2014/95/EU Directive impact on the quality of non-financial information

Anastasiia Borisova

Prof. Paul André

Expert Prof. Michael Rockinger

This draft: January 2019

Abstract

We analyse corporate social responsibility (CSR) publications for large European companies ranging 2011-2017. Using Directive 2014/95/EU as an exogenous shock, we assess its impact on CSR disclosure textual quality (presented in annual and stand-alone reports) as formulated in the guidelines accompanying new mandatory legislation. We operationalize key principles of quality reporting using Natural Language Processing metrics such as tone, readability, report length, topic coverage (environmental, social, bribery, human rights, horizon and strategic orientation), boilerplates and cosine similarity. The analysis is carried out with a difference-in-difference (DID) approach. There is a substantial difference in the reporting quality as formulated by the guidelines between the control (UK and Scandinavian countries) and treatment groups (Other EU countries). Our results of DID indicate a significant increase in the quantity of disclosures, however, we also witness the increase in boilerplate, with the latter capturing an uninformative growth of the disclosures. Overall, we conclude that during the transitional period between the adoption of the directive and its implementation, companies do not show signs of improved quality but rather experience a lack of clarity when preparing for the new legislation.

1. Introduction

Starting in 2018 the European directive on the disclosure of non-financial and diversity information requires large European companies to publish non-financial information in their annual or stand-alone reports covering the 2017 financial year. It is an unprecedented regulatory change in terms of one-time transition of several countries to the mandatory reporting of non-financial information.

The directive passed on April 15th, 2014 and entered into force on December 6th, 2014 giving Member States four years to implement the former. Based on country level selection criteria, the approximate amount of companies in scope of the directive amounts to 18,000 large public, private and financial undertakings.

The key objective of the directive is to enhance transparency of non-financial information by enhancing its quantity and quality. In this paper, we focus on the achievement of this objective evaluated through the impact on textual characteristics of CSR disclosures.

Regulatory initiatives such as transitions from voluntary to mandatory reporting should be analysed in two steps (Leuz & Wysocki, 2016). The first stage consists in the analysis of disclosure outcomes stated by the regulatory act while the second stage investigates the economic consequences of the regulatory changes. We follow this research path to analyse the impact of the new Directive 2014/95/EU on the quality of non-financial information. The research question of the current paper represents the first part of the analysis and is formulated as follows: does the new Directive 2014/95/EU improve textual quality of CSR disclosures?

The directive allows companies to rely on several reporting frameworks while providing its own guidelines. We evaluate disclosure outcomes based on the guidelines issued by the EU Commission which determine the principles of high-quality disclosure.

These guidelines were elaborated based on best practice and experience gained through the implementation of a large set of international CSR frameworks.

According to the directive, there is no additional verification required for non-financial information which results in a weakly enforced legal act. The study of the European directive will allow us to cover a broad set of institutional settings and dissect nuanced impacts of poorly enforced legislation. Most studies generally focus on the US and its unique regulatory setting (Jegadeesh & Wu, 2013; Lawrence, 2013; Loughran & McDonald, 2011, 2014; Loughran & McDonald, 2015; Nazari, Hrazdil, & Mahmoudian, 2017). Studies of non-US settings remain quite limited yet richer at many levels.

For the analysis of disclosure outcomes, we rely on Natural Language Processing (NLP) techniques which allow measuring of CSR reporting textual attributes and capture disclosure quality as formulated in the guidelines. We further provide a detailed overview of those principles and explain the choice of the corresponding NLP metrics.

Our appeal to NLP techniques is motivated by several factors. First, CSR disclosures primarily consist of narrative information. Second, European disclosures lack uniformity in their presentation which is often an impediment to non-US studies. NLP techniques make the analysis of a large set of disclosures possible despite their structural inconsistencies.

Third, the use of NLP is also supported in the literature as an appropriate method for the evaluation of disclosure quality. According to Beyer et al. 2010 p.312 “analyzing disclosures using natural language processing techniques seems most promising in creating meaningful disclosure quality measures for large samples.” They also add that “non-quantitative parts of corporate disclosure will likely provide us with a better understanding of management’s disclosure choices and the resulting economic consequences…”

Finally, evaluating textual attributes of CSR disclosures before and after the regulatory change will shed light on the effectiveness of weakly enforced regulation.

Meanwhile, it will open the possibility to investigate the usefulness of textual CSR disclosures for financial users which is assumed by the directive but lacks empirical evidence. Indeed, most studies cover environmental disclosures and only few assess the usefulness of the narratives. The majority focuses on the mere existence of the disclosure or on scores (e.g. Bloomberg CSR Disclosure score), neglecting the impact of what is communicated and how.

On one side, it has been shown that: - the issuance of stand-alone CSR reports is associated with lower analyst forecast errors (Dhaliwal, Radhakrishnan, Tsang, & Yang, 2012) and reduces the cost of equity capital for the best CSR performers (Dhaliwal, Li, Tsang, & Yang, 2011); - carbon emissions disclosures directly affect firm value (Matsumura, Prakash, & Vera-Muñoz, 2014); - environmental disclosures are positively evaluated by investors (Plumlee, Brown, Hayes, & Marshall, 2015) and are value-relevant (Clarkson, Fang, Li, & Richardson, 2013).

On the contrary, environmental narratives of annual reports are shown to be ignored by sell-side bank analysts (Campbell & Slack, 2011) and to be negatively related to CSR performance (Cho & Patten, 2007).

In summary, we believe that this stage of research is necessary not only for the evaluation of regulatory impact but also for further analysis of the usefulness of CSR disclosures for financial users.

In order to assess the effects of an exogenous shock on disclosure, we apply a difference-in-difference design using a set of companies defined according to the selection criteria as determined by each Member State. We classify companies from UK and Scandinavian countries as our control group as these countries have the most similar regulatory regimes to the directive which will be discussed later. Given the combination of lack of enforcement and a relatively long transitional period, we expect that regulatory effects

will start to more clearly manifest in the reporting by 2017 – the year of actual implementation of the directive requirements. We test the impact of the directive before and after 2017 rolling the year of impact back to 2014. The first signs of the directive appear in 2015 and strengthen over the consequent years. There is no indication of quality improvement during this transitional period, which rather results in the increase of complexity in CSR reporting captured by significant increase in the length of reported information in parallel with the increasing use of boilerplates.

2. Institutional setting

The first attempt to regulate CSR reporting in European companies was undertaken in the Account Modernization Directive 2003/51/EC amended to report on environmental and employee affairs. It was followed by the Accounting Directive 2013/34/EU. The implementation of these acts was not supported by any standards or guidelines, hence, resulted in non-financial issues being rarely reported.

As for Member States, only a few European countries introduced mandatory disclosure rules before 2014. Amongst the most prominent adopters of mandatory regimes were Scandinavian countries (from 2008), France (from 2001) and United Kingdom (2006). The rest of the European states regarded CSR as a purely voluntary practice up to the new directive. Non-financial reporting in the beginning of the decade could have been described as being highly dispersed across countries – a quality drawback heavily affecting the usefulness of such information.

Overall, the lack of harmonization in regard to CSR disclosures is a major reason for introducing a mandatory standard at the EU level to improve comparability, transparency and reliability.

The new directive clearly determines the content of disclosed information and imposes a framework with a more structured presentation and specific requirements. The directive determines a benchmark for the transposition of the directive in Member State legislation. According to the directive, organizations must disclose if they cumulatively have:

- an average number of over 500 employees during the financial year
- a balance sheet total exceeding EUR 20 million or a net turnover exceeding EUR 40 million

The minimum covered topics are as follows:

- environmental
- social and employee matters
- respect for human rights
- anti-corruption and bribery matters

Member States are entitled to self-determine the set of undertakings falling under the directive, as well as the penalties for non-compliance and audit requirements. We present a more detailed comparison of the Member States' requirements vs. the benchmark determined by the directive in Table **1**. On the whole, we can see that Member States either maintained the same conditions or apply a stricter framework.

Insert Table **1** here

According to the directive, Member States only impose that auditors verify the existence/presence of such disclosures in annual or stand-alone reports. There is no qualitative requirement as to the content prescribed by the directive. However, several Member States introduced such a requirement by including a content check for directive compliance in the audit procedure. Concerning enforcement, article 33.1 of the directive stipulates that the members of the administrative, management and supervisory bodies have

collective responsibility for the management report. It also ensures that responsibility encompasses stand-alone CSR reports. The consequences of non-compliance differ across countries and vary from purely administrative and/or monetary penalties to imprisonment.

3. Theoretical background and hypothesis development

The outcome of the directive can be twofold. Strategic CSR theories (e.g. signalling theory, resource-based theory, stakeholder theory) provide a rationale in favour of quality improvement while theories seeing CSR as market failure (e.g. agency theory) predict its inefficiency.

Strategic CSR theories offer several channels for an economically rational application of CSR initiatives which motivates companies to go beyond mere legal compliance and boilerplate disclosures. Within this view, companies invest in CSR to increase their reputation and competitive advantage (resource-based theory (Hart, 1995)), improve relationships with the beneficial stakeholders (stakeholder theory (Jones, 1995)) or to signal the ethical nature for the parties willing to initiate or maintain the relationships with the firm and avoid a "lemon problem" (signalling theory (Spence, 1973)).

