Apart from leverage and return on equity, all other bank variables are important drivers in explaining the variation in NPLs. Adding bank variables improved the model fit to 23%. Model 3 reports a significant and positive relationship between NPL_TXT and NPLs at the p-value of 0.000. We observe a switch in significance for leverage and return on equity at the p-values of 0.000. We ascribe the change in significance to country-level effects. The positive significance of corruption at the p-value of 0.000 suggests that banks are likely to experience higher exposure to credit risks as the corruption level increases (Park, 2012). Government effectiveness also exhibits a negative significant association with NPLs at the p-value of 0.003, suggesting that banks operating in a less well-governed economic environment tend to be experience deterioration in their credit risk (Barth et al., 2013; Chan and Mohd, 2016). Financial stress is not significant in explaining the variation in NPLs. Most importantly, the adjusted R^2 is observed at 31%, which suggests that NPL_TXT is economically informative. This result is consistent with our expectation and therefore supports H1.

Panel B of Table 9 shows that NPL_TXT is a significant driver in explaining credit risk as measured by OSCORE. NPL_TXT is positively associated with OSCORE in all three models. Model 1 reveals that the variation captured by NPL_TXT in explaining OSCORE is 17% (adjusted R^2). In Model 2, the adjusted R^2 is 18%. The slight improvement in model fit signifies the importance of NPL_TXT in explaining OSCORE. Model 3 shows that NPL_TXT is positively associated with OSCORE at the p-value of 0.000. The adjusted R^2 improves by 7% (25% - 18%), which suggests that NPL_TXT contains economically significant information, making it a good measure of credit risk. This finding complements prior studies. giving credence to the statement that disclosures from annual reports are informative and present a good measure of banks' (Campbell et al., 2014; Donovan et al., 2019). This support H1.

Panel C of Table 9 shows the relationship between NPL_TXT and ZSCORE. Model 1 reports a positive and significant association of NPL_TXT at the p-value of 0.000. We observe the adjusted R^2 at 6%. This confirms our earlier findings. From Model 2, we find that the positive significance of NPL_TXT continues to hold after augmenting bank variables. Leverage is positive and significant at a p-value of 0.000. The positive association of NPL_TXT still holds after augmenting both bank and country variables. The total variation captured by this model is 10% (adjusted R^2).

[Insert Table 9]

In Model 1 of Table 10, NPL_TXT showed a positive association with a future rating downgrade at the p-value of 0.00. This indicates that the text-based credit risk measure contains economically significant information for predicting rating downgrades. The Pseudo R^2 is 5.9%. Model 2 augments bank variables; the results reveal most of the coefficients are anticipated. In Model 3, the positive association of NPL_TXT continues to hold after adding country variables. In order to assess the economic worth of the model, we observe from the NPL_TXT that issuers which fall into the bottom quartile have an approximate chance of 5.02% of being downgraded, whiles issuers in the top quartile have a 6.16% chance, approximately, of being downgraded with regard to the marginal effect. Moreover, it can be observed that the Pseudo R^2 improves significantly from Model 1 to Model 3, an incremental change of approximately 2% (7.5% - 5.9%). In Model 3, the coefficient of NPL_TXT suggests that a standard deviation could result in approximately 8.9% chance in future downgrade by S&P ratings. This suggests that NPL_TXT is a powerful measure for capturing credit risk information that relates to issuer credit risk, as evident in credit rating downgrades. Regarding country variables, apart from government effectiveness, all the variables are significant in explaining the variations in future credit rating downgrades.

[Insert Table 10]

6.2 Alternative credit risk measure (Provision for loan loss)

Current research shows that the NPLs reported in the statements of firms' financial position may be influenced by the variations in the accounting policies and the definition of default adopted across

Europe. In addition, since NPLs remain a severe and major way of exposing banks to credit risk in Europe, scholars tend to disregard the credit risk embodied in cases of exposure that have not deteriorated. However, the risk associated with non-deteriorated exposures affects capital ratios and risk weighted averages and may represent an incentive for a bank to request the validation of internal models. Because overlooking this component of credit risk may result in bias in our analysis, our first robustness test is to estimate our model using an alternative credit risk measure, specifically, provision for loan loss ratio (PROV_LL). One benefit of this variable over the NPL ratio is that it captures the credit risk embodied in performing loans (Cucinelli et al., 2018). This indicates that PROV_LL has a positive link with NPLs and therefore serves as a means of reviewing capital structure. Its disadvantage is that it responds more quickly to the deterioration of a loan portfolio and may be more volatile.

As can be seen from Table 11, the results confirm a positive and strong statistical significance of NPL_TXT with PROV_LL. There are some differences to note in the country-level coefficients. We observe a change in the role of government effectiveness, here significant and negatively related to provision for loan loss. We can interpret this to mean that the greater the deterioration in a country's economic efficiency, the higher the credit risk exposure of banks in the country (Barth et al., 2013; Chan and Mohd, 2016). Following the argument above in respect of provision for loan loss as tantamount to ignoring non-deteriorated exposures, it seems reasonable to find this negative relationship. This evidence supports our hypothesis that text-based credit risk measures contain economically significant information which reflects actual credit risk, signifying that our findings are robust to the proxy of credit risk.

[Insert Table 11]

6.3 Large and small samples

To examine whether our results are driven by our sample, we run further analysis on two subsamples. Table 12 shows the complete model that we apply to the two subsamples: sample 1 includes countries with more than a hundred observations, which we label the large sample and sample 2 contains countries with fewer than a hundred observations, as the small sample. By analysing these two blocks of sample countries separately, we intend to verify whether the same drivers of credit risk as measured by NPL, OSCORE, and ZSCORE apply to both large and small samples. We are interested in observing whether findings in the large sample also apply to the small sample.

We observe a similar impact at the bank level. All the bank characteristics exhibit the same impact on explaining NPLs. Most importantly, NPL_TXT is positive and significant for both subsamples in relation to NPL, OSCORE, and ZSCORE in turn. The most significant differences between the two subsamples rest on some country variables. For example, financial stress and corruption exhibit different effects, depending on which of the two samples they occur in. In the large sample, financial stress shows no significant relationship in explaining NPLs, but it is negative and significant in the small sample. Conversely, corruption is positive and significant in explaining NPLs for the large sample but insignificant for the small sample.

[Insert Table 12]

6.4 Endogeneity problem

We further check whether our results are affected by the issue of endogeneity. The problem of endogeneity arises from omitted variables and/or from reverse causality. Omitted variables cause bias in estimates due to the effect of unobserved heterogeneity of a firm-specific and/or time-invariant nature (Elshandidy et al., 2015).[13] The reverse causality concern, on the other hand, arises when the effect of textual credit risk information on credit risk measures can simultaneously occur in the other direction, suggesting that firms with high credit risks are likely to reveal information on credit risk (e.g., Jones et al., 2018). In the following subsections, our paper provides three common approaches in dealing with specified causes for endogeneity problems.

6.4.1 Generalized Method of Moments (GMM)

To control for potential endogeneity and persistence in our dependent variable (credit risk estimates), we employ GMM to address potential selection bias without external instrumental variables (Kim and Frees, 2007). The application of GMM tackles inconsistencies in estimates due to unobserved heterogeneity across banking institutions (Nguyen et al., 2012; Wu and Bowe, 2012). In addition, it has the capability of dealing with reverse causality between credit risk and text-based credit risk. With our nested data, this

[13] This concern can be mitigated by applying the fixed-effect estimations that have been introduced in our previous Tables 4 -7. Furthermore, applying the RMMA accounts for firm-specific and time-invariant effects that can possibly affect credit risk measures (Elshandity et al., 2013, 2015).

technique employs both the within and between variations of exogeneous variables, but just the within variation of the variables is regarded endogenous (Kim and Frees, 2007). We re-estimate model 4 of our main analysis (Tables 4-6). From Panel A of Table 13, we find similar results to our previous results that NPL_TXT is positively associated with all credit risk measures confirming the significance of credit risk information disclosed in the narrative section of bank's annual report. Thus, our findings are not subject to endogeneity concern.

[Insert Table 13]

6.4.2 Propensity Score Matching (PSM)

To further address potential endogeneity concerns, we use the Propensity Score Matching (PSM) approach as one of the most common techniques to address potential endogeneity (Roberts and Whited, 2013; To et al., 2018). We re-estimate Model 4 of our main analyses. First, we form sub-samples from our data: banks with high credit risk disclosure (the treatment group) and banks with low credit risk disclosure (the control group) and match the treatment group to the control group. Specifically, we employ a logit model to estimate the propensity score for each observation which is used to predict the likelihood of high NPL_TXT (the likelihood of being treated) as a function of bank-level characteristics (LEV, DIV_INC, SIZE, and ROA). Next each high NPL_TXT observation is matched with the bank in the control group which has the closest propensity score in relation to the treated bank, using the nearest neighbours matching technique (Bonaventura et al., 2018). We also employ common support to remove all extreme boundaries by excluding banks in the control group whose propensity score is lower than the minimum or higher than the maximum propensity score among high NPL_TXT in the treatment group and vice versa.

We report the results in Panel B of Table 13 which shows the impact of the treated group on credit risk measures, using the multilevel propensity score approach. Columns 1 to 3 reveal that the coefficients of the NPL_TXT variable that reflects low credit risk disclosure of banks is positive and significant, confirming our previous findings that text-based credit risk is positively associated with credit risk after addressing potential selection bias concerns using propensity score matching.

6.4.3 Lag approach for principal independent variables

To address the reverse causality concern, we regress our dependent variable on the lag of principal independent variables (e.g., Elshandidy and Neri, 2015). Thus, we replicate Model 4, which includes both bank and country level variables, of our main analysis by regressing the current year's credit risk on the relevant previous year's variables, as indicated in the equation below:

$$CR_{tik} = \beta_{0ik} + \beta_1 T_{tik} + \sum_{n=1}^{Nr} \beta_{rn} Xbl_{nikt-1} + \sum_{n=1}^{Nc} \beta_{cn} Xcl_{nik} \quad (3)$$

From Panel C of Table 13 (Model 1), we observe that the lag coefficient of NPL_TXT at time *t-1* is positive and significantly associated with NPL and ZSCORE at 1% significance level. This implies that an increase in credit risk identified from annual reports in immediately prior years leads to an increase in NPLs the year after. We relate this effect to the information about credit risk disclosed by firms in their annual reports. We find that a one percent increase in standard deviation of NPL_TXT_{t-1} on average predicts an increase in NPL_t by 1.395 (0.292 * 4.78) and ZSCORE by 0.007 (0.025 * 0.29) which confirms our previous findings. Financial stress and GDP significantly explain variations in all credit risks. In summary, the results in Table 13 reveal the coefficients of NPL_TXT to be statistically significant and consistent with our models in Table 4 to 7, and, more importantly have theoretically plausible signs. Additionally, the weight of the coefficients and the explanatory power of the model remain consistent with our main models (i.e., Tables 4 to 7). This therefore supports the economic significance of a text-based credit risk measure in explaining credit risk.

7. Conclusion

We employ sLDA to extract relevant (soft) credit risk information from annual reports. We associate variations in credit risk as measured by NPL, OSCORE, ZSCORE, and DOWN with variations in both bank-level characteristics and country level characteristics across 19 EU countries, over the period from 2005 to 2017. We find that our text-based credit risk measure explains a substantial portion of the variation in all credit risk measures. This finding holds after several further analyses. In terms of country-level characteristics, we find that financial stress, inflation and GDP growth rate have significantly high explanatory power over all credit risk measures variations over time. Corruption level explains significant variations in NPL and DOWN, but it is not a significant driver in explaining variations in ZSCORE and OSCORE. Again, government effectiveness does not contribute to explaining variations in all the credit risk measures.

