Clapham, Benjamin; Bender, Micha; Lausen, Jens; Gomber, Peter

Policy making in the financial industry: A framework for regulatory impact analysis using textual analysis

Journal of Business Economics

Provided in Cooperation with:
Springer Nature

Suggested Citation: Clapham, Benjamin; Bender, Micha; Lausen, Jens; Gomber, Peter (2022) : Policy making in the financial industry: A framework for regulatory impact analysis using textual analysis, Journal of Business Economics, ISSN 1861-8928, Springer, Berlin, Heidelberg, Vol. 93, Iss. 9, pp. 1463-1514,
https://doi.org/10.1007/s11573-022-01119-3

This Version is available at:
https://hdl.handle.net/10419/312313

Journal of Business Economics (2023) 93:1463–1514
https://doi.org/10.1007/s11573-022-01119-3

ORIGINAL PAPER

Policy making in the financial industry: A framework for regulatory impact analysis using textual analysis

Benjamin Clapham[1] · **Micha Bender**[1] · **Jens Lausen**[1] · **Peter Gomber**[1]

Accepted: 11 October 2022 / Published online: 9 November 2022

Abstract
Regulators conduct regulatory impact analyses (RIA) to evaluate whether regulatory actions fulfill the desired goals. Although there are different frameworks for conducting RIA, they are only applicable to regulations whose impact can be measured with structured data. Yet, a significant and increasing number of regulations require firms to comply by specifying and communicating textual data to consumers and supervisors. Therefore, we develop a methodological framework for RIA in case of unstructured data following the design science research paradigm. The framework enables the application of textual analysis and natural language processing to assess the impact of regulatory actions that result in unstructured data and offers guidance on how to map suitable methods to the dimensions impacted by the regulation. We evaluate the framework by applying it to the European financial market regulation MiFID II, specifically the recent regulatory changes regarding best execution. Thereby, we show that MiFID II failed to improve informativeness and comprehensibility of best execution policies.

Keywords RegTech · Regulatory impact analysis · Unstructured data · Textual analysis · Natural language processing · Design science

JEL Classification G28 · K20

1 Introduction

Regulation of the financial industry and of financial markets is a fundamental tool of governments and policy makers to ensure customer and investor protection as well as market efficiency and market integrity. In order for these tools to be effective and to achieve high regulatory quality, it is critical to ensure that regulatory adjustments and new regulations in the financial industry actually meet their desired objectives

✉ Benjamin Clapham
clapham@wiwi.uni-frankfurt.de

[1] Goethe University Frankfurt, Theodor-W.-Adorno-Platz 4, 60323 Frankfurt, Germany

and result in the intended changes. The growing pace of technological progress and the increasing interdependencies between different financial regulations pose substantial challenges to policy makers and regulatory quality since the exact effects of a regulation are hard to assess. Therefore, analyzing the impact of regulatory actions is a crucial step for evidence-based policy making.

For this purpose, policy makers and regulators around the globe conduct regulatory impact analysis (RIA) to evaluate whether regulatory actions meet the desired goals. Although there exist different guidelines and frameworks for conducting RIA (e.g., OECD 1997; Radaelli 2004), they are only applicable to regulations whose impact can be measured with structured and quantifiable data. Yet, an increasing and significant number of regulatory actions aim at or result in vast amounts of documents representing textual data that is hard to evaluate manually. In the financial industry, regulatory actions aimed at unstructured data mainly result from increasing disclosure requirements such as additional prospectus requirements for mutual funds in the US (U.S. Securities and Exchange Commission 2009) or new requirements for key information documents for retail investment products in the European Union (European Parliament and Council 2014b).

Regulators become increasingly aware that they need process guidelines and information technology (IT) based solutions to deal with and analyze the masses of reports and textual data. Such innovative IT-solutions are also known under the term RegTech, and are already used by firms to manage their regulatory requirements, e.g., by automating regulatory reporting (Butler and O'Brien 2019). However, RegTech and RegTech literature comes short concerning the supply of appropriate IT-enabled solutions and methodologies for regulators. Yet, such solutions are necessary to assess the impact and effectiveness of regulations that result in massive amounts of unstructured data. They can also serve to assess the compliance of firms to those regulations. Systems and methodologies are needed to appropriately monitor and analyze regulatory documents that firms have to deliver. The analysis of these documents enables regulators to assess the effectiveness of their efforts to ensure economic stability, fair competition, and market integrity (Arner et al. 2017).

To enable researchers and regulators to assess the impact of regulatory actions aimed at unstructured data and to improve future evidence-based policy making with the help of regulatory intelligence and RegTech-solutions, this paper develops a methodological framework for RIA in case of unstructured data (also referred to as RIA-framework hereafter), which builds on methods from textual analysis (TA) and natural language processing (NLP)[1] (e.g., Loughran and McDonald 2016). Those methods as well as improvements in data processing and aggregation help to make unstructured data quantifiable and, thus, provide the methodological foundation for our framework.

[1] Textual analysis refers to the broad field of methods and tools to automatically extract the quantity and quality of information in a collection of text. Frequently, textual analysis falls into the categories of targeted phrases, sentiment analysis, topic modeling, readability analysis, or measures of document similarity (Loughran and McDonald 2016). Natural language processing is the subfield of computer science that uses artificial intelligence to learn and understand content in human language (Hirschberg and Manning 2015).

For the development of the RIA-framework, we follow the design science research paradigm, which aims to create methods, tools, and other artificial objects that meet pre-defined goals and provide utility to their users (Simon 1996). Specifically, we adhere to the guidelines for design science research by Hevner et al. (2004) and follow the methodology by Peffers et al. (2007), which builds on these guidelines. The RIA-framework provides an innovative and effective solution for an important practical problem, which are the crucial characteristics of a design science artifact to be a relevant contribution (Geerts 2011; Hevner et al. 2004). The RIA-framework details the necessary steps for the application of TA and NLP to assess both (i) the achievement of regulatory objectives and (ii) compliance of firms with the regulation in case firms have to comply by setting up textual data. It also offers clear guidance on how to map suitable TA and NLP methods to the dimensions impacted by the regulation.

Following Peffers et al. (2007), we evaluate the RIA-framework based on a demonstration of its applicability in a use case of a recent financial market regulation where investment firms, i.e., banks and brokers, have to generate huge amounts of unstructured textual data. Specifically, we use the RIA-framework to assess the impact of the recently enforced changes in best execution requirements of the Markets in Financial Instruments Directive II (MiFID II) in Europe (European Parliament and Council 2014a) that has to be applied since January 2018. These rule changes demand investment firms to provide more informative best execution policies, which also should be easier to understand. In best execution policies, investment firms have to describe their processes of order handling and routing to achieve the best possible order execution for their clients. Thus, these policies should enhance transparency for investors and protect them from potential downsides of the stock market fragmentation in Europe.

The use case confirms that our RIA-framework serves to effectively evaluate the impact of a financial market regulation resulting in unstructured data. Moreover, it shows that the analyzed best execution requirements in MiFID II did not achieve the desired goals. By comparing textual similarity, specificity, and boilerplate information of policies from German institutions before and after MiFID II, we find that the informational value of these policies actually decreased rather than increased as intended by the regulation. Also, we find that these policies became harder to read and are more difficult to understand after the regulatory change. Based on a second and broader sample of European best execution policies, we apply the benchmarking approach proposed in the developed framework and compare the readability of European best execution policies with texts from different contexts and with varying levels of readability (e.g., European legislative documents, companies' annual financial statements, Wikipedia articles, spoken language). The analysis shows that—although intended to be understood by retail investors—best execution policies are among the most difficult and complex documents and that they are as hard to read as companies' annual financial statements or legislative documents. Consequently, the analysis of the new regulatory requirements on best execution policies in MiFID II shows that they did not reach the desired goals of increased investor protection and competition between brokers by providing more informative and easily understandable best execution policies.