Signalling theory provides a mechanism justifying the expectation of quality improvement under the directive. Signalling equilibrium ensures that companies signal the "true type" under the two following conditions:

- the returns of signalling true type are greater than its costs
- the costs of falsification are greater than returns (Zerbini, 2017).

If either of the conditions is not satisfied, a company chooses to falsify or to avoid disclosing information. The directive would mostly foster the second condition. Prior to the directive it has been commonly recognized that the quality of the voluntary CSR disclosures is low (Michelon, Pilonato, & Ricceri, 2015). The directive aims to alleviate this problem by

directly affecting the scope of the disclosing companies. The Commission's impact assessment document announces an increase in the number of disclosing companies by about seven-fold (from ~2500 to ~18000). Theoretically, the increase in the number of reporters will improve stakeholders' understanding of the matter and make falsification more easily detectable or, in other terms, "costly". Such expectations would encourage companies to make a step towards a better-quality disclosure. A recent study by Fiechter, et al. (2017) provides us with support for this expectation claiming a significant increase in CSR expenditures by companies falling under the directive. That brings us to the first hypothesis:

Hypothesis 1.a: The quality of CSR disclosures will improve after the adoption of the directive.

Meanwhile theories explaining CSR initiatives as market inefficiencies would not provide a motivation for the improvement of CSR quality under weakly enforced regulation. Namely, agency theory argues in favour of the implementation of controls preventing such investments and being penalized for CSR activities. In this scenario, CSR disclosures will not be informative as they will increase proprietary costs. In other words, if CSR activities are seen as reducing shareholder value, managers will have little incentive to properly disclose CSR information (Gutsche, 2016). From this perspective, there is no reason to expect any improvements after the implementation of the directive.

Alternate hypothesis 1.b: The quality of CSR disclosures will not change after the adoption of the directive.

4. Quality definition and operationalization

The guidelines clarify EU understanding of non-financial reporting quality. It relies on six principles suggesting disclosing information, which are:

1. Material
2. Fair, balanced and understandable
3. Comprehensive but concise
4. Strategic and forward-looking
5. Stakeholder orientated
6. Consistent over time and coherent

We focus on the principles that can be evaluated with NLP techniques. Talking about materiality, article 1 of the directive stipulates that the information should be provided "to the extent necessary for an understanding of the [...] impact of (the company's) activity". It is unclear from who's side it should be assessed as "necessary to understanding" (Jeffery, 2018). The same lack of clarity impedes the analysis of the fifth principle. The guidelines comment by saying that "companies should provide relevant, useful information on their engagement with relevant stakeholders, and how their information needs are taken into account", leaving aside the definition of the relevant stakeholders. Thus, due to an unclear definition and vague extent of information that should be provided in the reports, we omitted the two above-enumerated principles: materiality and stakeholder orientation of the disclosures.

The final property which is not measurable with NLP techniques is coherence. Going back to the original meaning of the word "coherence", we find that it originates from the Latin meaning of "to stick together". This definition implies a logical and ordered relation of words. Given that we apply a "bag of words" approach which disregards grammar and word order, our methods are not applicable to the assessment of this property.

For the rest of the principles, we provide our reasoning for the choice of NLP proxies as summarized below:

	Guidelines principles	Corresponding NLP tools
1	Fair, balanced	Tone
2	Understandable	Readability
3	Comprehensive	Report length, topics coverage
4	Concise	Boilerplates
5	Strategic and forward-looking	Horizon and strategic vocabulary
6	Consistent over time	Similarity of texts over time

(1) Fair, balanced (Tone)

The guidelines state that "the non-financial statement should give fair consideration to favourable and unfavourable aspects, and information should be assessed and presented in an unbiased way" (European Commission 2016, page 7). Prior literature has shown that the tone changes users' appreciation of financial narratives independently of content. Tone can either signal future performance (e.g. Li, 2010b) or be a result of strategic manipulations of users' perceptions (e.g. (Huang, Y.Zang, & Zheng, 2014)). In a CSR context, disclosures were demonstrated to be optimistically biased in order to manage stakeholders' impressions (Cho, Roberts, & Patten, 2010). We evaluate the tone of CSR disclosures and analyse the impact of the directive on this attribute. We expect that bias will be reduced after the directive as negative CSR disclosures are considered to be more credible than positive CSR information. We rely on a dictionary based approach using classification provided by Loughran and McDonald (2015) as the one developed specifically for the financial context.

Following Loughran & McDonald, (2011) we multiply term frequency by the inverse document frequency in order to attribute more weight to the more unusual words. Term weighting alleviate the problem of attributing too much importance to the frequent words as "the word occurring ten times more frequently is most likely not 10 times more informative" Loughran & McDonald, (2011, p.18). 'Term frequency – inverse document frequency' score (TF-IDF score) computation is presented in the following formula:

$$\text{TF} - \text{IDF score} = \frac{n_{i,j}}{\sum_j n_{i,j}} * \log(\frac{N}{df_i}) \quad (1)$$

Where

$n_{i,j}$ – number of occurrences of term i in document j

$\sum_j n_{i,j}$ – document length

N – total number of documents

df_i – number of documents containing term i

We apply TF-IDF weighting using "TfidfVectorizer" from sklearn Python module.

Our final parameter of interest is bias which represents the difference between the corresponding TF-IDF scores of positive and negative words. We associate the decrease of bias with quality improvement.

According to Loughran & McDonald, (2016), the main caveat of the dictionary approach is the negligence of negations. This problem mainly occurs when counting positive words: "negative words seem unambiguous – rarely does management negate a negative word to make a positive statement" (Loughran & McDonald, 2016 p.35). The lack of a pre-trained database for the financial reporting makes Naïve Bayes approach non-applicable when dealing with sophisticated negations. We are only limited to the straightforward detection of negations surrounding positive words. Namely, if a positive word is preceded by one of the following negations: "not", "no", "none", "nothing", "nowhere", - we count it as negative. The difference between the bias calculated taking into account negations and the direct count of Loughran & McDonald dictionary will be analysed in the discussion of the descriptive statistics.

(2) Understandable (Readability)

The guidelines suggest that "the information may also be made more understandable by using plain language and consistent terminology…" (European Commission 2016, page 7). In the context of the guidelines, understandability is associated with style attributes that are captured in textual analysis by readability indexes. Readability of documents is associated with timely information processing (You & Zhang, 2009) and it has been demonstrated that managers use a more complex language to obfuscate users in case of poor performance (Li, 2008; Ajina, Laouiti, & Msolli, 2016). Impression management by means of language obfuscation was also confirmed in CSR context (Nazari et al., 2017).

The most common metric is the FOG index (Li, 2008; Miller, 2010) calculated as follows:

$$FOG\ grade\ level = (words\ per\ sentence + percent\ of\ complex\ words) \times 0.4 \quad (2)$$

Some argue (e.g. Loughran & McDonald, 2014) that this metric is not appropriate to assess financial texts known for their sophisticated vocabulary while still well understood among financial users. CSR texts are oriented towards a larger set of users which makes the use of FOG index more justified in this context. However, we also compute readability with other indexes, such as the SMOG or the Flesch-Kincaid grade level (FKGL). The latter uses the same components as FOG, but rather than using binary classification of complex words (words with more than three syllables classified as complex), relies on the explicit count of syllables:

$$FKGL = (11.8 * syllables\ per\ word) + (0.39 * words\ per\ sentence) - 15.59 \quad (3)$$

The SMOG index, developed by McLaughlin in 1969 focuses on text comprehension of the reader (higher weight is attributed to complex words) in contrast with the writing style assessment given by FOG index (higher weight is attributed to longer sentences), which are

considered to be more appropriate in an accounting context (Loughran & McDonald, 2016, page 1196). The corresponding formula of the SMOG index is presented below:

$$SMOG = 1.0430\sqrt{percent\ of\ complex\ words \times \frac{30}{number\ of\ sentences}} + 3.1291 \quad (4)$$

We use the *textstat* Python module for the computation of the readability indexes. The resulting values are measured in grades of education required to understand the text. The level of complexity increases with the increasing level of education. We associate the decrease in readability indexes with quality improvement.

(3) Comprehensive (Report length, Topic coverage)

The comprehensiveness of texts is referred in the guidelines as the breadth of information disclosed (i.e. coverage and dispersion of different topics, Beretta & Bozzolan, 2008). We will capture this attribute with the report length and topic coverage. Report length can have a twofold interpretation. Provision of more information is likely to signify higher transparency but can also go along with increased disclosure complexity due to higher information processing costs (Li, 2008). The ambiguous nature of this parameter requires additional information to be associated with quality. We abstain from predicting its direction and will assess it with other textual attributes. We follow Lang and Stice-Lawrence (2015) and measure report length as the natural logarithm of words contained in the annual report. To split our texts into words we apply nltk Python module. Topic coverage is evaluated based on word count using lexicons corresponding to each topic as presented in Appendix **1**. Similar to sentiment analysis, we apply TF-IDF weighting using "TfidfVectorizer" from sklearn Python module. We expect a broader discussion of the matters mandated by the directive after its implementation.