These results have theoretical and practical implications. Theoretically, our evidence adds significantly to machine learning in the financial literature by endorsing the current importance of widening this research scope to give more attention to the application of such algorithms in exploring various firm fundamentals. Furthermore, our result suggest that more attention should be given to variables that may explain the variations in credit risk for banks over time. Our suggestion is consistent with the recent trend (Barth et al., 2013; Beck et al., 2003; Chan and Mohd, 2016) in the literature which explores the impact of widely varying governance indicators on bank characteristics (e.g., efficiency, loan quality). Practically, our results signal that applying machine learning algorithms to disclosure content improves the identification of credit risk (modelling). This reveals additional information to market participants about the credibility of the risk information disclosed by firms. Our results also support the trend in EU regulation which encourages firms to disclose relevant information about their risk exposure rather than waiting for them to do so voluntarily.

Our paper has some limitations. First, although credit risk remains one of the key risks that EU banks face, there are essentially other risks (e.g., liquidity, market, and operational risks), which could be identified from the narrative sections of banks' annual reports. Further research might examine the extent of these different types of risk disclosure. Second, while our paper relies mainly on annual reports to extract credit risk information, there are other outlets (e.g., earnings releases, conference calls, and media coverage) that can be usefully used, then the effect of credit risk information extracted from each source can be integrated and its impact on credit risk observed. Further, with a different analytical lens, further research can employ informal sources of corporate information such as social media. Finally, our paper concerns the quantity of textual credit risk in determining credit risk in European banks. However, further research might usefully link the quality attributes (Miihkinen, 2013; Elshandidy et al., 2018b) or the compliance/non-compliance (e.g., Adam- Müller and Erkens, 2020) of credit risk disclosure to debt market indicators (i.e., cost of debt, credit risk, financial distress, and credit ratings).

References

Adam-Müller, A. F., Erkens, M.H., 2020. Risk disclosure noncompliance. J. Account. Pub. Pol, forthcoming.

Abbdullah, M., Shukor, Z.A., Mohamed, Z.M., Ahmad, A., 2015. Risk management disclosure: a study on the effect of voluntary risk management disclosure toward firm value. J. Appl. Account. Res. 16 (3), 400–432.

Agostino, M., Drago, D., Silipo, D.B., 2011. The value relevance of IFRS in the European banking industry. Rev. Quant. Financ. Account. 36 (3), 437–457.

Aiken, L. S., West, S. G., Reno, R. R., 1991. Multiple regression: testing and interpreting interactions. Sage.

Ali, A., Daly, K., 2010. Macroeconomic determinants of credit risk: Recent evidence from a cross country study. Int. Rev. Financ. Anal. 19 (3), 165–171.

Allini, A., Manes Rossi, F., Hussainey, K., 2016. The board's role in risk disclosure: an exploratory study of Italian listed state-owned enterprises. Public Money Manag. 36 (2), 113–120.

Altman, E.I., 1968. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. J. Finance 23 (4), 589–609.

Altman, E.I., Bharath, S.T., Saunders, A., 2002. Credit ratings and the BIS capital adequacy reform agenda. J. Bank. Financ. 26 (5), 909–921.

Altman, E.I., Hartzell, J., Peck, M., 1998. Emerging market corporate bonds: a scoring system. pp. 391–400.

Aziz, S., Dowling, M.M., Hammami, H., Piepenbrink, A., 2019. Machine Learning in Finance: A Topic Modeling Approach. SSRN Electron. J.

Balakrishnan, R., Danninger, S., Elekdag, S., Tytell, I., 2011. The transmission of financial stress from advanced to emerging economies, in: Emerging Markets Finance and Trade. 47(2), 40-68.

Bao, Y., Datta, A., 2014. Simultaneously discovering and quantifying risk types from textual risk disclosures. Manage. Sci. 60 (6), 1371–1391.

Baron, R.M., Kenny, D.A., 1986. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. J. Pers. Soc. Psychol. 51 (6).

Barth, J.R., Lin, C., Ma, Y., Seade, J., Song, F.M., 2013. Do bank regulation, supervision and monitoring enhance or impede bank efficiency? J. Bank. Financ. 37 (8), 2879–2892.

BCBS, 2015. Revised Pillar 3 disclosure requirements. Basel Committee on Banking Supervision, BIS, Basel, Switzerland.

Beck, T., Demirgüç-Kunt, A., Levine, R., 2003. Law, endowments, and finance. J. Financ. Econ. 70 (2), 137–181.

Berger, A.N., DeYoung, R., 1997. Problem loans and cost efficiency in commercial banks. J. Bank. Financ. 21 (6), 849–870.

Bharath, S.T., Sunder, J., Sunder, S. V., 2008. Accounting quality and debt contracting. Account. Rev. 83 (1), 1-28.

Bianco, T., Gramlich, D., Oet, M., Ong, S., 2012. Financial Stress Index: A Lens for Supervising the Financial System. Work. Pap.

Bitar, M., Pukthuanthong, K., Walker, T., 2018. The effect of capital ratios on the risk, efficiency and profitability of banks: Evidence from OECD countries. J. Int. Financ. Mark. Institutions Money 53, 227–262.

Blei, D.M., Ng A.Y., Jordan, M.I., 2003. Latent dirichlet allocation. J. Mach. Learn. Res. 3(4-5), 993-1022.

Blei, D.M., McAuliffe, J., 2007. Supervised topic models. Neu. Info. Proc. Sys. 20:121128.

Blei, D., Boyd-Graber, J., Zhu, X., 2007. A topic model for word sense disambiguation, in: EMNLP-CoNLL 2007 - Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. 1024-1033.

Bliese, P.D., Hanges, P.J., 2004. Being both too liberal and too conservative: The perils of treating grouped data as though they were independent. Organ. Res. Methods. 7 (4), 400-417.

Bonaccorsi di Patti, E., Sette, E., 2016. Did the securitization market freeze affect bank lending during the financial crisis? Evidence from a credit register. J. Financ. Intermediation.25, 54-76.

Bonaventura, M., Giudici, G., Vismara, S., 2018. Valuation and performance of reallocated IPO shares.

J. Int. Financ. Mark. Institutions Money. 54, 15-26.
Bonsall, S.B., Buskirk, V., Ahn, M., 2014. Asymmetric disclosure to credit rating agencies. University of Texas. Working Paper.
Bonsall, S.B., Holzman, E.R., Miller, B.P., 2017. Managerial ability and credit risk assessment. Manag. Science. 63 (5), 1425-1449.
Bonsall, S.B., Miller, B.P., 2017. The impact of narrative disclosure readability on bond ratings and the cost of debt. Rev. Account. Stud. 22 (2), 608–643.
Boumparis, P., Milas, C., Panagiotidis, T., 2019. Non-performing loans and sovereign credit ratings. Int. Rev. Financ. Anal. 64 (May), 301–314.
Bourkhis, K., Nabi. M.S., 2013. Islamic and conventional banks' soundness during the 2007–2008 financial crisis. Rev. Finan. Econ. *22* (2), 68-77.
Bozanic, Z., Roulstone, D.T., Van Buskirk, A., 2018. Management earnings forecasts and other forward-looking statements. J. Account. Econ. 65 (1), 1–20.
Bozanic, Z., Kraft, P., 2018. Qualitative corporate disclosure and credit analysts' soft rating adjustments. SSRN Electron. J.
Butaru, F., Chen, Q., Clark, B., Das, S., Lo, A.W., Siddique, A., 2016. Risk and risk management in the credit card industry. J. Bank. Financ. 72 (1), 218–239.
Campbell, J.L., Chen, H., Dhaliwal, D.S., Lu, H. min, Steele, L.B., 2014. The information content of mandatory risk factor disclosures in corporate filings. Rev. Account. Stud. 19 (1), 396–455.
Cardarelli, R., Elekdag, S., Lall, S., 2011. Financial stress and economic contractions. J. Financ. Stab. 7 (2), 78–97.
Chan, S.G., Mohd, M.Z., 2016. Financial market regulation, country governance, and bank efficiency: evidence from East Asian countries. Contemp. Econ. 10 (1), 39–54.
Chen, M., Jeon, B.N., Wang, R., Wu, J., 2015. Corruption and bank risk-taking: evidence from emerging economies. Emerg. Mark. Rev. 24 (4), 122–148.
Čihák, M., Hesse, H., 2010. Islamic banks and financial stability: an empirical analysis. J. Financ. Serv. Res. 38 (2), 95–113.
Cucinelli, D., Battista, M.L. Di, Marchese, M., Nieri, L., 2018. Credit risk in European banks: the bright side of the internal ratings based approach. J. Bank. Financ. 93, 213–229.
Çule, M., Fulton, M.E., 2013. Corporate governance and subjective well-being. Appl. Econ. Lett. 20, 364-367.
Das, S.R., Hanouna, P., Sarin, A., 2009. Accounting-based versus market-based cross-sectional models of CDS spreads. J. Bank. Financ. 33 (4), 719–730.
DeBoskey, D.G., Gillett, P.R., 2013. The impact of multi-dimensional corporate transparency on us firms' credit ratings and cost of capital. Rev. Quant. Financ. Account. 40 (1), 101–134.
Donovan, J., Jennings, J.N., Koharki, K., Lee, J.A., 2019. Measuring credit risk using qualitative disclosure, SSRN Electronic Journal.
Elamer, A.A., Ntim, C.G., Abdou, H.A., Owusu, A., Elmagrhi, M. Ibrahim, A., 2020. Are bank risk disclosures informative? Evidence from debt markets. Int. J. Finan. Econ, forthcoming.
Elsayed, M., Elshandidy, T., 2020. Do narrative-related disclosures predict corporate failure? Evidence from UK non-financial publicly quoted firms. Int. Rev. Finan. Anal., 71, 1-20.
Elshandidy, T. M. 2011. Risk reporting incentives: A cross-country study. PhD thesis, the University of Stirling, UK.
Elshandidy, T., Fraser, I., Hussainey, K., 2015. What drives mandatory and voluntary risk reporting variations across Germany, UK and US? Br. Account. Rev. 47 (4), 376–394.
Elshandidy, T., Fraser, I., Hussainey, K., 2013. Aggregated, voluntary, and mandatory risk disclosure incentives: Evidence from UK FTSE all-share companies. Int. Rev. Financ. Anal. 30, 320–333.
Elshandidy, T., Neri, L., 2015. Corporate governance, risk disclosure practices, and market liquidity: comparative evidence from the UK and Italy. Corp. Gov. An Int. Rev. 23 (4), 331-356.
Elshandidy, T., Shrives, P.J., 2016. Environmental incentives for and usefulness of textual risk reporting: evidence from Germany. Int. J. Account. 51 (4), 464–486.
Elshandidy, T., Shrives, P.J., Bamber, M., Abraham, S., 2018a. Risk reporting: a review of the literature and implications for future research. J. Account. Lit. 40, 54–82.
Elshandidy, T., Neri, L., Guo, Y., 2018b. Determinants and impacts of risk disclosure quality: evidence