Although the RIA-framework is developed against the background of the financial industry and demonstrated based on a use case from financial regulation, it can be applied to regulatory initiatives of other economic sectors as no step of the framework is unique to the financial industry. Rather, the RIA-framework represents a general principle to solve a class of real-world problems, i.e., conducting RIA in case of unstructured data.

This paper contributes to the literature streams of RegTech and RIA by equipping regulators and researchers with a new framework for conducting RIA in case of unstructured data based on methods from the fields of TA and NLP. The RIA-framework provides the necessary process steps, decisions, and data requirements, as well as the suitable methodologies to assess the impact of a regulation aimed at or resulting in unstructured data in an organized, largely automated, and objective manner. This research is one of the first studies that aims at using information systems (IS) research methodologies to support regulators to achieve their goal of assessing the effectiveness of regulations and to increase regulatory intelligence based on IT. It extends prevailing RegTech literature that, up to now, mainly focuses on RegTech for compliance by firms and for supervision by authorities (e.g., Butler and O'Brien 2019; Arner et al. 2016). Thereby, this paper also helps to close the gap between the relatively high usage of RegTech in the private sector and the still relatively low adoption of RegTech by regulators themselves (Arner et al. 2017).

The paper is organized as follows: Sect. 2 discusses literature regarding RegTech, outlines the research gap concerning RegTech for regulators and law-makers, and discusses the concept of RIA. Against this background and based on existing guidelines for RIA and methods from TA and NLP, we develop the framework for RIA in case of unstructured data in Sect. 3. Section 4 demonstrates the usability of our RIA-framework by applying it to assess the impact of regulatory changes for European best execution policies. Sect. 5 evaluates the proposed RIA-framework. We discuss our RIA-framework and findings in Sect. 6 and conclude in Sect. 7.

2 Literature review on RegTech and RIA

The RIA-framework contributes to the literature stream RegTech. Therefore, this section provides a short overview of relevant studies related to RegTech and outlines the lack of research on RegTech solutions supporting regulators and policy makers to improve regulatory intelligence. As a basis for the framework development, this section also discusses the concept of RIA, existing guidelines for RIA, and related research.

2.1 Regulatory technology

Regulatory Technology or RegTech refers to IT deployed in the context of regulatory compliance, reporting, and supervision. RegTech helps firms to manage their regulatory obligations (Butler and O'Brien 2019) and supports supervisory

authorities by enabling them to effectively monitor whether the economic activities of firms are compliant (Arner et al. 2017; Williams 2013).

RegTech supports *firms* to set up compliant business systems, to control risks, and to perform or automate regulatory reporting (Butler and O'Brien 2019). This advancement in IT adoption is strongly connected to the general technological change in the industry (Arner et al. 2017). Specifically for compliance management, the literature proposes many different use cases based on IT: For instance, Gozman et al. (2020) explore the potential of applying blockchain technology for regulatory reporting of mortgages or Moyano and Ross (2017) propose a new approach for the know-your-customer (KYC) due diligence process.

Besides helping firms to be compliant, RegTech and related research also support *supervisory institutions*. Thereby, RegTech enables supervisors to conduct more granular and effective supervision (Arner et al. 2016). This especially holds for financial markets, where supervisors have successfully used IT to monitor and analyze markets and market participants preventing insider trading, market manipulations, and fraud (Arner et al. 2016; Williams 2013; Siering et al. 2017). Furthermore, the literature proposes several applications of advanced methodologies such as predictive analytics and machine learning for supervisory institutions in the context of corporate fraud (Dong et al. 2018), credit card fraud (Bhattacharyya et al. 2011), accounting fraud (Kirkos et al. 2007; Glancy and Yadav 2011; Humpherys et al. 2011), and financial misconduct (Lausen et al. 2020).

However, while RegTech is quite advanced in supporting the compliance of firms and the respective investigations by supervisors, literature comes short concerning the supply of appropriate IT systems and methodologies for *regulators*, who need to assess and review whether their regulatory actions actually fulfill the desired goals. While there is a call for increasing the efforts regarding the evaluation of regulations' effectiveness by conducting RIA (Gai et al. 2019), regulators become increasingly aware that they need appropriate process guidelines and IT-solutions to automate the assessment of the masses of data provided in response to the reporting and disclosure requirements of firms (Arner et al. 2016, 2017). Thereby, IT and RegTech may unleash significant benefits for regulators to improve regulatory intelligence and to achieve their goal of facilitating a safe and resilient economic system based on RIA and evidence-based policy making.

2.2 Regulatory impact analysis

The primary goal of RIA is the optimization of policy making by ensuring that benefits to society from regulatory actions are maximized while costs, i.e., potential negative consequences, are minimized (OECD 1997). The (OECD 1997, p. 7) defines RIA as an approach for "systematically assessing the negative and positive impacts of proposed and existing regulations". All OECD member states (38 countries as of 2022) as well as the European Commission have adopted some form of RIA for their legislative processes, both concerning primary laws and subordinate regulations, to increase regulatory quality (OECD 2018). In particular, ex-post RIA improves the monitoring of existing

regulations and builds the basis for potential revisions or even complete cancellations of a regulation depending on the actual regulatory impact (Kirkpatrick and Parker 2004). Therefore, RIA is an analytical and systematic research and policy tool to assist decision makers in evidence-based policy making (OECD 2008).

Academic literature analyzes the impact of policy decisions based on RIA in various domains, however, it mostly aims at regulation that can be assessed by structured and quantifiable data (Radaelli 2004). The limitation to structured data also holds for existing frameworks for RIA such as the general guidelines for systematic impact assessment for OECD member states (OECD 1995), the more specific framework of the European Commission (2005), and the various improvements of OECD frameworks and guidelines (OECD 2008, 2018, 2020). These improved as well as newly developed guidelines and frameworks entirely consider the assessment of regulatory actions based on the analysis of structured and quantifiable data. Yet, research and existing frameworks come short in providing solutions for analyzing the regulatory impact of policies that aim at or result in unstructured data. However, as more and more policies target the creation and provision of unstructured data such as firms' disclosure requirements to customers or supervisors, the assessment of such regulations becomes increasingly relevant so that regulators and researchers need to be equipped with the necessary tools and frameworks. Data science methods like TA and NLP (e.g., different sentiment (Salton and Buckley 1988; Pierrehumbert 2001) and readability measures (Gunning 1969; Tan et al. 2002)) as well as improvements in data processing and aggregation (e.g., information content (Blei et al. 2003) and textual similarity (Jiang and Conrath 1997; Bag et al. 2019; Lau and Baldwin 2016) analyses) can help to make unstructured data quantifiable and, thus, serve as methodological foundation for RIA in case of unstructured data. Al-Ubaydli and McLaughlin (2017) provide first steps in this direction by developing a measure to quantify regulatory demands in published documents based on methods from TA. We go further and develop and evaluate a framework for RIA using IS research methods.