(4) Concise (Boilerplates)

The guidelines reiterate numerous times that reporting should avoid the use of boilerplates: "The non-financial statement is also expected to be concise and avoid

immaterial information. Disclosing immaterial information may make the non-financial statement less easy to understand since it would obscure material information. Generic or boilerplate information that is not material should be avoided" (European Commission 2016, page 8). Thus, conciseness is associated with the avoidance of boilerplates. Lang and Stice-Lawrence (2015) define a boilerplate as an uninformative standardized disclosure. They use four-word phrases which are extremely common among documents in order to identify boilerplates. They exclude sentences with common tetragrams occurring in at least 30% of documents but not more than 75%. The main assumption is that the duplication of other firms' disclosures is unlikely to communicate firm-specific information. We rely on this methodology for the detection of boilerplates. The final values are measured in percentage. We associate the decrease in boilerplates with quality improvement.

(5) Strategic and forward-looking (Horizon and strategic vocabulary)

The suggestion to present strategic and forward-looking information echoes the findings of recent studies which confirm that horizon content increases informativeness of CSR disclosures and is associated with better quality disclosures (Muslu, Mutlu, Radhakrishnan, & Tsang, 2017). Moreover, Brochet, Loumioti, and Serafeim (2015) analyse the time horizon of conference calls' narratives and reveal that managers who emphasise on the short-term demonstrate opportunistic behaviour at the expense of long-term performance. Based on the above-mentioned studies' word selections, we assess the fourth principle of CSR reporting. The corresponding lexicon is presented in Appendix **2**. Similar to previous dictionary-based word counts, we apply TF-IDF weighting using "TfidfVectorizer" from sklearn Python module. We expect a broader use of horizon and strategic lexicon after the directive.

(6) Consistent over time (Similarity of texts over time)

The final principle states: "The content of the non-financial report should be consistent over time. This enables users of information to understand and compare past and present changes in a company's development …" (European Commission 2016, page 9). We analyse the consistency over time using an NLP metric presenting texts as a set of vectors of word frequencies between which we can compute cosine similarity over time. The raw measure of cosine similarity is mechanically related to the document length: "the longer a pair of documents, the more probable a word is included in both document, leading to a lower likelihood that the document will differ" (Brown and Tucker 2011, p.317). We remove this mechanical relationship using the Brown and Tucker (2011) methodology consisting in the removal from the raw cosine score of the fitted score. The latter is obtained from the regression of the raw score on the first five polynomials of length. The raw scores of cosine similarity are obtained from Python sklearn module. The raw values range from 0 to 1 where higher parameters are associated with higher comparability. We expect an increase in the comparability from year to year between texts within each company after the directive.

Section's conclusion

The enumerated principles of CSR disclosure quality as well as chosen NLP proxies are likely to capture the construct of quality for several reasons.

First, even though there is no such notion as universal disclosure quality, International Accounting Standards Board

and Financial Standards Board consider understandability, comparability, relevance and reliability to be useful and reflecting good practice[1]. Second, textual characteristics such as readability, tone, topic coverage and report length have been shown to reflect the

[1] In our context understandability is proxied by the readability index, comparability by cosine similarity, relevance by the relevant to CSR vocabulary (forward-looking, long-term oriented) and reliability by the unbiased tone.

informativeness of financial disclosures. This means that financial texts communicate information beyond standards' explanation of the numeric data and generate investor reaction to the variations in its style and content. Computational linguistics concludes that less readable financial reports are less informative. Quantitative and future-oriented financial narratives are more informative and tone reflects future performance (Li, 2010a). In the CSR context, this topic is still emerging. A recent study (Muslu et al., 2017) analyses the narratives of CSR reports and finds that CSR reporters with high disclosure scores are associated with more accurate forecasts and better CSR performance ratings issued by KLD. High CSR disclosure scores were associated with the disclosures that had fewer optimistic and more pessimistic keywords, were more easily readable, were longer and contained more numeric and horizon content. A study by Nazari et al. (2017) finds that within the context of voluntary CSR disclosure, increased size of CSR reports indicates better CSR performance and confirms the tendency to use more complex language in case of poor CSR performance.

Based on the quality characteristics of the guidelines as well as prior literature, we associate the improvement of textual quality with the following change in our metrics:

The directive's implementation will lead to the decrease of obfuscation, increase of topic coverage, increase of comparability, reduction of boilerplates, greater use of forward-looking, long-term oriented vocabulary and decrease of optimism.

The summary of the operationalized variables is presented in Appendix 5.

5. Sample selection

Using Reuters database, we preselected a set of companies required to disclose non-financial information based on country-level criteria published in the overview of the directive prepared by the Global Reporting Initiative (GRI, 2017). The total amount of listed companies, excluding the financial sector, represents 1164 undertakings which results in 8148 firm-year observations for the 2011-2017 period. For our analysis we keep only those companies that have uninterrupted publications during the whole period. Our final dataset represents a balanced panel with 2051 firm-year observations. A detailed breakdown by country is presented in Table **2**.

Insert Table **2** here

Given that the directive concedes non-financial disclosures in different types of corporate publications, namely annual and stand-alone reports, we ensure a comprehensive analysis by collecting both types of documents. We considered several methods for a reliable collection of the documents: Corporate Register, GRI database, companies' websites. Corporate Register claims to assemble the majority of the reports (annual and stand-alone), however it is a private database with limited access. GRI database's main limitations are its outdated references to the documents. The most credible approach appeared to be a direct hand collection from the companies' websites. We assembled the links of .pdf files for the period ranging from 2011 to 2017 and downloaded it with Python. At this stage we included only English written reports.

Since we solely focus our analysis on CSR information, the second stage of data collection consisted in the extraction of CSR sections from annual reports. The automation of this process is explained in Appendix **3**. Our final dataset is summarized in Table **3**.

Insert Table **3** here

6. Design

The impact of the regulation is evaluated using the difference-in-difference (DID) design. First, we compare textual quality of CSR disclosures prior to and after 2014. Then, we roll up to 2017 to explore the impact power over the consequent years. We use a seven-year window from 2011 to 2017 ensuring 3 years before and after directive's adoption. Our sample is limited to the companies issuing CSR disclosures before the directive and continuing this practice after its implementation. The DID groups are schematized below:

Groups	Before treatment	After treatment
Control	Mandatory reporting	Mandatory
Treatment	Voluntary reporting	Mandatory

Scandinavian and UK undertakings operate in the environments where CSR disclosures were mandated long before 2014. We assign companies from UK and Scandinavia (Denmark, Sweden) to the control group, while firms from countries with voluntary frameworks are studied as a treatment group. The experience of Scandinavian countries served as a model for the directive (Szabó & Sørensen, 2015) while UK has been shown to have a higher degree of compliance with the new requirements (Salvatori et al, 2015) leading to the smallest post-directive change in these countries. UK companies have historically paid more attention to issues of long-term environmental and social risk (Chen & Bouvain, 2009). Among the EU member states, UK is considered to be the leader in CSR and has even appointed a CSR Minister. Meanwhile, Denmark and Sweden are called 'welfare states' as they have higher socially oriented political regimes and have attained a high CSR performance over the past years (Mullerat, 2013). We do not include France in the control group due to its low compliance with the directive (Salvatori et al, 2015).

The difference-in-difference method consists of analysing a regression of the form:

$$y = \alpha + \beta_1 Time + \beta_2 Treatment + \beta_3 Time * Treatment + \beta_4 * Controls + \beta_5 * IndustryFE + e$$

Where

y – textual outcome

β_1- expected mean change from before to after the shock among the control group (UK, Scandinavia)

β_2- estimated mean difference in y between the treatment and control groups before the shock

β_3- difference-in-difference estimator, shows whether the mean differences between the treatment and control groups after the shock are significantly different from the gap that existed before the shock between those same groups

β_4 – control variables isolating effects of companies' characteristics (such as growth (relative change in revenues), size (market value), financial performance (ROA) and CSR performance (A4IR)) on the content and presentation of reports

β_5 – industry fixed effects

The parallel trend assumption requires that, in the absence of treatment, the difference between the 'treatment' and 'control' group is constant over time. We verify this condition based on trends presented in Graph **1**.

Insert Graph **1** here

The variables that clearly violate parallel trend assumption are the percentage of words belonging to the human rights and environmental categories as well as the FOG index. We exclude these parameters from further analysis. Graphical representation also indicates a sharp decrease in texts' similarity for the implementation year showing some level of reaction to the new legislation. We further exclude 2017 from the comparability analysis.

7. Descriptive statistics

The average values of key variables computed for our final dataset along with other descriptive statistics are presented in Table **4**.

Insert Table **4** here

Panel A summarizes descriptive statistics for our primary variables: CSR reports length, tone (bias)[2], readability (FOG, Flesch, SMOG), topic coverage, boilerplates, horizon and strategic vocabulary and similarity of texts over time. All variables are adjusted for outliers with *winsor2, cuts(5,95)* Stata command.

The textual part of the average CSR section without digits has about 18,000 words ranging from 1,300 to 66,000 words. The readability indexes are around 18 meaning that for the understanding of those texts the reader require 18 years of education similar to the values found in the literature analysing 10-K reports (Lang & Stice-Lawrence, 2015; Loughran & McDonald, 2014): 18 years for FOG, 19 years for SMOG (Muslu et al., 2017); and slightly different for Flesch which is equal to 52 before conversion to grade level or 14 in grade equivalent (De Franco, Hope, Vyas, & Zhou, 2015). The percentage of positive words is higher than the percentage of negative words when computed before weighting for the inverse document frequency (idf). CSR section's similarity before Taylor adjustment varies from 0.54 to 0.94 showing a high dispersion of texts' comparability going from 'slightly' to 'very' similar. The boilerplate measure has a mean of 6.65%. The presence of words characterizing specific topics does not exceed 1.4%, before weighting for idf, with the highest observations attributed to the environmental issues.