from China. J. Appl. Account. Res. 19 (4), 518–536.
Erkens, M., 2012. Reporting incentives and enforcement: Impact on corporate risk disclosure (Working Paper, German Economic Association of Business Administration)
Feldman, R., Govindaraj, S., Livnat, J., Segal, B., 2010. Management's tone change, post earnings announcement drift and accruals. Rev. Account. Stud. 15 (4), 915–953.
Fiordelisi, F., Marques-Ibanez, D., Molyneux, P., 2011. Efficiency and risk in European banking. J. Bank. Financ. 35 (5), 1315–1326.
Frankel, R., Jennings, J., Lee, J., 2018. How well do machines mimic readers? comparing analyst revisions to machine-based revisions generated from conference-call narratives. Working paper.
García-Marco, T., Robles-Fernández, M.D., 2008. Risk-taking behaviour and ownership in the banking industry: the Spanish evidence. J. Econ. Bus. 60 (4), 332–354.
Ghosh, A., 2015. Banking-industry specific and regional economic determinants of non-performing loans: Evidence from US states. J. Financ. Stab. 20, 93–104.
Goddard, J., Molyneux, P., Wilson, J.O.S., Tavakoli, M., 2007. European banking: an overview. J. Bank. Financ. 31 (7), 1911–1935.
Gomes, A., Gorton, G., Madureira, L., 2007. SEC regulation fair disclosure, information, and the cost of capital. J. Corp. Financ.
Grassa, A., Moument, N., Hussainey, K., 2020. Is bank creditworthiness associated with risk disclosure behavior? evidence from Islamic and conventional banks in emerging countries. Pac. Bas. Financ. J. pp. 101327.
Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. R., Tatham, R. L., 1998. Multivariate data analysus. Upper Saddle River, NJ: Prentice hall. 5 (3), 207-219.
Hakkio, S., Keeton, R., 2009. Financial stress: what is it, how can it be measured, and why does it matter? Econ. Rev. 94 (2), 5-50.
Horton, J., Serafeim, G., Serafeim, I., 2013. Does mandatory IFRS adoption improve the information environment? Contemp. Account. Res. 30 (1), 388–423.
Hox, J.J., 2010. Multilevel analysis: Techniques and applications. Routledge. Second edition.
Hu, J.L., Li, Y., Chiu, Y.H., 2004. Ownership and nonperforming loans: evidence from Taiwan's banks. Dev. Econ. 42 (3), 405–420.
Huang, K.W., Li, Z., 2011. A multilabel text classification algorithm for labeling risk factors in SEC form 10-K. ACM Trans. Manag. Inf. Syst. 2 (3), 1-19.
Jaccard, J., Turrisi, R, Jaccard, J., 2003. Interaction effects in multiple regression. Sage. 72.
Jankowitsch, R., Pichler, S., Schwaiger, W.S.A., 2007. Modelling the economic value of credit rating systems. J. Bank. Financ. 31 (1), 181–198.
Jiménez, G., Saurina, J., 2006. Credit cycles, credit risk, and prudential regulation. Int. J. Cent. Bank.
Jones, M., Melis, A., Gaia, S., Aresu, S., 2018. Does graphical reporting improve risk disclosure? Evidence from European banks. J. Appl. Account. Res. 19 (1), 161–180.
Kalemli-Ozcan, S., Sorensen, B., Yesiltas, S., 2012. Leverage across firms, banks, and countries. J. Int. Econ. 88 (2), 284–298.
Kim, J.S., Frees, E.W., 2007. Multilevel modeling with correlated effects. Psychometrika. 72 (4), 505
Kuang, Y.F., Qin, B., 2013. Credit ratings and CEO risk-taking incentives. Contemp. Account. Res. 30 (4), 1524–1559.
Li, F., 2010. The information content of forward- looking statements in corporate filings: a naïve bayesian machine learning approach. J. Account. Res. 48 (5), 1049–1102.
Li, F., 2008. Annual report readability, current earnings, and earnings persistence. J. Account. Econ. 45 (2–3), 221–247.
Li, J., Shin, Y.S., Moore, W.T., 2006. Reactions of Japanese markets to changes in credit ratings by global and local agencies. J. Bank. Financ. 30 (3), 1007–1021.
Linsley, P.M., Shrives, P.J., 2005. Transparency and the disclosure of risk information in the banking sector. J. Financ. Regul. Compliance 13 (3), 205–214.
Loughran, T., McDonald, B., 2014. Measuring readability in financial disclosures. J. Financ. 69 (4), 1643–1671.
Loughran, T., McDonald, B., 2011. When is a liability not a liability? textual analysis, dictionaries, and 10-Ks. J. Financ. 66 (1), 35–65.

Louzis, D.P., Vouldis, A.T., Metaxas, V.L., 2012. Macroeconomic and bank-specific determinants of non-performing loans in Greece: A comparative study of mortgage, business and consumer loan portfolios. J. Bank. Financ. 36 (4), 1012–1027.

Marcato, G., Milcheva, S., Zheng, C., 2018. Market integration, country institutions and IPO underpricing. J. Corp. Financ. 53, 87-105.

Maffei, M., Aria, M., Fiondella, C., Spanò, R., Zagaria, C., 2014. (Un)useful risk disclosure: explanations from the Italian banks. Manag. Audit. J. 29 (7), 621–648.

Magnan, M., Markarian, G., 2011. Accounting, governance and the crisis: Is risk the missing link? Eur. Account. Rev. 20 (2), 215–231.

Mayew, W.J., Sethuraman, M., Venkatachalam, M., 2015. MD&A disclosure and the firm's ability to continue as a going concern. Account. Rev. 90 (4), 1621–1651.

Miihkinen, A., 2013. The usefulness of firm risk disclosures under different firm riskiness, investor-interest, and market conditions: new evidence from Finland. Adv. Account. 29 (2), 312–331.

Muslu, V., Radhakrishnan, S., Subramanyam, K.R., Lim, D., 2015. Forward-looking MD&A disclosures and the information environment. Manage. Sci. 61 (5), 931–948.

Nguyen, M., Skully, M., Perera, S., 2012. Market power, revenue diversification and bank stability: Evidence from selected South Asian countries. J. Int. Financ. Mark. Institutions Money. 22 (4), 897-912.

Ohlson, J.A., 1980. Financial ratios and the probabilistic prediction of bankruptcy. J. Account. Res. 18 (1), 109.

Oliveira, J., Rodrigues, L.L., Craig, R., 2011. Risk-related disclosure practices in the annual reports of Portuguese credit institutions: an exploratory study. J. Bank. Regul. 12 (2), 100–118.

Park, J., 2012. Corruption, soundness of the banking sector, and economic growth: A cross-country study. J. Int. Money Financ. 31 (5), 907–929.

Price, R., Román, F.J., Rountree, B., 2011. The impact of governance reform on performance and transparency. J. Financ. Econ. 99, 76-96.

Price, S.M.K., Doran, J.S., Peterson, D.R., Bliss, B.A., 2012. Earnings conference calls and stock returns: the incremental informativeness of textual tone. J. Bank. Financ. 36 (4), 992–1011.

Rajan, R.G., 1994. Why bank credit policies fluctuate: a theory and some evidence. Q. J. Econ. 109 (2), 399–441.

Roberts, M.R., Whited, T.M., 2013. Endogeneity in Empirical Corporate Finance, in: Handbook of the Economics of Finance. 493-572.

Robinson, T.J., 2003. Multilevel analysis: techniques and applications. J. Am. Stat. Assoc. 98 (462), 496–496.

Salas, V., Saurina, J., 2002. Credit risk in two institutional regimes: Spanish commercial and savings banks. J. Financ. Serv. Res. 22 (3), 203-224.

Stiroh, K.J., Metli, C., 2003. Now and then: The evolution of loan quality for US banks. Current Issues in Econs and Financ. 9 (4), 1-7.

Shorter, G., Seitzinger, M. V., 2012. Credit rating agencies and their regulation, in: Credit Rating Agencies: Regulation and Reform Act Review. pp. 1–32.

Stern, G.H., Feldman, R.J., 2004. Too big to fail: The hazards of bank bailouts. Brookings Institution Press.

To, T.Y., Treepongkaruna, S., Wu, E., 2018. Are all insiders on the inside? Evidence from the initiation of CDS trading and short selling in the financial sector. J. Int. Financ. Mark. Institutions Money. 54, 114-129.

Wei, L., Li, G., Li, J., Zhu, X., 2019. Bank risk aggregation with forward-looking textual risk disclosures. North Am. J. Econ. Financ. 50, pp. 101016.

Williams, J., 2004. Determining management behaviour in European banking. J. Bank. Financ. 28 (10), 2427–2460.

Wu, Y., Bowe, M., 2012. Information disclosure and depositor discipline in the Chinese banking sector. J. Int. Financ. Mark. Institutions Money. 22 (4), 855-878.

Zhang, D., Cai, J., Dickinson, D.G., Kutan, A.M., 2016. Non-performing loans, moral hazard and regulation of the Chinese commercial banking system. J. Bank. Financ. 63, 48–60.

Table 1
Sample extraction and country distribution

Panel A: Sample extraction using Thomson One

Extraction procedures	Firms	Remaining
Firms must be: (a) public (b) active (c) non-ADR	53,545	53,454
Firms must be part of the European Union, not only belonging to European community	5,613	5,613
Firms listed on a country's stock exchange	4,853	4,853
Firms must be classified under Europe's SIC ("602" and "603")	214	214
Firms without complete annual reports and data are excluded	(69)	145
Total number of firm-year observations from 2005 to 2017 (145 x 13)		**1885**

Panel B: Country distribution

Country	N	Firms	%
Austria	91	7	4.83
Belgium	39	3	2.07
Croatia	104	8	5.52
Denmark	195	15	10.34
Finland	26	2	1.40
France	169	13	9.00
Germany	117	9	6.21
Greece	52	4	2.76
Ireland	26	2	1.38
Italy	208	16	11.03
Netherlands	78	6	4.14
Norway	247	19	13.10
Poland	65	5	3.49
Portugal	13	1	0.69
Romania	26	2	1.38
Spain	52	4	2.76
Sweden	52	4	2.76
Switzerland	195	15	10.34
United Kingdom	130	10	6.90
Total	1885	145	100

This table reports the country distribution based on a sample of 145 banks from 19 countries within the EU communities from the period of 2005 to 2017.

Table 2
Descriptive statistics

Statistics	N	Mean	25%	Median	75%	Std Dev
Dependent variables:						
NPL	1885	5.72	1.54	4.41	9.17	4.78
OSCORE	1885	0.35	0.26	0.16	0.40	0.15
ZSCORE	1885	0.29	0.04	0.16	0.47	0.29
DOWN	1885	0.46	0.00	0.00	1.00	0.49
Bank-level variables:						
NPL_TXT	1885	0.07	-1.11	-0.38	0.98	1.60
DIV_INC	1885	0.39	0.19	0.36	0.52	0.27
LEV	1885	0.16	0.10	0.14	0.19	0.09
SIZE	1885	20.91	18.10	20.72	24.43	4.55
ROE	1885	0.10	0.05	0.10	0.15	0.11
Country-level variables:						
FIN_STR	1885	0.12	0.07	0.09	0.17	0.08
CI	1885	0.55	0.32	0.72	0.80	0.34
GOV_EFF	1885	0.15	0.10	0.17	0.19	0.06
GDP	1885	0.06	0.01	0.02	0.05	0.09
INF	1885	0.15	0.07	0.13	0.22	0.10

The table presents the descriptive statistics of the variables used in the full sample analysis. It reports the main dependent variables, bank-level variables and country level variables All continuous variables are winsorised at 1st and 99th percentiles. The description of all variables is presented in **Appendix A**.