3 A framework for analyzing regulatory impact based on unstructured data

Following the design science research paradigm, we develop a framework for the analysis and evaluation of regulatory actions that result in unstructured data such as text documents. To create the artifact, we build on existing RIA guidelines and on methods from TA and NLP. The RIA-framework provides detailed guidance and the required steps and tools to analyze the impact of a regulation targeting at or leading to unstructured data in a systematic and largely automated manner. Figure 1 presents the proposed framework and the six steps that are necessary for RIA in case of unstructured data.

3.1 Step 1: problem and goal identification

The first step of the regulatory impact analysis is the identification of the economic, social, or environmental *problems initiating a regulatory action* to put the goals of the new regulations or regulatory adjustments into perspective and to enable their precise evaluation. To receive a profound knowledge base of the issue at stake, it is crucial to identify the key aspects of the problem, reduce conceptual uncertainty, and estimate the potential impact of the problem (European Commission 2005). Once the problem is precisely described, the reasons leading to the problem and its magnitude have to be examined by also taking into account the different entities being affected and their connections to the problem (OECD 2008). This critical in-depth-analysis of the problem addressed by the regulatory action should be based on political statements at the origin of a policy initiative, legal discussions, and academic studies. If the problem is identified and defined well, a sound understanding shall exist regarding the problem itself as well as regarding the reasons requiring regulatory actions to correct the identified problem (OECD 1997), which provides the foundation for a clear identification of the actual objectives of the regulatory action (OECD 2020).

A clear identification and formal definition of the regulatory objectives is essential in order to assess the accomplishments of regulatory actions since they are evaluated according to the initial *goals of a regulation* in light of the identified problem. Valuable information for the identification and description of regulatory objectives can often be directly derived from legal documents themselves. For instance, legislative documents in the European Union provide the reasons for the provisions at the start of every regulatory act in so-called recitals. Furthermore, other related legal documents such as regulatory consultations can be used to identify the objectives of a regulation. A detailed and explicit description of the regulatory objectives builds the basis for the evaluation of a regulatory action to verify whether it actually achieved its goals and objectives (European Commission 2005).

Step 1 Clearly describe the problem that the regulatory action wants to solve and identify the intended goals of the regulation.

3.2 Step 2: identification of affected dimensions

After the identification of the regulatory objectives, the specific dimensions that are affected by the regulatory action need to be determined. Dimensions in this context refer to the means (e.g., informativeness or objectivity) that are targeted by the regulation in order to achieve the identified regulatory goal. They refer to objects (e.g., regulatory disclosures) and subjects (e.g., companies). Thereby, it is crucial to derive the full set of affected dimensions taking into account all stakeholders (European Commission 2005) and to identify all relevant *direct* as well as *indirect* dimensions affected by the regulatory change (OECD 2020). The specific dimensions, objects, and subjects targeted by the regulatory action and the explanation of the

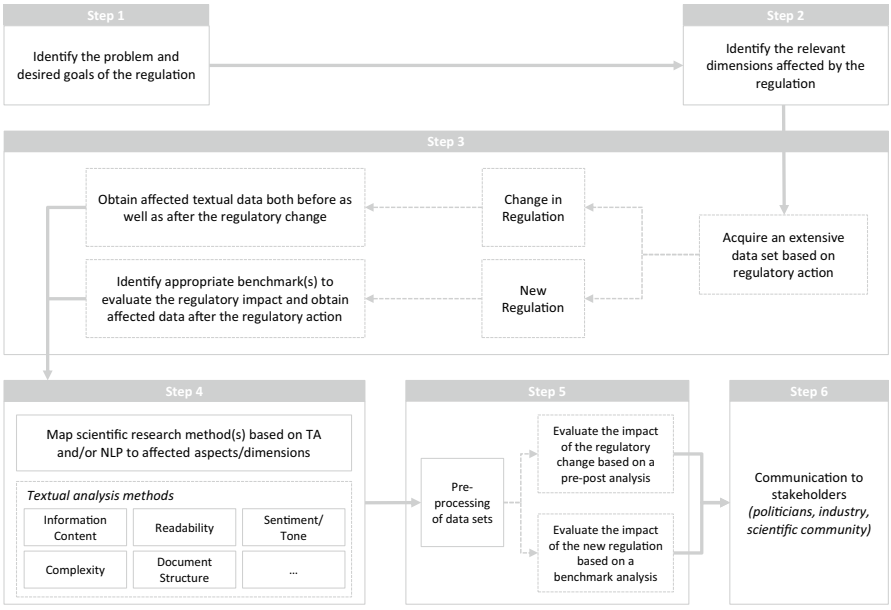

Fig. 1 RIA framework in case of unstructured data

corresponding operational changes can be extracted from the respective regulatory text, from related legal opinions, comments, and corresponding guidelines. Taking into account the perspective of different stakeholders and how they are affected by a regulation can support the identification of relevant dimensions (European Commission 2005). The identification of the targeted dimensions by a regulatory action is a crucial step in the RIA process since they build the basis according to which the impact of a regulation and thus its success or failure is evaluated.

Step 2 Identify the specific dimensions that are affected by the regulatory action.

3.3 Step 3: data acquisition

The third step of the RIA-framework outlines the data acquisition process and specifies the necessary data to assess the regulatory actions based on the derived dimensions determined in Step 2. To acquire an extensive data set for the assessment of regulatory impact, it is important to involve all relevant data holders and potential sources of unbiased data to guarantee that the RIA is conducted based on the most complete set of information (OECD 2020). The analysis and data acquisition approach for RIA in case of unstructured data differs dependent on whether a regulatory change or a new regulation is to be analyzed.

In case of the *revision of an already existing regulation*, the affected objects *before* as well as *after* the introduction of the regulatory change have to be collected. The inclusion of data from the pre-regulatory environment is important and

necessary to establish a baseline and reference against which potential changes are evaluated to ensure a sound assessment of the regulatory impact (OECD 2004). The collection of the affected objects *before* and *after* the regulatory action enables the assessment of the regulatory impact based on a *pre-post analysis* design. Thus, the impact and effects of the regulatory change can directly be derived from the changes in the dimensions targeted by the regulatory change. If possible, a control group should be used to exclude unobserved effects that might change over time independent of the regulatory change. An appropriate control group should not be affected by the regulatory action, however, jurisdictions should still be comparable.

In case of the *introduction of a new regulation*, the collection of any data before the regulatory change is mostly impossible for regulatory actions aimed at unstructured data since they often require firms to publish new textual documents which did not exist before. However, suitable references are necessary to assess the impact of new regulations aimed at unstructured data. Therefore, we propose a *benchmark approach* to assess the impact of such new regulations. To collect appropriate benchmarks (e.g., textual data generated in comparable regulatory areas), it is important to ensure comparability between affected objects and chosen benchmarks. The benchmarks should be selected considering the identified objectives and dimensions in Step 1 and Step 2. Furthermore, the background and area of the benchmarks should be matched to the affected objects as well as to the specific dimensions. Several benchmarks should be included in the data set to cover a wide range of different but comparable documents and to ensure an extensive analysis of the impacted aspects. In addition, each benchmark should contain sufficient data providing enough information to compare it with the objects targeted by the regulation.

Step 3 Acquire the necessary data for RIA dependent on whether a change in regulation (pre-post data set) or a new regulation (benchmark data set) is analyzed.

3.4 Step 4: map research method(s) to affected dimensions

In the fourth step, the affected dimensions identified in Step 2 need to be mapped with appropriate scientific research method(s) from TA and NLP that meet the requirements needed to examine the specific regulatory action.