Table **4**, Panel B reports descriptive statistics over time. The number of words increases during our seven-year window, the readability indexes go up as well as the use of boilerplates perhaps due to the lack of clear instructions from the Member States'

[2] The table below contains descriptive statistics for the bias computed with and without negations (bias_with_neg and bias_no_neg, respectively).

	mean	sd	min	p50	max
bias_no_neg	0.48	0.77	-4.55	0.43	3.78
bias_with_neg	0.47	0.76	-4.55	0.41	3.78

It follows that the inclusion of the negations in sentiment analysis as described in the previous section does not significantly change the value of the bias. We pursue our analysis using tone count without correction for negations.

governments as to the implementation of the directive between 2014 and 2017. Reporting also becomes more similar over time. The bribery and human rights coverage gains more importance. Finally, the bias, reflecting the level of manipulation, decreases which is probably due to the mandatory nature of the legislation with several countries introducing penalties for its breach.

Table **4**, Panel C presents descriptive statistics for our two groups of analysis. Our control group seems to have more concise reports with lower complexity given SMOG and Flesch indexes and a lower use of boilerplates.

Finally, our choice of the control group is consistent with the breakdown presented in Panel A of Table **5**. The latter demonstrates that the treatment group experiences significant changes after the shock compared to the previous years. The control group, on the contrary, has minor adjustments that confirms its status as a benchmark.

Insert Panel A Table **5** here

8. Results

Table **6** provides our set of results examining the effect of regulation on textual attributes of CSR disclosures. Panels A, B, C and D reflect different choices of the year of impact from 2014 to 2017.

Insert Table **6** here

The effect of the directive becomes more pronounced starting from 2015. The results get stronger as we approach the implementation date of the directive. The amount of reported CSR information increases significantly in parallel with the use of boilerplates. The results also indicate significant differences between control and treatment groups prior to the directive. The treatment group appears to have lengthier reports, use more obfuscating language, have less comparable reports and contain more boilerplates while having lower

coverage of bribery matters. These differences indicate lower reporting quality in the treatment group. However, we also obtained some surprising values for the latter. For example, we find a lower bias and more future-oriented and strategic information in the treatment group compared to the control group before the directive.

The time trend by country presented in Panel B of Table **5** provides additional insights into the effects of the directive.

Insert Panel B Table **5** here

We include in this breakdown only those countries that have significant changes for any variable of interest. The main contributors to the increase in report length are Germany, France and Austria. This increase goes in parallel with the increased complexity captured by readability indexes (Germany) or the use of boilerplates (France) – another indicator of low-quality growth of CSR reports. The only country demonstrating a reduction in complexity is Sweden which has implemented mandatory CSR reporting prior to the directive. Denmark confirms its status of the directive's benchmark as there is no change induced by the new legislation. Finland and Germany demonstrate a decreasing bias in the presentation of CSR information both countries paying more attention to the bribery matters. The case of Germany is particularly interesting as this country is still recovering from the corruption scandals in automotive industry. Moreover, Germany introduced a ten million fine for the breach of the new legislation.

9. Conclusion

Our study has several limitations. The nature of our variables implies a measurement error of the selected textual attributes. Nevertheless, measurement errors in the dependent variables do not affect consistency of estimation. We are also constrained to process only

English written reports which reduce our sample size along with systematically corrupted .pdf files.

Our results indicate a lack of quality improvement during the transitional period between the adoption of the directive and its implementation. Companies seem to be rather confused by the new legislation as the level of complexity increases when proxied by the length and the use of boilerplates. Mandatory framework forces companies to address CSR issues, however, the lack of enforcement results in a low quality of published information.

There is a clear distinction between the treatment and control groups. The latter experiences the smallest and only positive changes with a decreasing bias and increasing comparability over time, while the former presents significant time trend changes for almost all parameters.

The directive seems to have the highest impact in the countries with monetary penalties. It is interesting to note that countries which introduced possible jail sentences for breaches are those that determined a very limited scope of companies falling under the directive (below ten companies), thus, we could not asses the role of this legal enforcement.

Overall, our findings rather confirm that agency theory dominates corporate view on CSR activities, however, these results are preliminary and require further analysis of the directive impact after its actual implementation in 2017, which we plan to realise by complementing our dataset with 2018 reports. For further analysis of the economic consequences we would like to contrast the set of companies experiencing positive changes with the ones having a weak reaction to the directive. This stage will require developing a new quality measure based on the parameters evaluated in the current study.

Bibliography

Ajina, A., Laouiti, M., & Msolli, B. (2016). Guiding through the Fog: Does annual report readability reveal earnings management? *Research in International Business and Finance*. https://doi.org/10.1016/j.ribaf.2016.07.021

Beretta, S., & Bozzolan, S. (2008). Quality versus quantity: the case of forward-looking disclosure. *Journal of Accounting, Auditing & Finance*. https://doi.org/10.1177/0148558X0802300304

Beyer, A., Cohen, D. A., Lys, T. Z., & Walther, B. R. (2010). The financial reporting environment: Review of the recent literature. *Journal of Accounting and Economics*. https://doi.org/10.1016/j.jacceco.2010.10.003

Brochet, F., Loumioti, M., & Serafeim, G. (2015). Speaking of the short-term: disclosure horizon and managerial myopia. *Review of Accounting Studies*. https://doi.org/10.1007/s11142-015-9329-8

Brown, S. V., & Tucker, J. W. (2011). Large-Sample Evidence on Firms' Year-over-Year MD&A Modifications. *Journal of Accounting Research*. https://doi.org/10.1111/j.1475-679X.2010.00396.x

Campbell, D., & Slack, R. (2011). Environmental disclosure and environmental risk: Sceptical attitudes of UK sell-side bank analysts. *British Accounting Review*. https://doi.org/10.1016/j.bar.2010.11.002

Chen, S., & Bouvain, P. (2009). Is corporate responsibility converging? a comparison of corporate responsibility reporting in the USA, UK, Australia, and Germany. In *Journal of Business Ethics*. https://doi.org/10.1007/s10551-008-9794-0

Cho, C. H., & Patten, D. M. (2007). The role of environmental disclosures as tools of legitimacy: A research note. *Accounting, Organizations and Society*. https://doi.org/10.1016/j.aos.2006.09.009

Cho, C. H., Roberts, R. W., & Patten, D. M. (2010). The language of US corporate environmental disclosure. *Accounting, Organizations and Society*. https://doi.org/10.1016/j.aos.2009.10.002

Clarkson, P. M., Fang, X., Li, Y., & Richardson, G. (2013). The relevance of environmental disclosures: Are such disclosures incrementally informative? *Journal of Accounting and Public Policy*. https://doi.org/10.1016/j.jaccpubpol.2013.06.008

De Franco, G., Hope, O. K., Vyas, D., & Zhou, Y. (2015). Analyst report readability. *Contemporary Accounting Research*. https://doi.org/10.1111/1911-3846.12062

Dhaliwal, D. S., Li, O. Z., Tsang, A., & Yang, Y. G. (2011). Voluntary nonfinancial disclosure and the cost of equity capital: The initiation of corporate social responsibility reporting. *Accounting Review*. https://doi.org/10.2308/accr.00000005

Dhaliwal, D. S., Radhakrishnan, S., Tsang, A., & Yang, Y. G. (2012). Nonfinancial disclosure and analyst forecast accuracy: International evidence on corporate social responsibility disclosure. *Accounting Review*. https://doi.org/10.2308/accr-10218

European Commission. (2016). European Commission: Non-Financial Reporting.

Fiechter, P., Hitz, J.-M., & Lehmann, N. (2017). *Real Effects in Anticipation of Mandatory Disclosures: Evidence from the European Union's CSR Directive*. *SSRN*. https://doi.org/10.2139/ssrn.3033883

GRI. (2017). Member State Implementation of Directive 2014/95/EU. *CSR Europe*.

Gutsche, R. (2016). Firm-Value Effects of CSR Disclosure and CSR Performance. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2821691

Hart, S. L. (1995). A natural resource-based view of the firm. *Academy of Management Review*. https://doi.org/10.5465/AMR.1995.9512280033

Huang, A. H., Y.zang, A., & Zheng, R. (2014). Evidence on the information content of text in analyst reports. *Accounting Review*. https://doi.org/10.2308/accr-50833

Jeffery, C. (2018). *Comparing the Implementation of the EU Non-Financial Reporting Directive*. *SSRN*. https://doi.org/10.2139/ssrn.3083368

Jegadeesh, N., & Wu, D. (2013). Word power: A new approach for content analysis. *Journal of Financial Economics*. https://doi.org/10.1016/j.jfineco.2013.08.018

Jones, T. M. (1995). Instrumental stakeholder theory: a synthseis of ethics and economics. *Academy of Management Review*. https://doi.org/10.5465/AMR.1995.9507312924

Lang, M., & Stice-Lawrence, L. (2015). Textual analysis and international financial reporting: Large sample evidence. *Journal of Accounting and Economics*. https://doi.org/10.1016/j.jacceco.2015.09.002

Lawrence, A. (2013). Individual investors and financial disclosure. *Journal of Accounting and Economics*. https://doi.org/10.1016/j.jacceco.2013.05.001

Leuz, C., & Wysocki, P. D. (2016). The Economics of Disclosure and Financial Reporting Regulation: Evidence and Suggestions for Future Research. *Journal of Accounting Research*. https://doi.org/10.1111/1475-679X.12115

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

Li, F. (2010a). Textual Analysis of Corporate Disclosures: A Survey of the Literature. *Journal of Accounting Literature*. https://doi.org/10.1093/rfs/hhu089

Li, F. (2010b). The information content of forward- looking statements in corporate filings-A naïve bayesian machine learning approach. *Journal of Accounting Research*. https://doi.org/10.1111/j.1475-679X.2010.00382.x

Loughran, T., & McDonald, B. (2011). When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *Journal of Finance*. https://doi.org/10.1111/j.1540-6261.2010.01625.x

Loughran, T., & McDonald, B. (2014). Measuring readability in financial disclosures.