Table 3
Correlation matrix

	NPL	NPL_TXT	OSCORE	ZSCORE	DOWN	DIV_INC	LEV	SIZE	ROE	FIN_STR	CI	GOV_EFF	GDP	INF
NPL		**0.39**	**0.11**	0.10	**-0.08**	**-0.06**	**0.11**	-0.04	0.02	-0.02	**0.16**	**-0.08**	**0.23**	**-0.10**
NPL_TXT	**0.39**		**0.41**	**0.24**	**-0.04**	**-0.36**	**0.14**	**0.50**	**0.47**	**-0.05**	**0.14**	**-0.21**	**0.43**	**-0.42**
OSCORE	**0.11**	**0.41**		0.03	-0.01	**-0.14**	-0.01	**-0.23**	**0.10**	0.02	0.07	**-0.05**	**0.23**	0.01
ZSCORE	0.06	**0.24**	0.03		-0.01	0.10	-0.02	**-0.16**	0.04	0.01	-0.03	-0.01	**0.16**	**-0.07**
DOWN	**-0.08**	**-0.04**	-0.01	-0.01		0.17	0.01	-0.02	-0.02	0.04	**-0.01**	0.03	0.01	-0.02
DIV_INC	**0.11**	**0.14**	-0.01	-0.02	0.01		**-0.01**	0.13	**0.09**	0.08	-0.07	0.07	-0.06	0.09
LEV	**-0.08**	**-0.36**	**-0.14**	0.10	**0.00**	**-0.10**		**0.44**	**-0.24**	**0.12**	**-0.12**	0.06	-0.03	0.03
SIZE	0.11	-0.50	**-0.23**	**-0.16**	-0.02	**0.44**	**0.05**		**-0.17**	**0.04**	0.06	0.04	-0.19	**0.17**
ROE	-0.04	0.24	0.03	-0.06	-0.05	**-0.06**	**0.09**	-0.04		**-0.13**	**-0.08**	-0.13	0.01	-0.01
FIN_STR	0.02	-0.05	**0.02**	0.00	0.00	**0.12**	**-0.18**	0.04	**-0.13**		**-0.04**	-0.06	0.02	**-0.04**
CI	**0.16**	**0.14**	**0.07**	-0.03	-0.01	-0.12	0.08	**0.02**	**-0.08**	**-0.04**		-0.40	**0.11**	**-0.10**
GOV_EFF	**-0.08**	-0.21	-0.05	-0.01	**0.03**	0.06	-0.07	**0.04**	**-0.13**	**-0.06**	**-0.40**		**-0.17**	**0.11**
GDP	**0.23**	**0.48**	**0.23**	**0.16**	0.01	**-0.03**	**0.07**	**-0.19**	0.02	0.02	**0.11**	**-0.17**		**-0.35**
INF	**-0.10**	**-0.42**	0.01	**-0.07**	-0.02	0.03	**-0.06**	0.17	-0.01	**-0.04**	**-0.10**	**0.11**	**-0.35**	

Table 3 presents the correlation analysis for all regression variables. The numbers above the diagonal are the linear Pearson coefficients; the numbers below the diagonal are the Spearman coefficients, significant coefficients are highlighted in bold. The description of all variables is presented in **Appendix A.**

Table 4
Repeated measure multilevel analysis: NPL

Dependent variable: NPL					
	Ex. Sig	Model 1	Model 2	Model 3	Model 4
Intercept	(?)	5.049***	4.952***	4.590***	6.164***
		(0.000)	(0.000)	(0.000)	(0.000)
Firm variables					
NPL_TXT	(+)		1.135***	1.656***	2.738***
			(0.000)	(0.000)	(0.000)
LEV	(+)			4.126***	8.317***
				(0.002)	(0.000)
DIV_INC	(?)			0.667*	0.597
				(0.063)	(0.161)
SIZE	(?)			0.182***	0.260***
				(0.000)	(0.000)
ROE	(?)			-6.740**	-9.868***
				(0.032)	(0.000)
Country variables					
FIN_STR	(+)				-1.156***
					(0.000)
CI	(+)				1.199**
					(0.025)
GOV_EFF	(-)				4.645
					(0.319)
GDP	(?)				-19.067***
					(0.000)
INF	(?)				5.708***
					(0.000)
Time		0.123***	0.118***	0.069*	0.009
		(0.000)	(0.003)	(0.096)	(0.865)
Year fixed effect		No	Yes	Yes	Yes
Intra-class correlation					
Repeated Measures (σ^2)		80%***	92%***	93%***	92%***
		(0.000)	(0.000)	(0.000)	(0.000)
Intercept (τ_{00})		18%***	7%***	6%***	7%***
		(0.000)	(0.000)	(0.000)	(0.017)
Time (τ_{11})		2%***	1%***	1%	1%
		(0.000)	(0.003)	(0.287)	(0.459)
Model-fit statistics					
Adjusted R^2			21%	29%	35%
AIC		11009.91	10812.43	10706.28	11009.91
BIC		11043.15	10851.21	10767.19	11043.15
Δ -2 Log Likelihood			199.48	331.631	478.736
Δ Chi-square			(0.000)	(0.000)	(0.000)
Observations		1885	1885	1885	1885

This table reports the repeated measures multilevel analysis with fixed effects. The sample comprises 145 EU banks from the period of 2005 to 2017. Model 1 is the null model, which does not include any predictor variable. Model 2 shows the estimate of NPL_TXT on NPL. Model 3 shows the estimate of other bank-level determinants on credit risk (NPL). Model 4 augments country variables. The dependent variable is NPL which is defined as the sum of non-accrual, reduced, renegotiated past-due loans minus assets acquired in foreclosures scaled by total assets. The table also report ICC (intra-class correlation) which shows the proportion of variance at each level by dividing each level's variation by the total variation. Level 1 gives the variance (repeated measures) within banks over the years under consideration whereas level report the variance between banks (either on the intercept and/or time). Level 1 variance is calculated as (σ^2 at level 1 / (σ^2 at level 1 + σ^2 at level 2)). Adjusted R^2 is computed as 1 – [(1 - R^2) * n-1/ (n – k - l)] where R^2 is computed as ($\sigma^2 m0$ - $\sigma^2 m1$)/$\sigma^2 m0$. Hence, *m*1 is current model's variance component, whereas *m*0 is null model's variance component. *k* is total number of parameters; *n* is total sample size. Year fixed effects are included in model 2 and 3 and p-values are reported in parentheses. Model fit statistics together with the number of bank-year observations are reported at the bottom of the table. The change in -2Log Likelihood (Δ -2LL) is used to examine the improvement of each model in comparison with the null model, whereas change in chi-square is used to examine such improvement statistically. The description of all other variables is presented in **Appendix A.**

Table 5
Repeated measure multilevel analysis: OSCORE

Dependent variable: OSCORE					
	Ex. Sig	Model 1	Model 2	Model 3	Model 4
Intercept	(?)	0.336***	0.330***	0.321***	0.308***
		(0.000)	(0.000)	(0.000)	(0.000)
Firm variables					
NPL_TXT	(+)		0.044***	0.047***	0.079***
			(0.000)	(0.000)	(0.000)
LEV	(+)			0.056	0.183***
				(0.177)	(0.000)
DIV_INC	(?)			-0.037***	-0.039***
				(0.001)	(0.000)
SIZE	(?)			0.000	0.001
				(0.756)	(0.168)
ROE	(?)			-0.105***	-0.208***
				(0.001)	(0.000)
Country variables					
FIN_STR	(+)				0.032***
					(0.002)
CI	(+)				-0.013
					(0.441)
GOV_EFF	(-)				-0.004
					(0.968)
GDP	(?)				-0.383***
					(0.000)
INF	(?)				0.335***
					(0.000)
Time		0.002	0.002	0.002	0.002
		(0.109)	(0.132)	(0.152)	(0.332)
Year fixed effect		No	Yes	Yes	Yes
Intra-class correlation					
Repeated Measures (σ^2)		91%***	67%***	94%***	99%***
		(0.000)	(0.000)	(0.000)	(0.000)
Intercept (τ_{00})		9%***	33%***	6%***	1%**
		(0.000)	(0.000)	(0.000)	(0.046)
Time (τ_{11})		0%	0%	0%	0%
		(0.109)	(0.132)	(0.243)	(0.332)
Model-fit statistics					
Adjusted R^2			87%	23%	30%
AIC		-1896.990	-2202.634	-2190.394	-2316.478
BIC		-1863.746	-2163.853	-2129.476	-2227.913
Δ -2 Log Likelihood			307.643	303.404	439.488
Δ Chi-square			(0.000)	(0.000)	(0.000)
Observations		1885	1885	1885	1885

This table reports the repeated measures multilevel analysis with fixed effects. The sample comprises 145 EU banks from the period of 2005 to 2017. Model 1 is the null model, which does not include any predictor variable. Model 2 shows the estimate of NPL_TXT on OSCORE. Model 3 shows the estimate of other bank-level determinants on credit risk (OSCORE). Model 4 augments country variables. The dependent variable is OSCORE which is defined following Ohlson's (1980) O-score formulae. The table also report ICC (the intra-class correlation) which shows the proportion of variance at each level by dividing each level's variation by the total variation. Level 1 gives the variance (repeated measures) within banks over the years under consideration whereas level report the variance between banks (either on the intercept and/or time). Level 1 variance is calculated as (σ^2 at level 1 / (σ^2 at level 1 + σ^2 at level 2)). Adjusted R^2 is computed as 1 – [(1 - R^2) * n-1/ (n – k - l)] where R^2 is computed as ($\sigma^2 m0$ - $\sigma^2 m1$)/$\sigma^2 m0$. Hence, *m*1 is current model's variance component, whereas *m*0 is null model's variance component. *k* is total number of parameters; *n* is total sample size. Year fixed effects are included in model 2 and 3 and p-values are reported in parentheses. Model fit statistics together with the number of bank-year observations are reported at the bottom of the table. The change in -2Log Likelihood (Δ -2LL) is used to examine the improvement of each model in comparison with the null model, whereas change in chi-square is used to examine such improvement statistically. The description of all other variables is presented in **Appendix A**

Table 6
Repeated measure multilevel analysis: ZSCORE

Dependent variable: ZSCORE					
	Ex. Sig	Model 1	Model 2	Model 3	Model 4
Intercept	(?)	0.308***	0.302***	0.237***	0.206***
		(0.000)	(0.000)	(0.000)	(0.000)
Firm variables					
NPL_TXT	(+)		0.047***	0.056***	0.091***
			(0.000)	(0.000)	(0.000)
LEV	(+)			0.398***	0.526***
				(0.000)	(0.000)
DIV_INC	(?)			-0.032	-0.039*
				(0.179)	(0.092)
SIZE	(?)			-0.005***	-0.003
				(0.008)	(0.207)
ROE	(?)			-0.332***	-0.449***
				(0.000)	(0.000)
Country variables					
FIN_STR	(+)				0.068***
					(0.003)
CI	(+)				-0.036
					(0.311)
GOV_EFF	(-)				-0.004
					(0.987)
GDP	(?)				-0.501***
					(0.000)
INF	(?)				0.124*
					(0.078)
Time		-0.000	-0.000	-0.000	0.000
		(0.885)	(0.863)	(0.869)	(0.823)
Year fixed effect		No	Yes	Yes	Yes
Intra-class correlation					
Repeated Measures (σ^2)		84%***	87%***	89%***	90%***
		(0.000)	(0.000)	(0.000)	(0.000)
Intercept (τ_{00})		15%***	12%***	11%***	10%***
		(0.000)	(0.000)	(0.000)	(0.000)
Time (τ_{11})		1%	1%	0%	0%
		(0.885)	(0.863)	(0.878)	(0.979)
Model-fit statistics					
Adjusted R^2			6%	11%	13%
AIC		654.575	576.158	558.788	550.611
BIC		687.819	614.938	625.237	639.175
Δ -2 Log Likelihood			80.417	107.787	123.964
Δ Chi-square			(0.000)	(0.000)	(0.000)
Observations		1885	1885	1885	1885