The mapping of TA and NLP methodologies to affected regulatory dimensions is a major component for an effective RIA in case of unstructured data and needs to be adapted for each specific use case. However, the schematic sequence of the mapping process is similar. Starting with the mapping process, it is important to identify relevant methodologies suitable for the TA of the dimensions identified in Step 2. TA covers a wide range of different methodologies enabling the extraction of various information from textual data (Lacity and Janson 1994) and a variety of methods have been used in many different research areas. Focusing on techniques enabling an effective assessment of regulatory impact in case of unstructured data, we provide a set of common TA and NLP methodologies. Table 1 presents an overview of these methodologies for RIA in case of unstructured data. This collection of potential

Table 1 Overview of different TA and NLP measures, methods, and tools that can be used to classify and analyze textual documents

Measures/methods/tools	Reference	Description
Readability		
Fog Index	Gunning (1969)	The Fog index estimates the readability of a document based on the number of words with three or more syllables
Modified Fog Index	Kim et al. (2019)	The modified Fog Index reclassifies context-specific common words with three or more syllables as two-syllable words
Flesch Reading Ease Score	Kim et al. (2019)	The Flesch Reading Ease Score estimates the readability based on the average sentence length and the average number of syllables per word
Automated Readability Index (ARI)	Hu et al. (2012)	The ARI estimates the readability of a text based on the number of its characters, words, and sentences
Bog Index	Bonsall IV et al. (2017)	The Bog Index captures the plain English writing attributes recommended by linguistic experts and highlighted in the SEC's Plain English Handbook
Average word length	Barrón-Cedeño et al. (2010)	A high average number of characters per word indicates that a text is difficult to read
Average words per sentence	Debortoli et al. (2016)	A high average number of words per sentence is a signal of low readability
Complexity		
Document length	Singhal et al. (1996)	Word count of the overall document used as a measure of textual complexity
File size	Loughran and McDonald (2014)	The file size of a document as a proxy for textual complexity
Conditional statements (CS)	Li et al. (2015)	The frequency of conditional statements in a text is used to measure its cyclomatic complexity
Number of unique bigrams	Tan et al. (2002)	The number of unique bigrams captures the extend of textual variation within a document
Entropy of words	Bommarito and Katz (2010)	Entropy measures the diversity of language and concepts within a document
Sentiment & targeted phrases		
Term weighting	Salton and Buckley (1988)	Word count set in proportion to, e.g., importance of a word to a text corpus
Map analysis	Carley (1993)	Map analysis compares documents in terms of topics and the relationships between them
Word frequency	Pierrehumbert (2001)	Frequency of words in a document
Word lists/"bag-of-words"	Esuli and Sebastiani (2006)	Targeted phrases or dictionaries compare documents according to specific word lists or word combinations

Table 1 (continued)

Measures/methods/tools	Reference	Description
Textual similarity		
Semantic similarity	Jiang and Conrath (1997)	Semantic similarity measures similarity between words and documents by combining lexical taxonomy structure with a corpus of statistical information
Cosine similarity	Korenius et al. (2007)	Cosine similarity measures similarity between documents based on the frequency of terms within each document
Euclidean distance	Lee et al. (2012)	Euclidean distance measures distance between words based on the distance of words within a document
Jaccard similarity	Bag et al. (2019)	Jaccard similarity measures similarity between words based on the number of common words over all words
Numerical conversion of documents		
tf-idf weights	Ramos et al. (2003)	The term frequency-inverse document frequency (tf-idf) weights are used to reflect how important a word is to a document in a collection or corpus
word2vec	Mikolov et al. (2013)	The word2vec algorithm uses a neural network model to learn word associations from a large corpus of text
doc2vec	Le and Mikolov (2014)	The doc2vec algorithm builds on the word2vec algorithm but accounts for the semantics of a sentence
BERT	Devlin et al. (2018)	Bidirectional encoder representations from transformers (BERT) is designed to pre-train deep bidirectional representations from unlabeled text
Information content & topic modeling		
K-means	MacQueen et al. (1967)	K-Means can be used to classify documents into disjoint topics by calculating the cosine similarity of certain documents to a cluster centroid
Latent semantic analysis (LSA)	Landauer et al. (1998)	LSA classifies a document by reducing the term-document matrix using singular value decomposition
Latent dirichlet allocation (LDA)	Blei et al. (2003)	LDA classifies documents into a mixture of topics by calculating the probability distribution over the vocabulary for each topic
top2vec	Angelov (2020)	The top2vec algorithm leverages joint document and word semantic embedding to find topic vectors without having to specify the number of topics

Table 1 (continued)

Measures/methods/tools	Reference	Description
BERTopic	Grootendorst (2022)	BERTopic is a topic modeling technique that uses BERT embeddings to create clusters which result in easily interpretable topics since important words are kept as descriptions
Named-entity recognition (NER)	Finkel et al. (2005)	The NER algorithm seeks to locate and classify named entities in documents into pre-defined categories
Specificity	Hope et al. (2016)	The number of numerical values, dates, and entities identified by the NER algorithm scaled by the total words within a text
Boilerplate	Lang and Stice-Lawrence (2015)	The share of standard phrases in a document which are prevalent in many documents of a corpus and thus are unlikely to be informative

methodologies provides a guidance for RIA application, but can be extended with additional methodologies as necessary for the specific use case.

We group the TA and NLP methodologies in Table 1 into the categories *Readability*, *Complexity*, *Sentiment & Targeted Phrases*, *Textual Similarity*, *Numerical Conversion of Documents*, and *Information Content & Topic Modeling*: First, the category *Readability* summarizes methodologies that can be used to measure textual difficulty and the required ability of a reader to understand the content of documents. Second, the category *Complexity* contains measures that reflect the complexity and diversity of language in a text. Third, the category *Sentiment & Targeted Phrases* represents methodologies targeting certain words or phrases through word lists or dictionaries, which can be connected to specific contexts or common sentiments. Fourth, the category *Textual Similarity* includes methodologies providing insights on the similarity/distance of terms and documents. Fifth, the category *Numerical Conversion of Documents* describes methods to convert a text to a numerical representation, which is a crucial step especially for the analysis of topics and textual similarity. And last, the category *Information Content & Topic Modeling* contains methodologies allowing the identification and extraction of topics within a collection of documents. For further methodological overviews in specific contexts, we refer to the literature (Aggarwal and Zhai 2012; Reshamwala et al. 2013; Loughran and McDonald 2016; Kang et al. 2020).

Once suitable TA methods are extracted from scientific research, it is important to select the appropriate measures for the regulatory assessment and to ensure that the selected methods actually measure the impact on the affected dimensions. Based on the overview of TA and NLP methodologies for RIA in case of unstructured data, we provide a first mapping of TA methods to corresponding regulatory dimensions in Table 2. While these mapped research methods already cover many potential regulatory dimensions, the application of the research methods still needs to be checked for each regulatory action individually and the mapping as well as the methodologies can be further extended based on the required needs of specific use cases. We propose to trigger an academic debate on further suitable mappings as a future research step. Within the RIA-framework, the assignment of TA techniques to specific regulatory dimensions provides the methodological basis for the evaluation of textual data and enables the assessment of regulatory actions concerning specific affected objects.

Step 4 Select appropriate TA and NLP methods and map them to the dimensions affected by the regulation.