Journal of Finance. https://doi.org/10.1111/jofi.12162

Loughran, T., & McDonald, B. (2016). Textual Analysis in Accounting and Finance: A Survey. *Journal of Accounting Research*. https://doi.org/10.1111/1475-679X.12123

Loughran, T., & McDonald, B. (2015). The Use of Word Lists in Textual Analysis. *Journal of Behavioral Finance*. https://doi.org/10.1080/15427560.2015.1000335

Matsumura, E. M., Prakash, R., & Vera-Muñoz, S. C. (2014). Firm-value effects of carbon emissions and carbon disclosures. *Accounting Review*. https://doi.org/10.2308/accr-50629

Michelon, G., Pilonato, S., & Ricceri, F. (2015). CSR reporting practices and the quality of disclosure: An empirical analysis. *Critical Perspectives on Accounting*. https://doi.org/10.1016/j.cpa.2014.10.003

Miller, B. P. (2010). The effects of reporting complexity on small and large investor trading. *Accounting Review*. https://doi.org/10.2308/accr.00000001

Mullerat, R. (2013). *Corporate Social Responsibility: A European Perspective*. *Miami-Florida European Union Center of Excellence*.

Muslu, V., Mutlu, S., Radhakrishnan, S., & Tsang, A. (2017). Corporate Social Responsibility Report Narratives and Analyst Forecast Accuracy. *Journal of Business Ethics*. https://doi.org/10.1007/s10551-016-3429-7

Nazari, J. A., Hrazdil, K., & Mahmoudian, F. (2017). Assessing social and environmental performance through narrative complexity in CSR reports. *Journal of Contemporary Accounting and Economics*. https://doi.org/10.1016/j.jcae.2017.05.002

Plumlee, M., Brown, D., Hayes, R. M., & Marshall, R. S. (2015). Voluntary environmental disclosure quality and firm value: Further evidence. *Journal of Accounting and Public Policy*. https://doi.org/10.1016/j.jaccpubpol.2015.04.004

Salvatori et al. (2015). Transparency and Discretion : a comparative study of the transposition

of the Non-Financial Reporting Directive, 1–24.

Spence, M. (1973). Job Market Signaling. *The Quarterly Journal of Economics*. https://doi.org/10.2307/1882010

Szabó, D. G., & Sørensen, K. E. (2015). New EU Directive on the Disclosure of Non-Financial Information (CSR). *European Company & Financial Law Review*, *12*(3), 307. https://doi.org/182/6/3919 [pii]10.4049/jimmunol.0803834

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

Zerbini, F. (2017). CSR Initiatives as Market Signals: A Review and Research Agenda. *Journal of Business Ethics*. https://doi.org/10.1007/s10551-015-2922-8

Table 1: Member States requirements before and after the directive (Source: GRI 2017)
This table presents the comparison of the implementation requirements formulated in the directive and their transposition into the Member States legislation. We compare the scope of entities falling under the directive, the penalties and the level of involvement of the auditor.
= indicates the same level of requirements as determined by the directive.

Country	Staff	Balance Sheet	Net Turnover	Currency	Comparison*	Penalties	Audit
Directive	500	20 mio	40 mio	EUR	benchmark	not specified	existence check
Austria	500	20 mio	40 mio	EUR	= ***	Monetary	=
Belgium	500	17 mio	34 mio	EUR	stricter	Monetary	=
Bulgaria	500	38 mio	76 mio	BGN	=	Monetary	= +content check
Croatia	500	15 mio	30 mio	HRK	stricter	Monetary	=
Cyprus	500	20 mio	40 mio	EUR	=	Monetary	= +content check
Czech Republic	500	na**	1 bn	CZK	=	Monetary	=
Denmark	250	156 mio	313 mio	DKK	=	=	= +content check
Estonia	500	na	na	na	stricter	=	=
Finland	500	20 mio	40 mio	EUR	=	Monetary	=
France	500	20 mio	40 mio	EUR	=	Monetary	= +content check
Germany	500	20 mio	40 mio	EUR	=	Monetary	not mandatory
Greece	500	na	na	na	stricter	Monetary	=
Hungary	500	6 bn	12 bn	HUF	=	Monetary	=
Iceland	250	3 bn	6 bn	ISK	stricter	Monetary/ Imprisonment	= +content check
Italy	500	20 mio	40 mio	EUR	=	Monetary	= +content check
Latvia	500	20 mio	40 mio	EUR	=	Monetary	= +content check
Lithuania	500	20 mio	40 mio	EUR	=	Monetary	=
Luxembourg	250	20 mio	40 mio	EUR	stricter	=	=
Malta	500	20 mio	40 mio	EUR	=	Monetary	=
Netherlands	500	20 mio	40 mio	EUR	=	=	=
Norway	500	20 mio	40 mio	EUR	=	Monetary/ Imprisonment	=
Poland	500	85 mio	170 mio	PLN	=	Monetary	=
Portugal	500	na	na	na	stricter	Monetary	=
Republic of Ireland	500	20 mio	40 mio	EUR	=	Monetary/ Imprisonment	=
Romania	500	na	na	na	stricter	Monetary	= +content check
Slovak Republic	500	20 mio	40 mio	EUR	=	Monetary	=
Slovenia	500	20 mio	40 mio	EUR	=	Monetary	=
Spain	500	20 mio	40 mio	EUR	=	=	=
Sweden	250	175 mio	350 mio	SEK	stricter	Monetary	=
United Kingdom	500	na	na	na	stricter	Monetary	= +content check

* The comparison is made after conversion to EUR

** na: not applicable

*** =: similar to directive

Table 2: Companies falling under the directive
This table presents the number of companies preselected according to the criteria determined by each Member State.

Listed EU companies falling under the directive	
Country	**N GRI, firms**
United Kingdom	345
Germany	175
Poland	115
France	82
Spain	65
Denmark	48
Finland	47
Italy	47
Netherlands	44
Sweden	42
Greece	29
Austria	28
Portugal	19
Belgium	14
Norway	10
Republic of Ireland	7
Croatia	5
Czech Republic	5
Iceland	5
Lithuania	5
Slovenia	5
Romania	4
Slovak Republic	4
Hungary	3
Luxembourg	3
Cyprus	2
Estonia	2
Latvia	2
Malta	2
Total	**1164**

Table 3: Available documents

This table illustrates the process of sample selection. First, we download available .pdf files, then we convert it to .txt. There is an additional stage required for the annual reports consisting in the extraction of CSR sections explained in more details in Appendix 3. The final sample size consists of the balanced firm-year observations while the rest of the values are presented before conversion to the balanced panel.

Country	N GRI, firm-year	Annual reports			Stand-alone CSR reports		N final, firm-year (balanced)
		PDF	TXT	CSR section	PDF	TXT	
United Kingdom	2415	2012	1392	876	276	204	756
Germany	1225	850	399	270	203	140	119
Poland	805	317	266	125	18	18	77
France	574	371	280	261	87	68	203
Spain	455	317	208	159	112	94	70
Denmark	336	283	236	156	156	129	154
Finland	329	244	196	155	110	97	126
Italy	329	258	228	62	49	42	42
Netherlands	308	231	176	142	63	50	91
Sweden	294	211	191	170	80	79	140
Greece	203	124	108	72	51	42	63
Austria	196	139	91	68	23	17	35
Portugal	133	96	61	29	34	18	28
Belgium	98	84	73	38	5	5	14
Norway	70	61	60	51	7	7	49
Republic of Ireland	49	46	38	26	9	7	14
Croatia	35	19	19	9	4	4	7
Czech Republic	35	28	20	9	4	2	7
Iceland	35	13	10	6	0	0	0
Lithuania	35	33	17	8	0	0	7
Slovenia	35	35	27	27	4	3	21
Romania	28	16	12	11	10	7	7
Slovak Republic	28	20	20	10	0	0	14
Hungary	21	0	0	0	0	0	0
Luxembourg	21	6	0	0	0	0	0
Cyprus	14	11	10	0	1	1	0
Estonia	14	9	7	2	0	0	0
Latvia	14	7	7	4	1	1	7
Malta	14	14	7	1	0	0	0
Total	**8148**	**5855**	**4159**	**2747**	**1307**	**1035**	**2051**

Table 4: Descriptive statistics

This table reports the descriptive statistics for all the variables. Panel A provides the summary statistics. Panel B reports descriptive statistics over time. Panel C presents descriptive statistics for the control and treatment groups. All variables are adjusted for outliers with winsor2 at the 5th and 95th percentile. Detailed variables definition is presented in Appendix 5. The variables are presented before other transformations presented in Appendix 5.