This table reports the repeated measures multilevel analysis with fixed effects. The sample comprises 145 EU banks from the period of 2005 to 2017. Model 1 is the null model, which does not include any predictor variable. Model 2 shows the estimate of NPL_TXT on ZSCORE. Model 3 shows the estimate of other bank-level determinants on credit risk (ZSCORE). Model 4 augments country variables. The dependent variable is ZSCORE which is defined following Altman's Z-score formula. The table also report ICC (the intra-class correlation) which shows the proportion of variance at each level by dividing each level's variation by the total variation. Level 1 gives the variance (repeated measures) within banks over the years under consideration whereas level report the variance between banks (either on the intercept and/or time). Level 1 variance is calculated as (σ^2 at level 1 / (σ^2 at level 1 + σ^2 at level 2)). Adjusted R^2 is computed as $1 – [(1 - R^2) * n-1/ (n – k - l)]$ where R^2 is computed as $(\sigma^2 m0 - \sigma^2 m1)/\sigma^2 m0$. Hence, *m*1 is current model's variance component, whereas *m*0 is null model's variance component. *k* is total number of parameters; *n* is total sample size. Year fixed effects are included in model 2 and 3 and p-values are reported in parentheses. Model fit statistics together with the number of bank-year observations are reported at the bottom of the table. The change in -2Log Likelihood (Δ -2LL) is used to examine the improvement of each model in comparison with the null model, whereas change in chi-square is used to examine such improvement statistically. The description of all other variables is presented in **Appendix A.**

Table 7
Repeated measure multilevel analysis: Credit rating downgrade

Dependent variable: $DOWN_{t+1}$				
	Ex. Sig	Model 1	Model 2	Model 3
Intercept	(?)	-0.038**	0.029***	0.088**
		(0.012)	(0.000)	(0.000)
Bank-level variables				
NPL_TXT	(+)	4.218***	4.480***	6.208***
		(0.000)	(0.000)	(0.000)
LEV	(+)		2.302***	3.020
			(0.000)	(0.103)
DIV_INC	(?)		-3.384**	-6.370***
			(0.039)	(0.004)
SIZE	(?)		-0.157**	-0.154**
			(0.031)	(0.042)
ROE	(?)		-4.806***	-8.152***
			(0.000)	(0.000)
Country-level variables				
FIN_STR	(+)			-4.416***
				(0.007)
CI	(+)			-1.247***
				(0.005)
GOV_EFF	(-)			-1.832
				(0.565)
GDP	(?)			-21.413***
				(0.000)
INF	(?)			-14.667***
				(0.000)
Model-fit statistics				
Pseudo R^2		5.9%	6.5%	7.6%
AIC		972.7	868.8	601.6
BIC		994.9	919.7	679.2
-2 Log Likelihood		482.3	425.9	286.8
Observations		1885	1885	1885

This table reports the logit multilevel analysis with fixed effects. The sample comprises 145 EU banks from the period of 2005 to 2017. Model 1 reports the relationship of NPL_TXT with rating downgrade. Model 2 shows the estimate of bank-level determinants on rating downgrade. Model 3 augment country variables. The dependent variable is future credit rating downgrade which is coded as a dummy variable equal to one when a bank receives rating downgrade from S&P credit ratings or zero for otherwise. Year fixed effects are included in models 2 and 3 and p-values are reported in parentheses. Model fit statistics together with the number of bank-year observations are reported at the bottom of the table. The description of all other variables is presented in **Appendix A.**

Table 8
Cross-level interaction

Dependent variable: NPL					
	Ex. Sig	Model 1	Model 2	Model 3	Model 4
Intercept	(?)	5.049***	6.696***	5.703***	7.267***
		(0.000)	(0.000)	(0.000)	(0.000)
Bank-level variables					
NPL_TXT	(+)		2.871***	3.102***	1.945***
			(0.000)	(0.000)	(0.000)
LEV	(+)		7.655***	5.896***	2.786**
			(0.000)	(0.000)	(0.031)
DIV_INC	(?)		0.470	0.514	0.612*
			(0.158)	(0.131)	(0.082)
SIZE	(?)		0.354***	0.320***	0.220***
			(0.000)	(0.000)	(0.000)
ROE	(?)		-4.837***	-4.365***	-2.534**
			(0.000)	(0.000)	(0.024)
Country-level variables					
FIN_STR	(+)		-0.554*		
			(0.088)		
CI	(+)		1.241**		
			(0.012)		
GOV_EFF	(-)		3.194		
			(0.481)		
GDP	(?)		-21.994***		
			(0.000)		
INF	(?)		8.601***		
			(0.000)		
FACTOR	(?)			-0.290**	
				(0.027)	
AGGREGATE	(?)				-0.689***
					(0.000)
Cross-level Interaction					
FIN_STRESS*NPL_TXT	(+)		1.278***		
			(0.000)		
CI*NPL_TXT	(+)		-0.104		
			(0.621)		
GOV_EFF*NPL_TXT	(-)		0.394		
			(0.735)		
GDP*NPL_TXT	(-)		0.291		
			(0.750)		
INF*NPL_TXT	(+)		1.831**		
			(0.036)		
FACTOR*NPL_TXT	(?)			-0.447***	
				(0.000)	
AGGREGATE*NPL_TXT	(?)				0.008***
					(0.005)
Time			0.012	0.042*	0.074*
			(0.385)	(0.051)	(0.070)
Year fixed effect			Yes	Yes	Yes
Intra-class correlation					
Repeated Measures (σ^2)		80%***	91%***	86%***	95%***
		(0.000)	(0.000)	(0.000)	(0.000)
Intercept (τ_{00})		19%***	8%***	13%***	4%***
		(0.000)	(0.000)	(0.000)	(0.000)
Time (τ_{11})		1%*	1%	1%*	1%*
		(0.085)	(0.271)	(0.081)	(0.074)
Model-fit statistics					
Adjusted R^2			37%	41%	41%
AIC		11009.91	10403.04	10444.95	10643.52
BIC		11043.15	10524.75	10539.05	10721.03
Δ -2 Log Likelihood			51.911	216.165	201.633
Δ Chi-square			(0.000)	(0.000)	(0.000)

Observations	1885	1885	1885	1885

This table reports the repeated measures multilevel analysis with fixed effects. The sample comprises 145 EU banks from the period of 2005 to 2017. Model 1 is the null model without any predictors. Model 2report the cross-level interaction effects with individual country-level variables. Model 3 report the estimate using the Principal Component Factor of all country-level variables with the highest loadings (FACTOR). Model 4 estimate the same model with Country-Aggregate (AGGREGATE). The dependent variable is NPL which is defined as the sum of non-accrual, reduced, renegotiated past-due loans minus assets acquired in foreclosures scaled by total assets. The table also report ICC (the intra-class correlation) which shows the proportion of variance at each level by dividing each level's variation by the total variation. Level 1 gives the variance (repeated measures) within banks over the years under consideration whereas level report the variance between banks (either on the intercept and/or time). Level 1 variance is calculated as (σ^2 at level 1 / (σ^2 at level 1 + σ^2 at level 2)). Adjusted R^2 is computed as 1 – [(1 - R^2) * n-1/ (n – k - l)] where R^2 is computed as $(\sigma^2 m0 - \sigma^2 m1)/\sigma^2 m0$. Hence, *m*1 is current model's variance component, whereas *m*0 is null model's variance component. *k* is total number of parameters; *n* is total sample size. Year fixed effects are included in model 2 and 3 and p-values are reported in parentheses. Model fit statistics together with the number of bank-year observations are reported at the bottom of the table. The change in -2Log Likelihood (Δ -2LL) is used to examine the improvement of each model in comparison with the null model, whereas change in chi-square is used to examine such improvement statistically. The description of all other variables is presented in **Appendix A.**

Table 9

Further analysis: OLS regression of the predictive power of text-based credit risk

		Panel A: NPL			Panel B: OSCORE			Panel C: ZSCORE		
Dependent variable:	Ex. Sig	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Intercept	(?)	5.721***	1.241**	1.725**	0.348***	0.382***	0.272***	0.287***	0.370***	0.351***
		(0.000)	(0.019)	(0.009)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Bank-level variables										
NPL_TXT	(+)	1.172***	1.867***	3.358***	0.038***	0.043***	0.081***	0.044***	0.052***	0.084***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
LEV	(+)		0.174	6.013***		-0.006	0.128*		0.414***	0.519***
			(0.892)	(0.000)		(0.877)	(0.018)		(0.000)	(0.000)
DIV_INC	(?)		0.662*	0.401		-0.034**	-0.033**		-0.035	-0.041*
			(0.067)	(0.248)		(0.013)	(0.013)		(0.142)	(0.098)
SIZE	(?)		0.254***	0.344***		-0.001	0.008**		-0.005**	-0.007
			(0.000)	(0.000)		(0.873)	(0.034)		(0.031)	(0.459)
ROE	(?)		-1.628	-3.676***		-0.030	-0.078**		-0.328***	-0.414***
			(0.145)	(0.000)		(0.398)	(0.032)		(0.000)	(0.000)
Country-level variables										
FIN_STR	(+)			-1.100			0.053			-0.069
				(0.348)			(0.159)			(0.399)
CI	(+)			1.152***			0.014			-0.059**
				(0.000)			(0.149)			(0.015)
GOV_EFF	(-)			-4.954***			0.126**			-0.010
				(0.003)			(0.018)			(0.927)
GDP	(?)			-22.089***			-0.419***			-0.533***
				(0.000)			(0.000)			(0.000)
INF	(?)			6.527***			0.635***			0.182**
				(0.000)			(0.000)			(0.019)
Model-fit statistics										
Pseudo R^2		15%	23%	31%	17%	18%	25%	6%	9%	10%
Mean VIF		1.13	1.52	1.92	1.00	1.52	1.92	1.00	1.52	1.92
Max VIF		1.00	1.88	5.35	1.00	1.78	5.35	1.00	1.78	5.35
Observations		1885	1885	1885	1885	1885	1885	1885	1885	1885

This table reports the OLS regression output. The sample comprises 145 EU banks from the period of 2005 to 2017. Model 1 is the baseline model which shows the estimate of credit risk and NPL_TXT only. Model 2 shows the estimate of other bank variables on credit risk. Model 3 augments country variables. The dependent variable is NPL, OSCORE, and ZSCORE in Panel A, Panel B, and Panel C respectively. NPL is defined as the sum of non-accrual, reduced, renegotiated past-due loans minus assets acquired in foreclosures scaled by total assets. OSCORE is defined following Ohlson's O-score formula. ZSCORE is defined following Altman's Z-score formula. P-values are reported in parentheses. The description of all other variables is presented in **Appendix A.**

Table 10
Further analysis: Logistic regression of the predictive power of text-based credit risk

Dependent variable: $DOWN_{t+1}$				
	Ex. Sig	Model 1	Model 2	Model 3
Intercept	(?)	-0.315***	0.390	0.478***
		(0.000)	(0.124)	(0.000)
Bank-level variables				
NPL_TXT	(+)	3.558***	-0.095**	5.563***
		(0.000)	(0.016)	(0.000)
LEV	(+)		0.285	-5.493**
			(0.642)	(0.024)
DIV_INC	(?)		0.193	2.806***
			(0.265)	(0.000)
SIZE	(?)		-0.029**	-0.135***
			(0.023)	(0.000)
ROE	(?)		-1.185**	-7.031***
			(0.029)	(0.000)
Country-level variables				
FIN_STR	(+)			-3.315**
				(0.015)
CI	(+)			-1.089***
				(0.002)
GOV_EFF	(-)			-1.249
				(0.512)
GDP	(?)			-14.244***
				(0.000)
INF	(?)			-18.305***
				(0.000)
Model-fit statistics				
Pseudo R^2		5.9%	1.5%	7.5%
AIC		1060	2603	653
Mean VIF		1.24	1.53	1.69
Max VIF		1.47	1.80	2.61
Observations		1885	1885	1885