3.5 Step 5: analyze and evaluate the impact of the regulatory change

This step of the RIA-framework comprises the actual analysis and evaluation of the regulation's impact. Depending on the structure of the respective data (e.g., simple text files, documents in PDF format, textual information from web pages, XML, JSON), several preprocessing steps need to be performed to make the data machine-readable. Thereby, relevant textual information needs to be extracted

Table 2 Overview of dimensions affected by regulatory actions mapped to suitable TA and NLP methods

Affected dimension	Method & relevant literature	Description/example of application
Differentiation	Similarity measures Hoberg and Phillips (2016)	If a regulation requires firms to publish more specific descriptions regarding their product offerings or business model, textual similarity can be used to find competitors and the level of competition by clustering documents according to their similarity. Furthermore, within a group of similar firms textual similarity can also be used as a measure of overall product differentiation between firms.
Informativeness	Similarity measures Hanley and Hoberg (2010)	Similarity between documents indicates whether a new document provides additional information compared to other documents—higher similarity indicates lower informativeness of new documents. Similarity as a measure of informativeness is in particular useful for RIA if a regulation requires additional information disclosure for public reportings.
	Information content Huang et al. (2018)	Information content methodologies such as topic modeling or named entity recognition can be used to extract the thematic content or relevant entities of documents. Comparing documents can then give indications whether a new document provides new thematic content or different entities.
Objectivity	Sentiment/targeted phrases Pak and Paroubek (2010)	When the regulatory goal is to reduce subjectivity in documents, methods of sentiment analyses can be applied to evaluate the tone of documents. In specific, methods differentiating objective and subjective text can be used to provide indications/levels of objectivity in documents.
Relevance	Information content Raghuveer (2012)	Measures of information content enable the evaluation of specific subjects within texts. For instance, if regulatory authorities require certain parties to mention specific relevant topics/entities within a regulatory disclosure document, information content measures can be applied to retrieve specific information in a document.
Comprehensibility	Readability & complexity measures Li et al. (2015), Crossley et al. (2011)	Readability measures can be applied to evaluate how easily a document can be understood. Furthermore, complexity measures aim to analyze how easily the text of a document can be processed. While readability and complexity measures analyze different characteristics and aspects within a document, both measures can be used to evaluate the comprehensibility as a regulatory goal.

and the handling of textual information from figures, tables, and lists needs to be determined. In addition, and depending on the methodology used for the analysis, further text cleaning steps such as removing stopwords and numerical characters or stemming need to be performed. Because preprocessing depends strongly on the data, goal, and methodology of the analysis, we do not provide a technical overview of different preprocessing techniques but refer to the literature (e.g., Kannan et al. 2014; Vijayarani et al. 2015; Kathuria et al. 2021).

Once preprocessing is completed, the assessment of the regulatory action can be started. At first, the TA and NLP methods determined and mapped to the affected regulatory dimensions in Step 4 are applied. In case the goal of the RIA is to assess a change in regulation, pre-post analysis is conducted and textual data affected by the regulatory change is analyzed using the selected methods on samples both before and after the regulatory change went into force. In case the analysis is conducted to evaluate the regulatory impact of a new regulation, benchmark analysis is performed and textual data resulting from the regulation is compared to suitable benchmarks using the selected TA and NLP methods.

Once these analyses are conducted, the results need to be compared, e.g., by using data visualization and statistical tests. In doing so, researchers and regulators should evaluate whether changes in the analyzed measures can be observed and whether these changes correspond to the regulatory objectives (defined in Step 1). Moreover, they should evaluate whether any undesired effects can be detected. The evaluation of the results should indicate whether the regulator has achieved the objectives of the regulatory action and whether further regulatory requirements might be necessary if the objectives are not met.

Step 5 Conduct necessary data preprocessing steps and analyze the regulatory impact based on the obtained data and the selected TA and NLP methods.

3.6 Step 6: communication to stakeholders

The final step of the framework represents the communication of the results of the RIA to relevant stakeholders, e.g., policy makers, regulators, reporting standards providers, industry and consumer protection associations, and the scientific community, by publishing a policy white paper or a research report. The report should briefly address each step of the framework, describe the TA and NLP methods used to analyze the impact of the regulation, summarize the key findings of the analysis, and elaborate on whether the regulation achieved the desired goals. Moreover, the results of the RIA can be used to discuss potential improvements of the analyzed regulation and the report can elaborate on potential further regulatory actions. The steps of our RIA-framework can be used to structure the report.

Step 6 Communicate the results of the RIA to relevant stakeholders.

4 Framework application: assessment of the change in best execution requirements in MiFID II

To demonstrate its applicability, we make use of the RIA-framework and perform an ex-post regulatory impact analysis of the best execution requirements outlined in MiFID II (European Parliament and Council 2014a). According to Peffers et al. (2007), demonstrating the applicability of an artifact is an important step in design science research to evaluate how well the developed artifact provides a solution to the problem.

4.1 Problem and goal identification (Step 1)

With more than 300 trading venues as of September 2022,[2] the European securities market is highly fragmented. Consequently, there is a large choice of venues to which an order to buy or sell a stock or other financial instruments could be sent to. The selection of the appropriate trading venue to execute a specific order is one of the key tasks of investment firms to obtain the best possible result for their clients taking into account a range of factors such as price, costs, speed, likelihood of execution ("best execution"). Already in MiFID I that went live in November 2007, the European regulator defined principles for investment firms concerning best execution and required investment firms to publish so-called best execution policies,[3] which describe their processes to achieve best execution (European Parliament and Council 2014a). The best execution regime allows investment firms to implement individual approaches and strategies (execution arrangements) to achieve best execution in compliance with the statutory minimum requirements (Gomber et al. 2012). These execution arrangements are summarized in best execution policies. Yet, the actual implementation of execution policies revealed significant shortcomings since most policies are limited to minimum information, do not comprehensively describe the whole best execution process, and, most importantly, are difficult to understand (Gomber et al. 2012, MiFID II, Recital 97). With MiFID II, that went live in January 2018, European authorities intended to address these shortcomings. One crucial amendment of this revision is that execution policies are required to become more informative and comprehensible in order to provide value to clients (European Parliament and Council 2014a).

Specifically, MiFID II Recital 97 states that "*In order to enhance investor protection it is appropriate to specify the principles concerning the information given by investment firms to their clients on the execution policy [...]*". To achieve this goal, MiFID II Art. 27(5) includes a new paragraph requiring investment firms to specify their execution policies so that the provided "*information shall explain clearly, in sufficient detail and in a way that can be easily understood by clients, how orders*

[2] For the number of trading venues in Europe, see https://registers.esma.europa.eu/publication/searchRegister?core=esma_registers_upreg.

[3] As an example for European best execution policies, please refer to the policy of Deutsche Bank (available at https://www.db.com/legal-resources/order-execution-policy).

will be executed by the investment firm for the client" (European Parliament and Council 2014a).[4] Consequently, the primary desired goal of the regulatory change is to foster investor protection by increasing the informational value and ease of understanding of best execution requirements for clients. Moreover, the regulatory change also aims at increasing the competitive aspect of best execution policies (Committee of European Securities Regulators 2007; Laruelle and Lehalle 2018). Because the new requirements for best execution policies should provide investors a better understanding of how banks and brokers handle their orders, the policies can serve investors as a basis to select the investment firm that best suits their needs and preferences. In summary, the analyzed regulatory change in MiFID II aims to solve the problem of best execution policies being not informative and difficult to understand in order to achieve the desired goals of *investor protection* and *competition between brokers*.