Panel A: Primary variable descriptive statistics

	mean	sd	min	p50	max
Length	18271.63	19132.28	1375.00	9601.00	66322.00
FOG	19.12	2.29	15.29	19.02	23.71
Flesch	17.37	2.32	13.90	17.00	22.40
SMOG	17.96	1.50	15.60	17.70	21.20
Bias	0.51	0.66	-0.66	0.45	1.78
Similarity	0.80	0.11	0.54	0.83	0.94
Boilerplate	6.65	3.50	1.52	6.05	14.22
Horizon	0.03	0.03	0.00	0.02	0.09
Bribery	0.06	0.08	0.00	0.03	0.29
Environmental	1.39	0.53	0.51	1.35	2.43
Human Rights	0.32	0.16	0.07	0.29	0.68
Social	1.17	0.39	0.45	1.15	1.89
Control variables					
lnMV	6.93	2.21	1.05	6.86	13.62
DiffREV	0.20	5.91	-1.00	0.02	221.21
ROA	5.74	7.77	-72.58	5.21	60.55
A4IR	51.31	40.65	0.00	71.53	96.51
N	1404				

Panel B: Text characteristics by pre, post directive period

	Mean (post 2014)	Mean (pre 2015)	Diff. (post-pre)
Length	19998.44	17212.04	2786.40**
FOG	18.95	19.13	-0.17
Flesch	17.52	17.19	0.33**
SMOG	18.06	17.85	0.21**
Bias	0.47	0.54	-0.08**
Similarity	0.81	0.79	0.02***
Boilerplate	7.02	6.36	0.66***
Horizon	0.03	0.03	-0.01***
Bribery	0.07	0.05	0.02***
Environmental	1.38	1.39	-0.00
Human Rights	0.35	0.30	0.05***
Social	1.18	1.15	0.03
Observations	879	1172	2051

Panel C: Text characteristics by control, treatment group

	Mean (treatment)	Mean (control)	Diff. (treatment -control)
Length	24773.37	12336.18	12437.19***
FOG	18.87	19.22	-0.35***
Flesch	17.95	16.75	1.21***
SMOG	18.37	17.52	0.85***
Bias	0.35	0.67	-0.32***
Similarity	0.81	0.79	0.02***
Boilerplate	7.85	5.50	2.35***
Horizon	0.02	0.04	-0.02***
Bribery	0.04	0.08	-0.04***
Environmental	1.49	1.29	0.20***
Human Rights	0.33	0.32	0.01
Social	1.16	1.18	-0.02
Observations	1001	1050	2051

See Appendix 5 for a description of variables

Table 5: Time trends (before and after 2015)

This table presents the time trends. Panel A provides information about time changes in the control and treatment groups. Panel B illustrates the breakdown of the time trend by country. This analysis only includes variables selected for the final DID fulfilling parallel trend assumption. Detailed variables definition is presented in Appendix 5.

Panel A: by group

Group	lnLen_w	Flesch_w	SMOG_w	bias_w	TadjCos_w	lnBoiler_w	Hor_w	lnBrb_w	social_w
Control	0.082	0.122	0.071	**-0.010***	**0.012****	0.031	0.020	0.069	-0.002
	(0.050)	(0.175)	(0.092)	(0.006)	(0.005)	(0.020)	(0.029)	(0.060)	(0.002)
Treatment	**0.264*****	**0.374*****	**0.260*****	**-0.012****	0.011	**0.150*****	**0.066*****	**0.160****	0.001
	(0.058)	(0.111)	(0.079)	(0.004)	(0.006)	(0.020)	(0.013)	(0.061)	(0.002)
All	**0.165*****	**0.233****	**0.152****	**-0.011*****	**0.012****	**0.087*****	**0.042****	0.075	-0.001
	(0.030)	(0.084)	(0.052)	(0.004)	(0.005)	(0.009)	(0.016)	(0.050)	(0.002)

Panel B: by country

Country	lnLen_w	Flesch_w	SMOG_w	bias_w	TadjCos_w	lnBoiler_w	Hor_w	lnBrb_w	social_w
Denmark	0.049	0.354	0.232	-0.033	0.004	0.107	0.081	0.261	-0.006
	(0.085)	(0.384)	(0.230)	(0.024)	(0.023)	(0.097)	(0.055)	(0.172)	(0.005)
Finland	0.106	**0.837*****	**0.547*****	**-0.024***	-0.002	**0.209*****	0.045	**0.365****	0.011
	(0.097)	(0.140)	(0.090)	(0.010)	(0.011)	(0.038)	(0.038)	(0.113)	(0.007)
Sweden	-0.163	**-0.749****	**-0.427***	-0.029	0.003	0.134	**-0.150*****	-0.134	**0.008***
	(0.214)	(0.250)	(0.204)	(0.017)	(0.018)	(0.182)	(0.026)	(0.229)	(0.004)
UK	0.110	0.189	0.093	-0.001	**0.015****	0.008	0.027	0.048	-0.003
	(0.064)	(0.210)	(0.100)	(0.007)	(0.006)	(0.040)	(0.043)	(0.076)	(0.002)
Germany	**0.241***	**0.808****	**0.582*****	**-0.029****	**-0.038*****	0.213	0.011	**0.245****	0.004
	(0.097)	(0.299)	(0.123)	(0.011)	(0.008)	(0.113)	(0.062)	(0.069)	(0.003)
France	**0.286*****	0.270	0.263	-0.011	**0.036****	**0.169****	-0.000	0.090	-0.004
	(0.039)	(0.220)	(0.137)	(0.014)	(0.014)	(0.049)	(0.015)	(0.076)	(0.004)
Austria	**0.669****	0.200	0.312	-0.022	0.007	-0.080	0.048	0.074	-0.017
	(0.126)	(0.359)	(0.281)	(0.030)	(0.014)	(0.089)	(0.104)	(0.151)	(0.007)
Italy	-0.720	-0.059	-0.213	-0.031	**-0.067****	**0.316*****	-0.199	0.033	**-0.010***
	(0.416)	(1.189)	(0.921)	(0.033)	(0.015)	(0.042)	(0.130)	(0.180)	(0.003)

See Appendix 5 for a description of variables

Table 6: Difference-in-difference

This table presents DID analysis results regarding the impact of the directive on the metrics capturing textual quality of CSR disclosures. Detailed variables definition is presented in Appendix 5. Panels A, B, C and D reflect different choices of the year of impact from 2014 to 2017. We exclude from the DID analysis those variables of interest that do not fulfil parallel trend assumption illustrated in Graph 1.

Panel A: year of impact – 2014

VARIABLES	(1) lnLen_w	(2) Flesch_w	(3) SMOG_w	(4) bias_w	(5) TadjCos_w	(6) lnBoiler_w	(7) Hor_w	(8) lnBrb_w	(9) social_w
time	0.109*	0.184	0.115	-0.011**	0.005*	0.075***	0.040	0.021	-0.003
	(0.050)	(0.148)	(0.070)	(0.004)	(0.003)	(0.018)	(0.029)	(0.057)	(0.003)
treated	0.786***	1.137***	0.790***	-0.026	-0.016***	0.328***	0.070*	-0.705***	-0.006
	(0.108)	(0.350)	(0.211)	(0.015)	(0.005)	(0.038)	(0.038)	(0.082)	(0.006)
did	0.089	0.171	0.134	-0.002	0.003	**0.078****	0.016	0.040	0.004**
	(0.085)	(0.242)	(0.138)	(0.007)	(0.007)	(0.034)	(0.029)	(0.095)	(0.002)
lnMV	0.052***	-0.046	-0.028	0.002	0.004***	0.003	0.006	-0.011	-0.001
	(0.012)	(0.040)	(0.020)	(0.003)	(0.001)	(0.008)	(0.007)	(0.019)	(0.001)
DiffREV	-0.004***	0.002**	0.005***	0.001***	0.000***	-0.001***	-0.001***	0.046	-0.000***
	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.091)	(0.000)
ROA	0.000	-0.006	-0.003	0.001	0.000	-0.001	-0.001	0.004	-0.000
	(0.003)	(0.007)	(0.005)	(0.001)	(0.000)	(0.002)	(0.001)	(0.003)	(0.000)
A4IR	0.013***	0.004	0.002	0.000	-0.000	-0.001*	0.004***	-0.002**	0.000
	(0.001)	(0.004)	(0.002)	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)
Constant	7.768***	16.867***	17.573***	-0.006	-0.019**	1.579***	1.612***	-4.266***	0.133***
	(0.098)	(0.456)	(0.260)	(0.028)	(0.009)	(0.086)	(0.060)	(0.149)	(0.011)
Observations	1,692	1,692	1,692	1,692	1,404	1,664	1,692	1,228	1,692
R-squared	0.339	0.079	0.090	0.030	0.018	0.129	0.148	0.141	0.008
Industry FE	YES	YES	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