This table reports the logistic regression analysis with fixed effects. The sample comprises 145 EU banks from the period of 2005 to 2017. Model 1 reports the relationship of NPL_TXT with rating downgrade. Model 2 shows the estimate of bank-level determinants on rating downgrade. Model 3 augments country variables. The dependent variable is future credit rating downgrade which is coded as a dummy variable equal to one when a bank receives rating downgrade from S&P credit ratings or zero for otherwise. Year fixed effects are included in models 2 and 3 and p-values are reported in parentheses. The description of all other variables is presented in **Appendix A.**

Table 11

Robustness checks: Repeated measures multilevel analysis – provision for loan loss

Dependent variable: Provision for loan loss		
	Ex. Sig	Model 1
Intercept	(?)	2.294***
		(0.000)
Bank-level variables		
NPL_TXT	(+)	3.596***
		(0.000)
LEV	(+)	-2.264***
		(0.000)
DIV_INC	(?)	-17.746***
		(0.000)
SIZE	(?)	0.446
		(0.712)
ROE	(?)	-11.887***
		(0.000)
Country-level variables		
FIN_STR	(+)	0.145*
		(0.099)
CI	(+)	-0.078**
		(0.015)
GOV_EFF	(-)	-0.116**
		(0.048)
GDP	(?)	15.001***
		(0.000)
INF	(?)	-0.667***
		(0.000)
Model-fit statistics		
AIC		488.242
BIC		510.381
-2 Log Likelihood		48.242
Observations		1885

This table shows the robustness checks taking provision for loan loss (PROV_LL) as a credit risk alternative measure. PROV_LL is defined as the loan default at the end of the fiscal year multiplied by total outstanding loans. We test the model of random intercept. Year fixed effects are added to the model. The sample comprises 145 EU banks from the period of 2005 to 2017. P-values are reported in parentheses. Model fit statistics together with the number of bank-year observations are reported at the bottom of the table. The description of all other variables is presented in **Appendix A.**

Table 12
Robustness checks: Repeated measures multilevel analysis for sub-samples

		Panel A: Large Sample			Panel B: Small Sample		
	Ex. Sig	NPL	OSCORE	ZSCORE	NPL	OSCORE	ZSCORE
		Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Intercept	(?)	0.629	0.365***	0.232**	4.648***	0.302***	0.314***
		(0.000)	(0.000)	(0.019)	(0.000)	(0.000)	(0.000)
Bank-level variables							
NPL_TXT	(+)	2.870***	0.137***	0.069***	2.654***	0.125***	0.057***
		(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)
LEV	(+)	4.919**	0.166***	0.144	11.809***	0.301***	0.248
		(0.011)	(0.016)	(0.262)	(0.000)	(0.004)	(0.182)
DIV_INC	(?)	-0.604	-0.041**	-0.022	-0.583	-0.049**	-0.034
		(0.129)	(0.035)	(0.335)	(0.424)	(0.057)	(0.495)
SIZE	(?)	0.246***	0.008***	0.002	0.264**	0.014**	-0.132
		(0.000)	(0.000)	(0.649)	(0.003)	(0.012)	(0.339)
ROE	(?)	-6.454***	0.192***	-0.158**	-1.208	-0.073	-0.011*
		(0.000)	(0.025)	(0.013)	(0.552)	(0.337)	(0.072)
Country-level variables							
FIN_STR	(+)	-1.368	0.066	0.096	-3.800**	0.094	0.085
		(0.273)	(0.221)	(0.225)	(0.034)	(0.173)	(0.487)
CI	(+)	0.830**	0.014	0.008	0.552	-0.041	0.024
		(0.018)	(0.303)	(0.746)	(0.487)	(0.178)	(0.662)
GOV_EFF	(-)	2.478	0.078	0.079	0.892	-0.081	0.138
		(0.630)	(0.104)	(0.791)	(0.906)	(0.751)	(0.726)
GDP	(?)	-19.821***	-0.597***	-0.417***	-15.689**	-0.603***	-0.093
		(0.000)	(0.000)	(0.000)	(0.034)	(0.000)	(0.583)
INF	(?)	5.385***	0.478***	0.098*	4.634**	0.506***	0.277**
		(0.000)	(0.000)	(0.077)	(0.016)	(0.000)	(0.026)
Year fixed effect		Yes	Yes	Yes	Yes	Yes	Yes
Intra-class correlation							
Repeated Measures		52%***	71%***	63%***	63%***	78%***	82%***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Intercept		44%***	24%	36%	18%**	18%	17%
		(0.000)	(0.303)	(0.000)	(0.017)	(0.354)	(0.083)
Time		4%	5%	1%	19%***	17%***	1%
		(0.574)	(0.000)	(0.354)	(0.000)	(0.000)	(0.312)
Model-fit statistics							
AIC		6833.298	-1954.380	-728.949	2669.913	-648.570	-96.834
BIC		6858.109	-1929.540	-708.109	2686.827	-631.656	-79.812
-2 Log Likelihood		6825.268	1962.380	736.949	2661.913	656.570	104.834
Observations		1365	1365	1365	520	520	520

This table reports the repeated measures multilevel analysis for subsamples with bank-year observations from the period of 2005 to 2017. The list of samples is based on classification where countries with more than 100 observations are grouped as the large sample and countries with less than 100 observations are grouped as the small sample. We run analysis for both samples. The dependent variables are NPL, OSCORE and ZSCORE. Adjusted R^2 is calculated as $1 – [(1- R^2) n-1/ (n – k - 1.)]$, where R^2 is computed as $1 – [(\sigma^2 m1 + \tau^2 m0)/ (\sigma^2 null + \tau^2 m0)]$. Hence, *m*1 is current model's variance component, whereas *m*0 is null model's variance component. *k* is total number of parameters; *n* is total sample size. Year fixed effects are reported added in both models. P-values are in parentheses. Model fit statistics are only reported for reference, but they cannot be used for comparison since the samples are different. The number of observations is reported at the bottom of the table. The definition of the explanatory variables is described in **Appendix A.**

Table 13
Robustness checks: Endogeneity checks

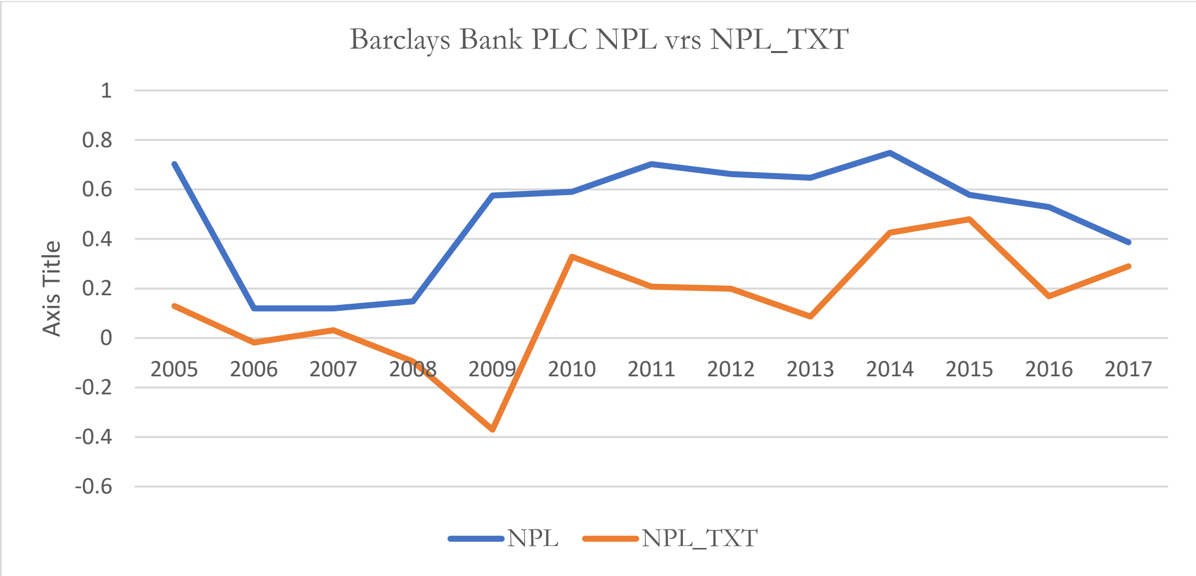
Dependent variable:		Panel A: GMM			Panel B: PSM		
		NPL	OSCORE	ZSCORE	NPL	OSCORE	ZSCORE
	Ex. Sig	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Intercept	(?)	5.902***	0.346***	0.251***	5.779***	0.242***	0.287***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Bank-level variables							
NPL_TXT	(+)	3.342***	0.088***	0.091***	4.975***	0.182***	0.054***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
LEV	(+)	7.243***	0.136***	0.500***	-0.594	0.384***	-0.034
		(0.000)	(0.001)	(0.000)	(0.768)	(0.002)	(0.594)
DIV_INC	(?)	0.553	-0.034***	-0.033	0.926*	-0.041	-0.002
		(0.105)	(0.002)	(0.163)	(0.077)	(0.205)	(0.887)
SIZE	(?)	0.317***	0.002***	-0.002	0.219***	-0.009***	-0.003**
		(0.000)	(0.005)	(0.233)	(0.000)	(0.000)	(0.015)
ROE	(?)	-4.725***	-0.098***	-0.403***	-1.585	-0.116	0.024
		(0.000)	(0.006)	(0.000)	(0.328)	(0.244)	(0.642)
Country-level variables							
FIN_STR	(+)	-1.015	0.055	-0.096	0.188	-0.168	0.176***
		(0.380)	(0.141)	(0.241)	(0.919)	(0.136)	(0.003)
CI	(+)	0.865***	0.0135	-0.025	1.075*	0.007	-0.021
		(0.008)	(0.205)	(0.268)	(0.082)	(0.880)	(0.345)
GOV_EFF	(-)	-0.003	0.068	-0.011	3.550	-0.056	-0.034
		(0.999)	(0.570)	(0.962)	(0.375)	(0.883)	(0.769)
GDP	(?)	-22.73***	-0.410***	-0.507***	-5.682***	-0.298**	0.218***
		(0.000)	(0.000)	(0.000)	(0.004)	(0.013)	(0.000)
INF	(?)	6.019***	0.365***	0.157**	6.442***	-0.036	0.149**
		(0.000)	(0.000)	(0.024)	(0.000)	(0.743)	(0.011)
Year fixed effect		Yes	Yes	Yes	Yes	Yes	Yes
Intra-class correlation							
Repeated Measures (σ^2)		63%***	77%***	79%***	72%***	81%***	93%***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Intercept (τ_{00})		35%***	22%***	20%***	29%***	9%***	6%***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Time (τ_{11})		2%**	1%	1%	1%	0%	1%**
		(0.031)	(0.131)	(0.105)	(0.301)	(0.898)	(0.028)
Model-fit statistics							
Adjusted R^2		32%	19%	27%	58%	11%	76%