4.2 Identification of affected dimensions (Step 2)

The crucial amendment with respect to best execution policies in MiFID II is that they are required to become more informative and comprehensible in order to reach the desired goals, i.e., strengthen investor protection and foster competition between brokers.[5] Consequently, two dimensions need to be analyzed in order to assess the impact of changed best execution requirements:

- Informativeness of execution policies (Dimension 1): Derived from the legal text, which states that execution policies need to *"explain clearly, in sufficient detail"* (MiFID II Art. 27(5)) how the broker handles clients' orders to achieve best execution.
- Comprehensibility of execution policies (Dimension 2): Again derived from the legal text, which says that the policies should be *"easily understood by clients"* (MiFID II Art. 27(5)).

4.3 Data acquisition (Step 3)

According to the RIA-framework, data acquisition depends on whether a new regulation or whether a change in regulation has to be analyzed. Because the amendments to best execution requirements in MiFID II represent a change in regulation,

[4] In this context, also the so-called level 2 documents supporting the regulation were clarified. Specifically, the provisions for how the execution policies should be designed and articulated in MiFID I Art. 21 and Art. 46(2) of its accompanying implementing Directive 2006/73/EC were revised within MiFID II and Art. 66 of the Delegated Regulation (EU) 2017/565 to clarify the requirements for investment firms' execution policies.

[5] The content that banks and brokers are required to disclose in best execution policies did not change from MiFID I to MiFID II. Under both regulatory regimes, banks and brokers need to obtain "the best possible result for their clients taking into account price, costs, speed, likelihood of execution and settlement, size, nature or any other consideration relevant to the execution of the order" (MiFID I, Article 21(1); MiFID II, Article 27(1)) when executing clients orders.

we need to obtain data on the affected dimensions both before as well as after the regulatory change (see Case 1 below). Nevertheless, and as the benchmark approach is a major contribution of the paper, we also evaluate the impact of MiFID II on best execution policies as if it was a new regulation (see Case 2 below). Moreover, this alternative approach can also be used in case data before the regulatory change cannot be obtained, which, however, should regularly not be the case since regulators as the main potential users of the framework can request the necessary documents from the regulated entities (here: investment firms).

Case 1: change in regulation (MiFID I to MiFID II)

Because the amendments to best execution requirements in MiFID II represent a change in regulation, we need to obtain data on the affected dimensions both before and after the regulatory change. Consequently, best execution policies of banks and brokers before as well as after the application of MiFID II need to be obtained. Specifically, we build upon the execution policy examination of Gomber et al. (2012)[6] analyzing 75 execution policies of the largest German financial institutions and online brokers written in German from 2009. These policies are then matched with the corresponding firm's execution policies post-MiFID II from 2020. Mergers and acquisitions as well as insolvencies within the time from 2009 to 2020 reduce the sample of the analysis to 50 firms. Thus, the final data set includes a total of 100 execution policies aimed at retail clients (50 from 2009 and 50 from 2020).[7]

Case 2: MiFID II as new regulation

For this second part of the analysis, we collect execution policies from trading members (i.e., banks and brokers) of the largest European stock exchanges.[8] The execution policies valid as of May 2020 are downloaded from the investment firms' websites provided that an English version is available to ensure comparability across the different countries. Because MiFID II requires banks and brokers to account for the different characteristics and needs of retail and professional investors (MiFID II Art. 27(9a)), we follow this differentiation and sort the policies in these two groups, i.e., retail and professional clients. This results in a total of 124 execution policies addressing retail investors and 167 execution policies applying to professional investors.[9]

[6] We thank the authors for providing us the best execution policies included in their sample.

[7] We use execution policies from 2020, i.e., two years after the applicability of MiFID II so that the post-sample fits to the pre-sample obtained of Gomber et al. (2012) from 2009, i.e., two years after the introduction of the previous regulation in MiFID I. As banks and brokers only provide the currently valid execution policy to the public due to compliance reasons, no history of policies is available to analyze changes over several years. Yet, regulators could request and obtain these policies when conducting such an analysis.

[8] LSE, Xetra, Euronext Paris, Six Swiss Exchange, Nasdaq Nordic, Borsa Italiana (now Euronext Milan), Bolsa de Madrid, Oslo (now Euronext Oslo), Luxembourg, Warsaw.

[9] Some best execution policies apply to both retail and professional investors, so that they are included in both samples. We repeat our analysis by separating the policies into three groups (i.e., retail, professional, and retail & professional).

For the benchmark analysis, we choose five different benchmarks from different contexts with varying textual complexity to be able to evaluate the readability and complexity of best execution policies against these benchmarks. Specifically, we (i) use the textual content of the Management Discussion and Analysis (MD&A) section of US Form 10-K filings. 10-K filings represent a standardized form of listed US companies' annual reports regulated by the Securities and Exchange Commission (SEC). The MD&A section has widely been used in the finance and accounting literature (e.g., Lundholm et al. 2014; Loughran and McDonald 2016). In our context, using 10-K filings is particularly interesting because similar to best execution policies, they represent reporting obligations of companies. We choose a random sample of 10-K filings for the year 2019 of 100 constituents of the S&P 500. Furthermore, we (ii) use the textual content of EU regulatory documents. Regulatory documents not only contain the rules for companies, especially regarding their reporting obligations, but also serve a as benchmark for high complexity texts due to their legal language and conditional statements. The textual content of regulations has been analyzed in various academic studies (e.g., Bommarito and Katz 2010; Katz and Bommarito 2014 for the United States Code). We concentrate on the key EU financial services legislation in the context of financial markets and market infrastructures (European Parliament 2020) and separate the text according to the chapters of the respective documents.[10] In addition, we follow Hassan et al. (2019) and (iii) include readability and complexity measures of spoken language based on the Santa Barbara Corpus of Spoken American English from Du Bois et al. (2000). We further follow Hassan et al. (2019) and (iv) use chapters of a standard financial accounting textbook (Libby et al. 2004) to cover general financial terms and financial jargon. And last, for general language we (v) use a random sample of 1000 Wikipedia articles.

4.4 Map research methods to affected dimensions (Step 4)

Mapping research methods from the fields of TA and NLP to the affected dimensions (identified in Step 2 of the framework) is one of the central steps in the assessment of regulatory actions aimed at or resulting in unstructured data. The first dimension to be analyzed is the impact of the regulation on the informativeness of best execution policies. In order to assess the informational content of the policies, we rely on three different measures: textual similarity, the percentage of boilerplate information, and specificity. Textual similarity analysis is an appropriate method which has already been applied in other studies analyzing the informational content and the amount of new information in documents (Hoberg and Phillips 2016; Kelly et al. 2018). If the policies only copy the legal text or copy from each other, the policies do not provide informational value to investors. The similarity analysis can reveal such a

[10] We include the Markets in Financial Instruments Directive (MiFID II), the Markets in Financial Instruments Regulation (MiFIR), the European Market Infrastructure Regulation (EMIR), and the Prospectus Regulation (European Parliament 2020).