See Appendix 5 for a description of variables

Panel B: year of impact – 2015

VARIABLES	(1) lnLen_w	(2) Flesch_w	(3) SMOG_w	(4) bias_w	(5) TadjCos_w	(6) lnBoiler_w	(7) Hor_w	(8) lnBrb_w	(9) social_w
time	0.093*	0.118	0.064	-0.011*	0.011**	0.028	0.021	0.054	-0.003
	(0.045)	(0.168)	(0.087)	(0.005)	(0.005)	(0.019)	(0.029)	(0.065)	(0.002)
treated	0.770***	1.131***	0.788***	-0.027*	-0.015***	0.318***	0.059	-0.724***	-0.005
	(0.101)	(0.325)	(0.201)	(0.014)	(0.004)	(0.038)	(0.036)	(0.096)	(0.006)
did	**0.150***	0.239	0.182	-0.000	0.000	**0.121*****	0.043	0.083	0.004
	(0.082)	(0.250)	(0.145)	(0.006)	(0.006)	(0.036)	(0.034)	(0.093)	(0.002)
lnMV	0.052***	-0.045	-0.027	0.002	0.004***	0.004	0.007	-0.012	-0.001
	(0.012)	(0.040)	(0.020)	(0.003)	(0.001)	(0.008)	(0.007)	(0.019)	(0.001)
DiffREV	-0.004***	0.003**	0.005***	0.001***	0.000***	-0.001***	-0.001***	0.036	-0.000***
	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.096)	(0.000)
ROA	0.000	-0.006	-0.003	0.001	0.000	-0.001	-0.001	0.004	-0.000
	(0.003)	(0.007)	(0.005)	(0.001)	(0.000)	(0.002)	(0.001)	(0.003)	(0.000)
A4IR	0.013***	0.004	0.002	0.000	-0.000	-0.001*	0.004***	-0.002**	0.000
	(0.001)	(0.003)	(0.002)	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)
Constant	7.793***	16.923***	17.612***	-0.008	-0.020**	1.610***	1.627***	-4.276***	0.132***
	(0.099)	(0.446)	(0.256)	(0.028)	(0.009)	(0.085)	(0.060)	(0.165)	(0.012)
Observations	1,692	1,692	1,692	1,692	1,404	1,664	1,692	1,228	1,692
R-squared	0.342	0.079	0.090	0.031	0.020	0.128	0.148	0.143	0.008
Industry FE	YES	YES	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
See Appendix 5 for a description of variables

Panel C: year of impact – 2016

VARIABLES	(1) lnLen_w	(2) Flesch_w	(3) SMOG_w	(4) bias_w	(5) TadjCos_w	(6) lnBoiler_w	(7) Hor_w	(8) lnBrb_w	(9) social_w
time	0.124**	0.146	0.098	-0.010*	0.003	0.024	0.033	0.002	-0.002
	(0.047)	(0.161)	(0.074)	(0.005)	(0.008)	(0.024)	(0.031)	(0.046)	(0.002)
treated	0.783***	1.155***	0.813***	-0.026*	-0.017***	0.333***	0.072*	-0.731***	-0.004
	(0.099)	(0.315)	(0.194)	(0.014)	(0.003)	(0.039)	(0.039)	(0.103)	(0.006)
did	**0.186****	0.287	0.199	-0.004	**0.013***	**0.138*****	0.028	0.140	0.002
	(0.083)	(0.249)	(0.142)	(0.008)	(0.007)	(0.042)	(0.043)	(0.116)	(0.003)
lnMV	0.052***	-0.046	-0.027	0.002	0.004***	0.004	0.006	-0.012	-0.001
	(0.012)	(0.039)	(0.020)	(0.003)	(0.001)	(0.008)	(0.007)	(0.018)	(0.001)
DiffREV	-0.004***	0.003**	0.006***	0.001***	0.000***	-0.001***	-0.001***	0.046	-0.000***
	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.091)	(0.000)
ROA	-0.000	-0.006	-0.003	0.001	0.000	-0.001	-0.001	0.004	-0.000
	(0.003)	(0.007)	(0.005)	(0.001)	(0.000)	(0.002)	(0.002)	(0.003)	(0.000)
A4IR	0.013***	0.004	0.003	0.000	-0.000	-0.001*	0.004***	-0.002**	0.000
	(0.001)	(0.004)	(0.002)	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)
Constant	7.799***	16.933***	17.612***	-0.010	-0.017*	1.615***	1.626***	-4.258***	0.132***
	(0.097)	(0.444)	(0.254)	(0.029)	(0.009)	(0.087)	(0.058)	(0.179)	(0.012)
Observations	1,692	1,692	1,692	1,692	1,404	1,664	1,692	1,228	1,692
R-squared	0.345	0.080	0.092	0.031	0.018	0.129	0.148	0.143	0.007
Industry FE	YES	YES	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
See Appendix 5 for a description of variables

Panel D: year of impact – 2017

VARIABLES	(1) lnLen_w	(2) Flesch_w	(3) SMOG_w	(4) bias_w	(6) lnBoiler_w	(7) Hor_w	(8) lnBrb_w	(9) social_w
time	0.187**	0.196	0.108	-0.014**	0.017	0.046**	0.071	-0.003
	(0.070)	(0.194)	(0.101)	(0.005)	(0.038)	(0.020)	(0.063)	(0.002)
treated	0.782***	1.192***	0.834***	-0.028**	0.351***	0.068*	-0.693***	-0.003
	(0.094)	(0.304)	(0.193)	(0.013)	(0.041)	(0.033)	(0.094)	(0.007)
did	**0.382*** **	0.360	0.274	0.006	**0.170*** **	0.076	0.066	-0.000
	(0.113)	(0.256)	(0.158)	(0.006)	(0.045)	(0.045)	(0.086)	(0.002)
lnMV	0.051***	-0.045	-0.027	0.002	0.004	0.006	-0.012	-0.001
	(0.011)	(0.039)	(0.020)	(0.003)	(0.008)	(0.007)	(0.018)	(0.001)
DiffREV	-0.004***	0.003**	0.005***	0.001***	-0.001**	-0.001***	0.036	-0.000***
	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.091)	(0.000)
ROA	-0.001	-0.007	-0.003	0.001	-0.001	-0.001	0.004	-0.000
	(0.003)	(0.007)	(0.005)	(0.001)	(0.002)	(0.002)	(0.003)	(0.000)
A4IR	0.013***	0.004	0.003	0.000	-0.001	0.004***	-0.002**	0.000
	(0.001)	(0.004)	(0.002)	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)
Constant	7.799***	16.934***	17.616***	-0.010	1.615***	1.627***	-4.270***	0.132***
	(0.096)	(0.446)	(0.257)	(0.028)	(0.090)	(0.053)	(0.174)	(0.012)
Observations	1,692	1,692	1,692	1,692	1,664	1,692	1,228	1,692
R-squared	0.354	0.080	0.092	0.029	0.127	0.152	0.143	0.008
Industry FE	YES	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
See Appendix 5 for a description of variables

Appendix 1: CSR Vocabulary

General CSR terms
2014/95/eu
aa1000
csr
gri
integrate
non-financial
oecd
sa800
sasb
sdg
ungc
voluntary

Environmental
biodiversity
biogenic
carbon
climate
conservation
dioxide
diversity
eco
ecological
ecosystem
electricity
emission
energy
environment
environmental
environmentally
extinction
freshwater
fuel
ghg
green
greenhouse
groundwater
natural
non-renewable
organic
ozone
ozone-depleting
pollutant
pollution
rainwater
recycle
renewable
resource
warming
waste
water

Social and employee matters
absenteeism
citizenship
community
cultural
culture
dialogue
donate
donation
employability
employee
employment
ethics
fatalities
friendly
gender
health
labor
labour
multi-stakeholder
non-discrimination
personnel
philanthropy
retirement
safety
school
social
training
union
user-friendly
volunteer
woman
women
worker
workforce
workplace

Respect for human rights
asylum
attack
childhood
color
colour
cruel
destruction
detention
dignity
disability
discrimination
education
equal
equality
exile
family
free
freedom
honor
honour
human
inhuman
liberty
marriage
moral
motherhood
nationality
peace
peaceful
persecution
personality
privacy
protection
punishment
race
religion
servitude
slave
slavery
society
tolerance
torture
widowhood

Anti-corruption and bribery matters
anti-corruption
bribe
bribery
condemn
corrupt
corruption
crime
dishonest
unethical
whistleblowing