		Panel C: Lagged independent variables		
		NPL	OSCORE	ZSCORE
	Ex. Sig	Model 1	Model 2	Model 3
Intercept	(?)	6.224***	0.312***	0.215***
		(0.000)	(0.000)	(0.000)
NPL_TXT$_{T-1}$	(+)	0.292***	0.004	0.025***
		(0.001)	(0.160)	(0.000)
LEV	(+)	-0.207	-0.066	0.212**
		(0.880)	(0.126)	(0.015)
DIV_INC	(?)	1.431***	-0.015	-0.011
		(0.000)	(0.215)	(0.641)
SIZE	(?)	0.045	-0.006***	-0.008***
		(0.120)	(0.000)	(0.000)
ROE	(?)	-1.386	0.039	-0.187***
		(0.168)	(0.211)	(0.003)
Country-level variables				
FIN_STR	(+)	-1.832***	0.008	0.058**
		(0.000)	(0.492)	(0.013)
CI	(+)	1.352**	0.000	-0.025
		(0.016)	(0.995)	(0.457)
GOV_EFF	(-)	-1.094	-0.105	-0.152
		(0.837)	(0.299)	(0.526)
GDP	(?)	4.572***	0.358***	0.212**
		(0.002)	(0.000)	(0.024)
INF	(?)	0.513	0.181***	-0.054
		(0.638)	(0.000)	(0.436)
Year fixed effect		Yes	Yes	Yes
Intra-class correlation				
Repeated Measures		89%***	94%***	87%***
		(0.000)	(0.000)	(0.000)
Intercept		10%***	5%***	12%***
		(0.000)	(0.000)	(0.000)
Time		1%*	1%	1%
		(0.053)	(0.201)	(0.147)
Model-fit statistics				
Adjusted R^2		14%	10%	7%

AIC	9444.95	2033.65	7513.21	5383.374	230.146	-949.587	AIC	10931.20	-1987.29	631.79
BIC	9453.07	2142.66	7566.07	5464.407	312.179	-867.554	BIC	11019.76	-1898.73	720.36
Observations	1885	1885	1885	1885	1885	1885	Observations	1740	1740	1740

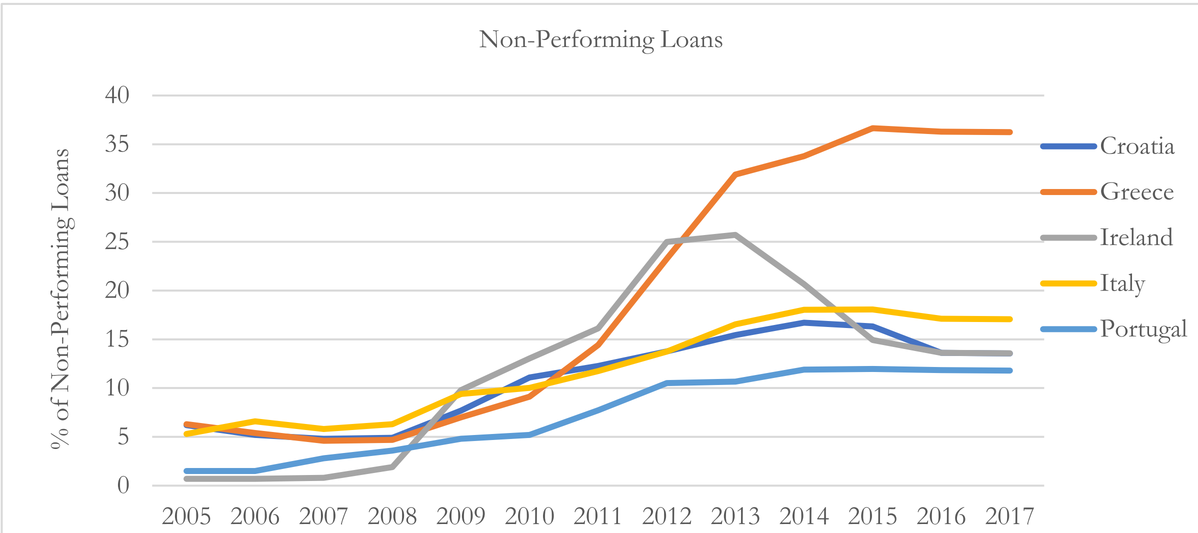
This table reports the results for addressing endogeneity concerns Panel A reports the results using multilevel Generalized Method of Moments (GMM), Panel B reports the results using multilevel Propensity Matching Score (PMS), and Panel C reports the estimate of lag NPL_TXT together with firm and country characteristics. The dependent variables in Model 1 and 2 are NPL, OSCORE, and ZSCORE respectively in Panel A, B, and C. NPL is defined as the sum of non-accrual, reduced, renegotiated past-due loans minus assets acquired in foreclosures scaled by total assets. ZSCORE is defined following Altman's Z-score formula whereas OSCORE is defined following Ohlson's O-score formula. The table also report ICC (the intra-class correlation) which shows the proportion of variance at each level by dividing each level's variation by the total variation. Level 1 gives the variance (repeated measures) within banks over the years under consideration whereas level report the variance between banks (either on the intercept and/or time). Level 1 variance is calculated as (σ^2 at level 1 / (σ^2 at level 1 + σ^2 at level 2)). Adjusted R^2 is computed as $1 – [(1 - R^2) * n-1/ (n – k - l)]$ where R^2 is computed as $(\sigma^2 m0 - \sigma^2 m1)/\sigma^2 m0$. Hence, *m*1 is current model's variance component, whereas *m*0 is null model's variance component. *k* is total number of parameters; *n* is total sample size. Year fixed effects are included in all models and p-values are reported in parentheses. Model fit statistics together with the number of bank-year observations are reported at the bottom of the table. The description of all other variables is presented in **Appendix A**.

Figure 1. Non-performing loans and text-based credit risk for Barclays Bank Plc.



This figure shows the non-performing loans and the credit risk (soft) information for Barclays over the period of the study 2005-2017.

Figure 2. Non-performing loans of sample countries



This figure shows the non-performing loans of five countries from total sample countries. The countries are Croatia, Greece, Ireland, Italy, and Portugal for the period of 2005 to 2017.

Appendix A

Variable descriptions, measures, examples of prior literature, and sources

Variable	Ex.Sig	Definition	Sample paper(s)	Source
Dependent variables: **Credit risk characteristics**				
NPL		((Non-accrual loans + reduced rate loans + renegotiated loans + past due loan at least 90 days) – (assets acquired in foreclosures + repossessed personal property))/total assets	(Berger and DeYoung, 1997; Cucinelli et al., 2018; Ghosh, 2015; Zhang et al., 2016)	Bloomberg
OSCORE		$1.32 - 0.407 \times \log(TA/GNP) + 6.03 * (TL_t / TA_t) - 1.43 * (WC/TA_t) + 0.0757 * (CL_t/CA_t) - 1.72X - 2.37 * (NI_t/TA_t) - 1.83(FFO_t/TL_t) + 0.285 * Y - 0.521 (\Delta NI / NI_t + NI_{t-1})$. Where X is 1 when TL > TA and 0 if otherwise and Y is 1 if net loss for the last two years and 0 when otherwise	(Ohlson, 1980)	authors' computation
ZSCORE		6.56(WC/TA) + 3.26(RE/TA) + 6.72(EBIT/TA) + 1.05(MVE/TL)	(Altman et al., 1998; Donovan et al., 2019; Li et al., 2006)	authors' computation
DOWN		This is a dichotomous variable which takes the value of one if a bank receives ratings downgrade from S&P credit ratings or zero if otherwise	(Donovan et al., 2019; Li et al., 2006)	Bloomberg
Independent variables: **Bank characteristics**				
NPL_TXT	(+)	The out-of-sample sLDA prediction of $NPL_{i,t}$ using the frequency of terms found in the annual reports for bank *i* in year *t*. The training data include all annual reports of banks with reported NPL		sLDA output
LEV	(+)	Calculated as total debt scaled by total assets, representing change in financial leverage	(Donovan et al., 2019; Kalemli et al., 2012)	Thomson One
DIV_INC	(?)	Total non-interest income scaled by total income	(Ghosh, 2015)	Authors' computation
SIZE	(?)	Natural log of bank's total assets at the end of the fiscal year in which annual report was prepared	(Elshandidy and Shrives, 2016; Louzis et al., 2012; Salas and Saurina, 2002)	Authors' computation
ROE	(?)	Profit before tax ((if available, otherwise profit after tax)/worth)) x 100	(Garcia and Robles, 2008)	Thomson One
Independent variables: **Country characteristics**				
FIN_STR	(+)	Country level index of financial stress	(Cardarelli et al., 2011)	ECB
CI	(+)	The original score of Transparent International's Corruption Perception Index (TI index) indicates the perceived level of Prevailing corruption on a scale of 0-10, with higher score suggesting a higher economic and political integrity. We use 10 minus this TI index, rendering CI with a higher score indicating more rampant corruption	(Chen et al., 2015; Park, 2012)	Transparency International and authors' computation
GOV_EFF	(-)	A measure which captures perceptions of the quality of public and civil services, the degree of political independence, and policy formulation and implementation on a scale of -2.5 – 2.5 with higher score indicating more efficient policies	(Barth et al., 2013; Beck et al., 2003)	World Bank
GDP	(?)	A measure of the economic conditions under which a bank operates. It is measured as the real annual growth in GDP	(Ali and Daly, 2010; Boumparis et al., 2019; Louzis et al., 2012; Ghosh, 2015)	World Bank
INF	(?)	Annual rate of inflation measured as the percentage change in consumer price index	(Jankowitsch et al., 2007)	World Bank

Appendix B
Non-performing loan ratios of sample countries

	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Austria	2.12	2.74	2.24	1.90	2.25	2.83	2.71	2.81	2.87	3.47	3.39	2.70	2.37
Belgium	1.28	1.28	1.16	1.65	3.08	2.80	3.30	3.74	4.24	4.18	3.79	3.43	2.92
Croatia	5.00	5.19	4.75	4.87	7.66	11.1	12.3	13.8	15.4	16.7	16.3	13.6	11.2
Denmark	0.21	0.33	0.60	1.20	3.31	4.07	3.66	5.95	4.62	4.40	3.69	3.21	2.29
Finland	0.35	0.23	0.31	0.43	0.67	0.63	0.53	0.57	1.30	1.30	1.34	1.52	1.67
France	3.5	3.02	2.73	2.82	4.02	3.76	4.29	4.29	4.50	4.16	3.98	3.64	3.08
Germany	4.05	3.41	2.65	2.85	3.31	3.20	3.03	2.86	2.70	2.34	1.97	1.71	1.50
Greece	6.31	5.44	4.68	4.67	6.95	9.12	14.4	23.3	31.9	33.8	36.6	36.3	45.6
Ireland	0.48	0.53	0.63	1.92	9.80	13.0	16.1	24.9	25.7	20.6	14.9	13.6	11.5
Italy	7.00	6.57	5.78	6.28	9.45	10.0	11.7	13.8	16.5	18.0	18.1	17.1	14.4
Netherlands	1.25	0.80		1.68	3.20	2.83	2.71	3.10	3.23	2.98	2.71	2.54	2.31
Norway	0.73	0.61	0.53	0.72	1.28	1.52	1.68	1.51	1.34	1.13	1.05	1.18	1.00
Poland				2.82	4.29	4.91	4.66	5.20	4.98	4.82	4.34	4.05	3.94
Portugal	1.56	1.3	2.85	3.60	5.13	5.31	7.47	9.74	10.6	11.9	17.5	17.2	13.3
Romania			2.59	2.75	7.89	11.9	14.3	18.2	21.9	13.9	13.5	9.62	6.41
Spain	0.79	0.70	0.90	2.81	4.12	4.67	6.01	7.48	9.38	8.45	6.16	5.64	4.46
Sweden	0.85	0.10	0.08	0.46	0.83	0.78	0.65	0.70	0.61	1.24	1.17	1.06	1.12
Switzerland	1.29	0.96	0.77	0..95	1.12	0.92	0.84	0.79	0.78	0.72	0.75	0.74	0.64
UK	1.20	0.90	0.91	1.56	3.51	3.95	3.96	3.59	3.11	1.65	1.01	0.94	0.73