relation. To measure textual similarity, we follow Hanley and Hoberg (2010) as well as Cohen et al. (2020) and compute the cosine similarity of two documents based on the frequency of terms within each document. When counting the terms in each document, we use stemming[11] and adjust the term frequencies by the inverse document frequencies to give more weight to terms that occur less often, i.e., we apply the term frequency-inverse document frequency (tf-idf) approach to compute cosine similarities. Since this word weight approach does not account for the structure of a sentence, we also calculate the cosine similarity based on the *doc2vec* model developed by Le and Mikolov (2014), which accounts for semantics and was already applied to the financial context by Reichmann et al. (2022).[12,13] The use of similarity analysis as a measure of informativeness is further justified by the regulator's intention to foster competition between brokers based on their execution policies (Committee of European Securities Regulators 2007; Laruelle and Lehalle 2018). Again, if policies are meant to provide a meaningful basis for broker selection to clients, the policies should contain the specific differences between the brokers and, hence, need to be heterogeneous. To account for the drivers of similarity and to measure informativeness based on the standardization of best execution policies (whereas highly standardized policies would account as less informative), we compute a boilerplate measure following Dyer et al. (2017) and Lang and Stice-Lawrence (2015). Boilerplate information is defined as standard text that is prevalent in many documents and thus is unlikely to be informative. Specifically, we measure boilerplate information by counting all tetragrams, i.e., groups of four words within a single sentence, for each policy. We aggregate the tetragram-counts of each policy and then create a list with tetragrams which occur in at least 30% of the policies. Thereby, the assumption is that the use of common phrases in at least 30% of the policies is boilerplate information since disclosure of such common information is unlikely to be firm-specific. The boilerplate measure is then calculated as the number of words in sentences that include at least one boilerplate tetragram divided by the total number of words of the document. In addition to textual similarity and boilerplate information, we analyze the specificity of all policies following Hope et al. (2016) and Dyer et al. (2017), where more specific information indicates a higher level of informativeness. We calculate the specificity measure as the number of entities (locations, people, organizations, currency amounts, percentages, dates, or times) representing specific information within a policy using the Stanford Named Entity Recognizer (NER)[14] divided by the total number of words in each policy.

The second dimension that needs to be analyzed in order to assess the new requirements in MiFID II is the comprehensibility of best execution policies.

[11] We rely on the Porter stemming algorithm implemented with Python's *nltk* library.

[12] Following Lau and Baldwin (2016) and Reichmann et al. (2022), we use a model configuration with distributed memory (dm) = 1 to capture semantic information, vector size = 300, window size = 5, down-sampling threshold = 1e-6, negative sampling = 5, and ignore words occurring less than 5 times.

[13] We use the *scikit-learn* library in Python to implement the tf-idf approach and to calculate the cosine similarity. For implementing the doc2vec model, we rely on the *gensim* library in Python.

[14] We implement NER with the *stanza* library in Python. For the German best execution policies, we use the German extension of NER.

According to the literature (e.g., Loughran and McDonald 2016) and Step 4 of our framework, the ease to understand written text can be investigated by analyzing readability and textual complexity measures. Consequently, the comprehensibility of texts is determined by a multitude of different text-inherent features, which need to be analyzed. Since these features cannot be easily summarized in a single readability or complexity measure, we investigate different features separately when analyzing the comprehensibility of best execution policies.[15] In order to measure the policies' readability, we rely on the average word length, the average number of words per sentence, and the modified Fog index. While the original Fog index has proven to work well in previous research, Loughran and McDonald (2014) argue that the Fog Index is incorrectly specified in some cases because a large number of multi-syllabic words in a given context, such as company in a finance context, can be well understood and are not a signal of textual difficulty. Therefore, we follow Kim et al. (2019) and calculate a modified Fog measure. Since their list of context-specific multi-syllable words from companies' annual reports (10-K filings) does not fit to our context, we create our own word list for the modification of the Fog index. First, we construct a list of common complex words that are not difficult to understand in the context of securities trading and best execution by using securities trading and best execution related documents provided by the regulator.[16] We extract all complex words of these documents and count the occurrence of each word. When the word occurs at least twice, we consider it as a context specific common word. Then, we calculated the modified Fog index by labeling the context specific common words as non-complex words. As an additional readability measure that does not rely on the identification of complex words, we calculate the Flesch reading ease score, which is another popular readability index (e.g., Li 2008; Kim et al. 2019).[17]

With respect to textual complexity, we analyze document length[18] as a proxy for the effort necessary to process the content of a policy. We also investigate the file size of the unprocessed text documents as proposed by Loughran and McDonald (2014). Further, we follow Li et al. (2015) and apply the concept of cyclomatic complexity, i.e., the frequency of using conditional statements in a document.[19]

[15] We use the standard Python libraries *NumPy* and *pandas* for the calculation of readability and complexity measures unless otherwise stated.

[16] For the German list of common complex words in our context, we use https://www.bafin.de/DE/Verbraucher/GeldanlageWertpapiere/Wertpapiergeschaefte/wertpapiergeschaefte_artikel.html, for the English word list, we rely on https://www.esma.europa.eu/sites/default/files/library/2015/11/07_320.pdf.

[17] The Flesch reading ease score is calculated as 206.835−(1.015 × words per sentence)−(84.6 × syllables per word) for English policies and as 180-words per sentence−(58.5 × syllables per word) for German policies. A higher Flesch reading ease score indicates that the policy is more readable.

[18] Document length and file size are only used for the pre-post analysis and not for the benchmark analysis. Since both measures are in absolute terms which vary across different document types, a comparison of the length or file size of best execution policies to the length or file size of other documents is not meaningful.

[19] We follow Li et al. (2015) and use the following conditional terms: 'if', 'except', 'but', 'provided', 'when', 'where', 'whenever', 'unless', 'notwithstanding', 'in no event' and 'in the event'. The analysis of cyclomatic complexity is only conducted for the policies written in English (Case 2) since there is no comparable dictionary for German texts.

Additionally, we analyze the entropy of words (Bommarito and Katz 2010; Katz and Bommarito 2014) and the number of unique bigrams to capture the extent of used terms and variations in vocabulary within a document.[20] To calculate the number of unique bigrams in a document, we use the set of all bigrams that appear at least once in a document. For normalization, we divide the measures cyclomatic terms and unique bigrams by document length.

4.5 Analyze and evaluate MiFID II's impact on best execution policies (Step 5)

Before generating quantitative linguistic features from the textual content of the execution policies, we perform several common text preprocessing steps. First, as execution policies are mostly in PDF format, we extract the textual content from the documents and delete lists and tables. We further remove stop words from the derived text. For documents in English language, we use stop words from the Natural Language Processing Toolkit (NLTK) (Bird et al. 2009). For German policies, we use the stop words list of the German BPW Dictionary (Bannier et al. 2019). Because execution policies typically include a large number of execution venues and the respective location of the venues' operators, we additionally exclude text that includes the name and location of these venues. We derive country and city names from an exhaustive list of "major cities of the world"[21] and all common forms of the respective exchange names from the execution policies themselves.[22] We also remove parts of the text that did not contain any relevant information, such as email addresses, website URLs, numbers, non-text characters, and single-character words. And last, we convert the text into lower case letters and split it into individual words.

Case 1: Longitudinal analysis based on pre-post comparison

Since we investigate a change in regulation, we perform a longitudinal analysis considering best execution policies before and after the introduction of MiFID II to gain insights on how new regulatory requirements concerning the investment firms' (subjects) information provision to their clients improved informativeness and comprehensibility (dimensions) of the execution policies (objects). For the analysis, we use the data set consisting of 100 execution policies (50 from 2009 and 50 from 2020) of the largest German banks and brokers obtained in Step 3.