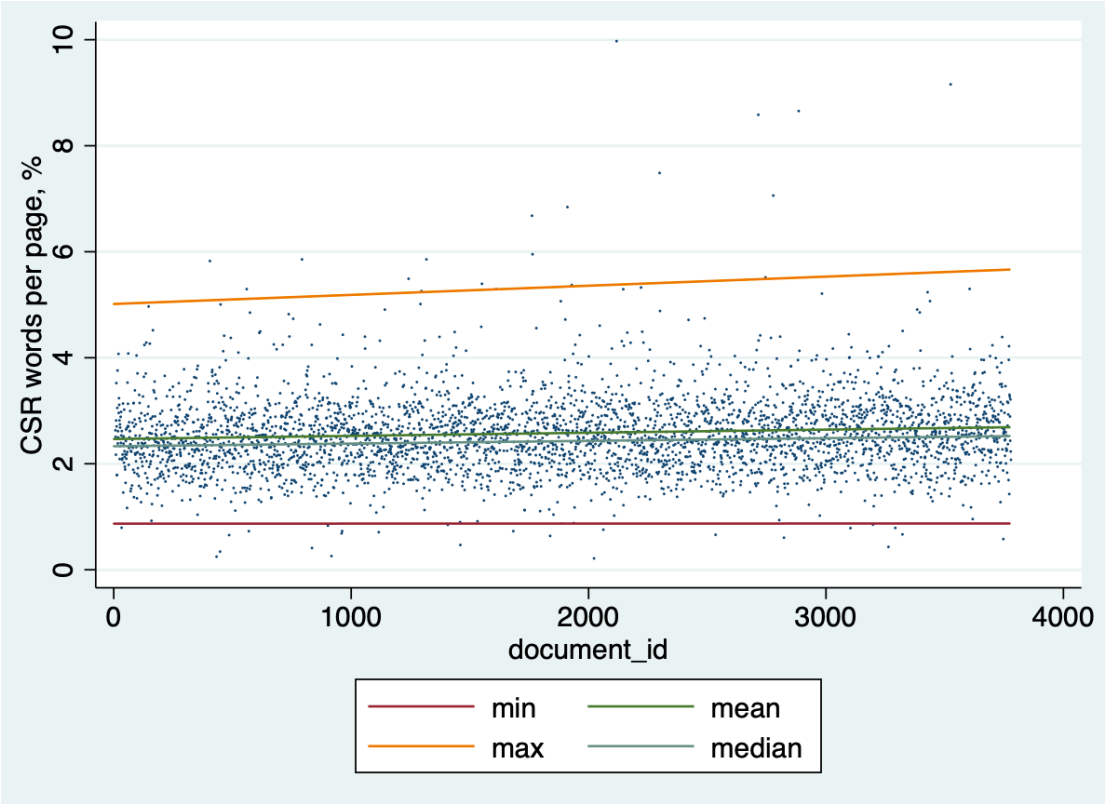
Appendix 2: Strategic and forward-looking vocabulary

foreseeable	coming year
future	following year
long-term	incoming year
long-run	next year
outlook	subsequent year
strategic	upcoming year
objective	long run
target	looking ahead
benchmark	k year(s), where k ranges from 1-20 or from ‘one’ to ‘twenty’
commitment	
prospect	
forward-looking	

Appendix 3: Automatized extraction of CSR sections from annual reports

Algorithm:

1. Conversion of .pdf files to .txt format
2. Computation of the distribution function of CSR vocabulary per 20 sentences (an average equivalent of one page)



Plot 1. Distribution of CSR vocabulary

3. Pages containing more than 1.5% of CSR words were extracted as being part of CSR sections. The vocabulary used for the identification of sections was constructed based on CSR standards and frameworks such as AA1000, UNGC, GRI, ISO14001, SA8000, SASB, SDG, Directive 2014/95/EU, OECD. The final set of words is presented in Appendix **1**.

Appendix 4: Details on parsing process (Python)

1. **Density**

Split text into sentences with “nltk sent_tokenize”, split each sentence into words with “nltk word_tokenize”, transform to lower case with “.lower”

2. **Report length**

Split text into words with “nltk word_tokenize”, remove punctuation, remove digits

3. **Tone**

Apply “TfidfVectorizer” from sklearn to a list of all documents that splits texts into words, transforms to lower case and removes punctuation. TF-IDF stands for “term frequency-inverse document frequency”. This module assigns a weight to each token which depends on its frequency in a document and in the entire corpora

4. **Readability**

Apply “textstat” from textstat to each document splitting text into words and transforming to lower case

5. **Topic coverage**

Apply “TfidfVectorizer”

6. **Forward-looking and long-term oriented vocabulary**

Apply “TfidfVectorizer”

7. **Similarity**

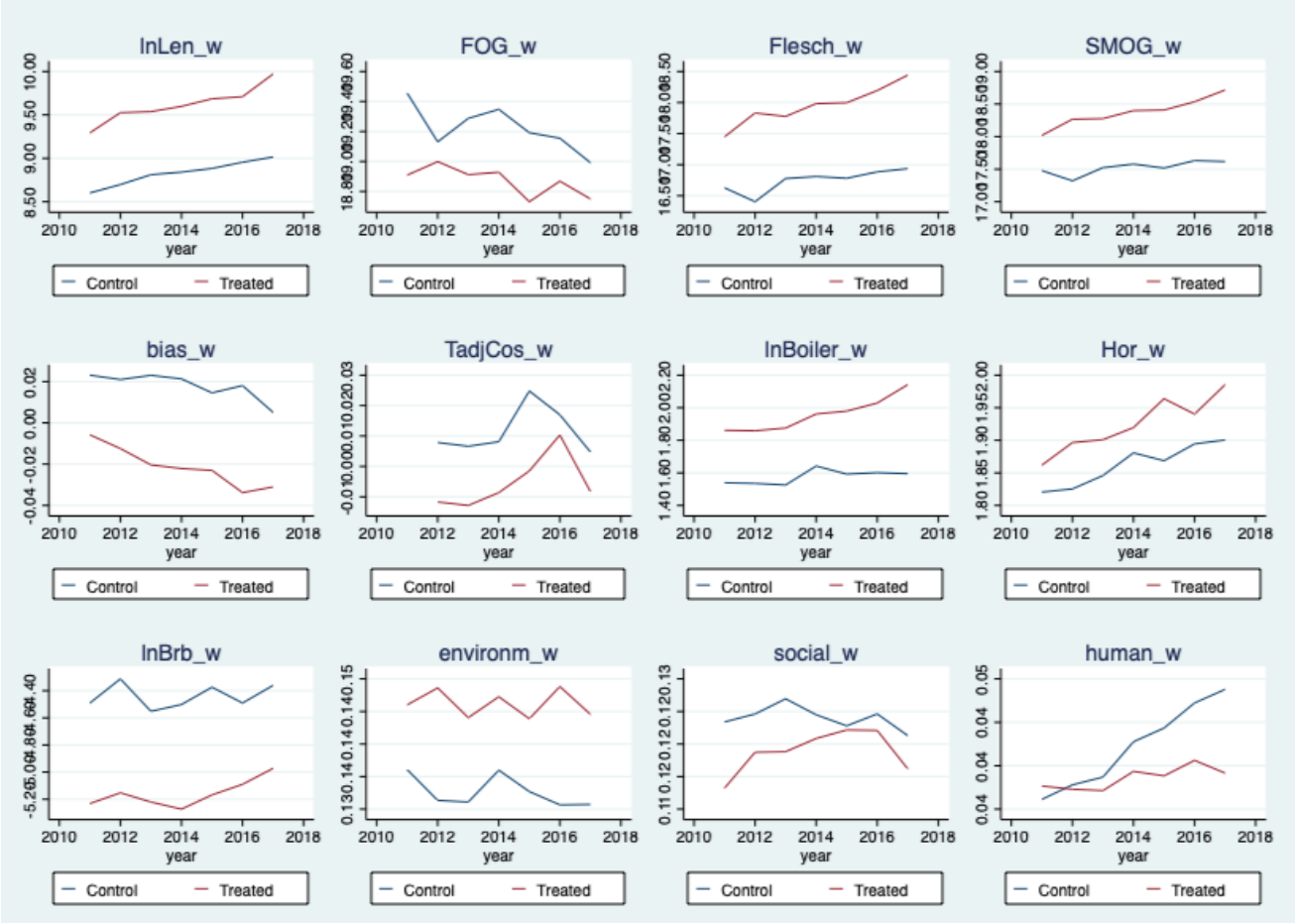
Apply “CountVectorizer” from sklearn to a list of all documents that splits texts into words, transforms to lower case and removes punctuation. This module assigns a weight to each token equal to its frequency in a document

Appendix 5: Variable descriptions

Parameter	Variable	Computation	Measurement	Denomination before transformation	Transformation	Denomination after transformation
Tone	Bias	%Positive words - %Negative words	Percentage	Bias	winsor2, cuts (5,95); idf	bias_w
Readability	FOG, SMOG, Flesch	f(complex words, words per sentence)	Years of education	FOG, SMOG, Flesch	winsor2, cuts (5,95)	FOG_w, Flesch_w, SMOG_w
Report length	Length	Split text into words with python nltk module	Number of words	Length	winsor2, cuts (5,95); ln	lnLen_w
Topic coverage	Environmental, social, bribery, human rights	Percentage of words from topic lexicon	Percentage	Environmental, social, bribery, human rights	winsor2, cuts (5,95); idf	environm_w, social_w, lnBrb_w, human_w
Boilerplates	Boiler	Four extremely common words among documents	Percentage	Boilerplate	winsor2, cuts (5,95); ln	lnBoiler_w
Horizon and strategic vocabulary	Horizon	Percentage of words from horizon lexicon	Percentage	Horizon	winsor2, cuts (5,95); idf	Hor_w
Similarity of texts over time	Cosine similarity	Cosine between vectors of word counts	Range of values from 0 to 1	Similarity	winsor2, cuts (5,95), Taylor adjustment	TadjCos_w

Graph 1: Parallel Trend Assumption

These graphs are built for visual inspection of the dependent variables ensuring the validity of the DID. The difference between the treatment and control groups prior to 2014 should be constant. Based on this representation we exclude FOG index, environmental topic and human rights matter from our DID.



See Appendix 5 for a description of variables