Source: World Bank, https://data.worldbank.org/indicator/FB.AST.NPER.Z3?end=2016&locations=XC&most recent year desc=true&start=2005 (Note: some values are missing in the World Bank Tables)

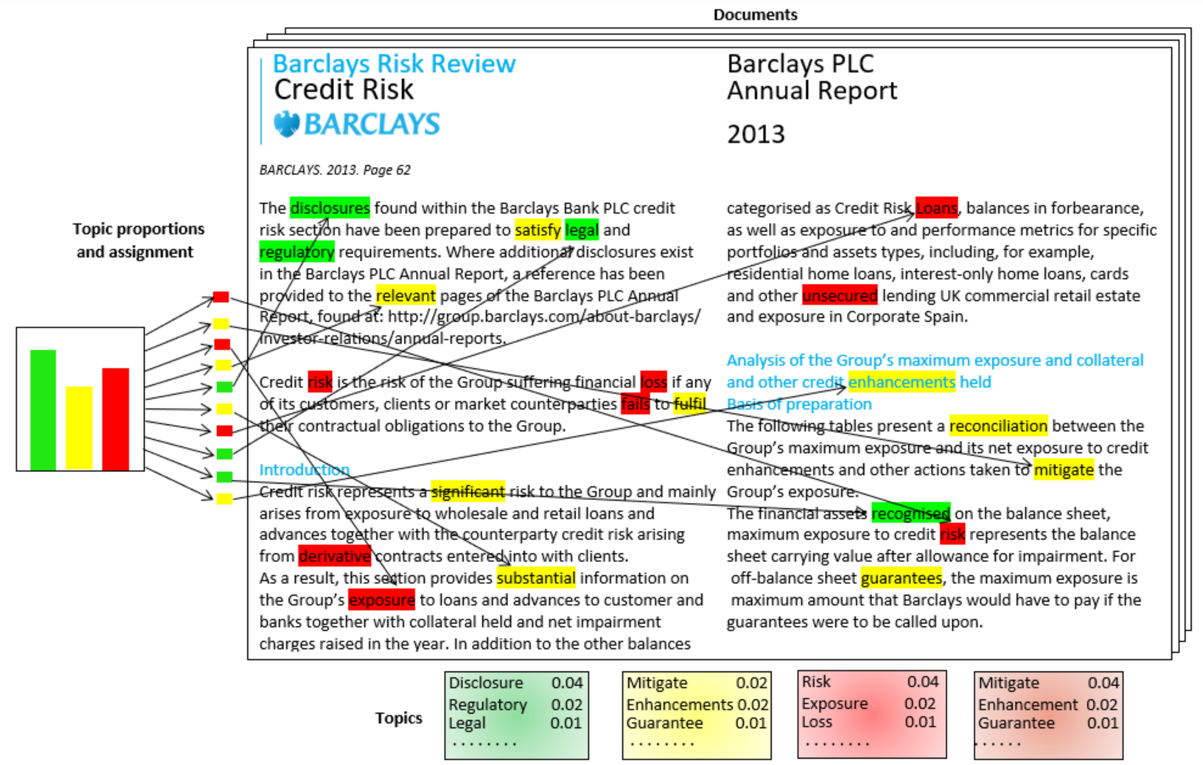
Appendix C
Detailed description of LDA
1. Intuition behind Latent Dirichlet Allocation

LDA is a statistical generative process which takes a random form of modelling over a set of documents. Blei et al. (2007) define LDA as a statistical method that analyses the words of documents to discover trends that run across them. Specifically, LDA demonstrates how these trends are correlated with each other and change over time. LDA has so far been the most widely employed and accepted topic modelling technique due to its great advantages (Aziz et al., 2019). One distinguishing feature of LDA is that no prior knowledge of topics is needed as topics are originated from screening original documents such as annual reports.

Banks' annual reports contain different topics that can be indicated by different relevant words. Figure 3 considers the annual report of Barclays PLC for the year ended 2013 where words pertaining to risk management (i.e., *mitigate, enhancement, guarantee),* risk (i.e., *risk, exposure, loss),* and regulation (i.e., *disclosure, regulatory, legal*) are highlighted in yellow, red, and green respectively. Figure 3 illustrates that, as each word is drawn from each topic, the distribution over topics shows the proportion of the topics – *risk management, risk,* and *regulatory standard.* Thus, this is the unique feature of LDA, where all documents share in the same corpus, but each document exhibits topics with different proportions.

The topic structure entails three stages, which include the *topics, per-document topic distribution,* and *per-document per-word topic assignment.* As the goal of LDA is to automatically identify topics from a collection of documents, it makes the documents observable and the topic structure hidden. It is therefore the computational responsibility of LDA to use the observed documents to infer the hidden topic structure.

Figure 3. The intuition behind LDA



This figure shows the generative process of LDA. It is assumed that the number of topics (K) is set for the whole collection of annual reports. This number of topics is the distribution over words as shown on the far left. A Dirichlet of topic proportion is chosen (i.e., the histogram on the far left) and for each word a topic assignment (coloured words), choses and finally words chosen from the corresponding topic. It must be noted that the topic structure (i.e., the topics, topic assignments and topic distribution) in this figure it's a graphics and not a real generated model.

2. Generative process

LDA is a generative process that forms part of the broad field of probabilistic modelling, where the outcome is seen as arising from hidden variables. The hidden variables ideally are the topic structures, while the observed variables are the collection of documents. Both observed and hidden variables form a joint distribution which then computes the conditional distribution of the hidden variables given the observed variables, which is called posterior distribution. Formally, LDA is explained using the following notation:

Observed variables collection of documents

W_d : observed words in document *d*

$W_{d,n}$: the nth word in document d. This is an element from the fixed vocabulary

Hidden variables (Topic assignment)

Topics:

$\beta_{1:k}$: topics

β_k : distribution over the corpus (vocab)

Topic distribution:

θ_d : topic proportion for *dth* document

$\theta_{d,k}$: topic proportion for topic *K*, in document *d*

Topic assignment:

$Z_{d,}$: topic assignment for the *dth* document

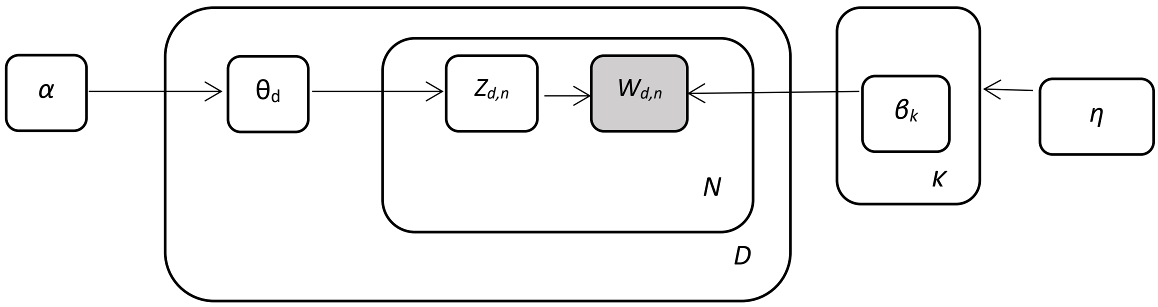
$Z_{d,n}$: topic assignment for the nth word in document *d*

Based on the above notation, the LDA infers the following joint distribution of hidden and observed variables:

$$P(\beta_{1:K,}\theta_{1:D,}z_{1:D,}w_{1:D}) = \prod_{i=1}^{k} P(B_i) \prod_{d=1}^{D} P(\theta_d) \prod_{n=1}^{N} P(z_{d,n}|\theta_d)\, P(w_{d,n}|\beta_{1:K,}z_{d:n})$$

It can further be observed that the distribution exhibits some level of dependencies in relation to: (1) topic assignment $Z_{d,n}$ relies on the per-document topic proportion $\theta_{d;}$ and (2) observed word $W_{d,n}$ depends on topic assignment $Z_{d,n}$ and all the topics $\beta_{1:}$ Identified words are based on correlated topics and probability of the word within that topic. These dependencies explain LDA. The probabilistic graphical model for LDA is presented in Figure 4 below:

Figure 4. LDA graphical model



This figure shows the nodes that explain the generative process of LDA. Each node represents a random variable. Hidden nodes (topics, topic proportion, topic assignment) are unshaded. Observed nodes (words in collection of documents) are shaded. The *D* plate represents the collection of documents, the *N* plate represents the collection of words. The *K* plate represents the predefined topics.

3. Supervised LDA

sLDA is an extension of LDA by adding an additional layer which captures the response variable. LDA certainly will not be a reliable option when the goal is to make predictions. The goal of sLDA is to predict a response variable for unseen documents using a fitted model. In this study the response variable (i.e., NPL) is our main interest In sLDA the response variable and the documents are modelled together to find the hidden topics that best predict the response variable for any unseen documents.

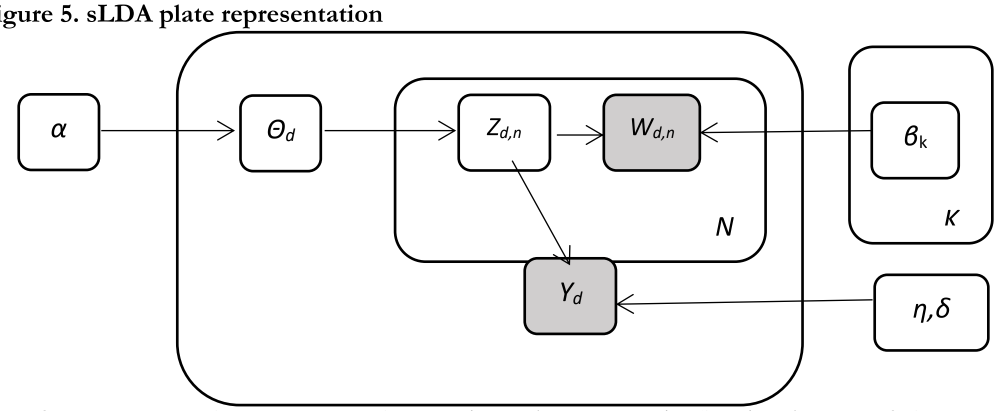
As shown in Figure 5, assume a collection of documents and terms (corpus) which has a fixed number of latent topics. Each document and topic consist of a discrete probability distribution over topics and terms respectively. On this note, a document can be created by repetitively sampling the topic distribution to list a topic. The same process is done on word distribution for the given topic to draw a word. The following assumptions are made for the generative process for each document and response:

1. Draw topic proportions $\theta \mid \alpha \sim$ Dirichlet (α)
2. For every word
    a. Draw topic assignment $Z_n \mid \theta \sim$ *Multinomial* (θ).
    b. Draw term $w_n \mid z_n, \beta_{1:k} \sim$ Multinomial (βz_n)
3. Draw response variable y | $z_{n:N}$, η, $\delta \sim$ GLM [($\bar{y}$), η, δ) where,

    $\bar{y} = (1/N)\, \Sigma^{N}_{n=1}\, z_{n.}$

where: α represents a dirichlet parameter, $\beta_{1:k}$ is a vector of term proportion , η and δ are response parameters.

Figure 5. sLDA plate representation



This figure presents the parameters that needs, to be estimated. The distribution of the response variable is a generalised linear model, which has a random component and a systematic component, with the assumption that it belongs to an exponential family. In contrast to the terms and documents, the responses are treated as non-exchangeable and response is treated depending on the topic frequencies which occur in the generated document.