Informativeness—case 1

First, we analyze the textual similarity of best execution policies by calculating the cosine similarity of the German policies in 2009 and 2020 respectively to investigate whether the regulation led to a change in similarity and thus informativeness of the policies. Table 3 provides the descriptive statistics for this analysis. The results clearly show that the regulatory action indeed led to a change of the policies'

[20] We use the *nltk* library in Python for the implementation of both measures.

[21] See https://datahub.io/core/world-cities.

[22] For the calculation of specificity and the associated NER tagging, we keep the execution venues' names and locations as well as the information on countries and cities because NER tagging enables to analyze these textual elements.

similarity. However, the change is contrary to the desired goals of the regulation. While the policies pre-MiFID II show an average tf-idf (doc2vec) cosine similarity of 0.60 (0.86), this score increases to 0.72 (0.89) post-MiFID II, which is also statistically significant as shown by the Wilcoxon Rank Sum (WRS) test. Furthermore, the results show that the best execution policies have a large portion of boilerplate information both pre- and post MiFID II with an average of 0.47 in 2009 and 0.44 in 2020. Also, specificity of the best execution policies remains low with an average of 0.04 in 2009 and 2020.

The increase in textual similarity is further demonstrated by Fig. 2, which visualizes the similarity of execution policies based on the tf-idf approach. Similar results are obtained when computing cosine similarities based on doc2vec (see Fig. 4 in the Appendix). Again, we see a substantial increase in similarity over the whole sample as all squares get darker, which is particularly true for the German savings banks (dark red part in the right corner). Specifically, all saving banks within the sample report the same content in their execution policy in 2020 in contrast to the disclosure of individual execution policies in 2009. Consequently, our results show that the regulatory change in MiFID II did not reach the desired goal of increasing informativeness of best execution policies and, therefore, this prevents that they can serve clients as a sound basis for broker selection. Instead, the policies became more homogeneous, remain relatively unspecific, and still include a large share of boilerplate information, mainly by reciting parts of the regulation,[23] which is not informative for clients and does not foster competition between brokers based on how they handle client orders to achieve best execution.

Additionally, we also examine the extent to which banks actually adjusted their best execution policies after MiFID II. Therefore, we measure the cosine similarity of the institutions' matched policies from 2009 and 2020 and compute the difference in the share of boilerplate and specific information. The results are reported in Table 4 together with differences in readability and complexity between the same institutions' policies. The high tf-idf (doc2vec) similarity scores of on average 0.84 (0.90) as well as the insignificant boilerplate and minor specificity score changes show that banks and brokers obviously did not substantially change their policies after the regulation, providing further indication that MiFID II failed to increase the informativeness of best execution policies.

Comprehensibility—case 1

Second, we investigate changes in the policies' comprehensibility due to MiFID II based on readability and textual complexity measures. Table 4 provides the differences in readability and complexity measures between matched polices of the years 2020 and 2009, i.e., we compare each bank's and broker's post-MiFID II policy with its pre-MiFID II policy. Descriptive statistics for pooled pre- and post-MiFID II best execution policies are provided in Table 8 in the Appendix. Also, Figs. 5 and 6 in

[23] The intensified reciting of the regulation was confirmed by manual inspection of the best execution policies.

Table 3 Descriptive statistics of similarity (tf-idf, doc2vec), boilerplate, and specificity scores for the 2009 and 2020 German best execution policies

	Year	Count	Mean	SD	Min	25%	50%	75%	Max	WRS
Cosine similarity	2009	2450.00	0.60	0.18	0.21	0.49	0.60	0.72	1.00	0.00
(tf-idf)	2020	2450.00	0.72	0.17	0.28	0.62	0.69	0.74	1.00	
Cosine similarity	2009	2450.00	0.86	0.08	0.63	0.81	0.86	0.90	1.00	0.00
(doc2vec)	2020	2450.00	0.89	0.07	0.69	0.85	0.89	0.92	1.00	
Boilerplate	2009	50.00	0.47	0.17	0.07	0.38	0.55	0.60	0.82	0.08
	2020	50.00	0.44	0.16	0.13	0.31	0.51	0.58	0.60	
Specificity	2009	50.00	0.04	0.04	0.00	0.02	0.02	0.04	0.17	0.06
	2020	50.00	0.04	0.02	0.02	0.03	0.03	0.04	0.09	

Count represents the number of observations, which is determined in the similarity analysis by $N \times (N-1)$, where N is the number of policies

the Appendix illustrates the distribution of the readability and complexity measures for the German policies pre- and post-MiFID II.

The readability and complexity analysis provides evidence that the comprehensibility of best execution policies did not improve after the regulatory change in MiFID II. The average number of words per sentence significantly increased by 1.11 words in 2020 suggesting that the policies are harder to read. This is further supported by a significant increase of the modified Fog by 0.60 in 2020. Also, the two readability measures average word length and the Flesch reading ease score did not change significantly indicating that MiFID II did not improve readability of best execution policies. The absence of improvements in readability becomes even more critical in light of the finding that best execution policies post-MiFID II have an average modified Fog index of 17.17 (see Table 8 in the Appendix), which classifies the majority of policies as unreadable and above the reading level of a college graduate (Li 2008). Generally, texts aiming at a wide audience[24] should have a Fog index less than 12 (Burke and Fry 2019). With respect to textual complexity, we observe a similar picture with no changes in the measures document length, file size, and relative bigrams, while entropy, and thus the diversity of language, significantly increases. Consequently, the analysis of readability and textual complexity shows that MiFID II also failed to improve readability and understandability of best execution policies.

With our sample of German best execution policies, we cover different bank types. Specifically, our sample consists of 20 savings banks (40% of the sample), eight state-owned banks (German "Landesbanken", 16%), five cooperative banks (10%), and 17 private universal banks (34%). Table 9 in the Appendix provides descriptive statistics per bank type for the years 2009 and 2020. The descriptive statistics show that the policies of the different bank types are on average highly

[24] E.g., leading magazines and newspapers such as the New York Times have a Fog index of around 11–12 (Burke and Fry 2019).

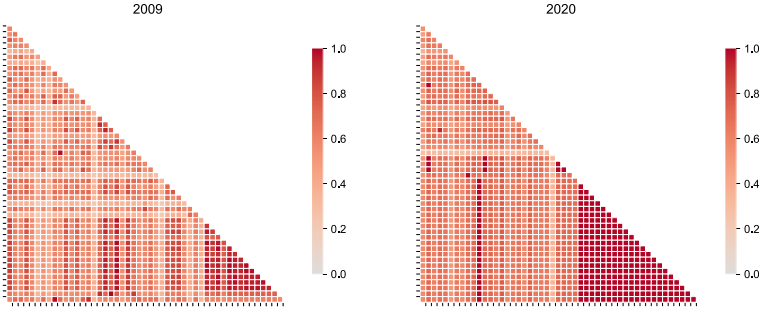

Fig. 2 Similarities (based on tf-idf) of German best execution policies of 2009 and 2020

Table 4 Descriptive statistics of the differences between readability and complexity measures as well as similarity of matched German best execution policies for the years 2020 and 2009

	Count	Mean	SD	Min	25%	50%	75%	Max	WRS
Readability measures									
delta_avg_word_len	50.00	0.00	0.20	−0.42	−0.12	−0.00	0.09	0.49	0.95
delta_wps	50.00	1.11	3.69	−9.07	−0.14	1.09	2.59	12.83	0.01
delta_modified_fog	50.00	0.60	1.91	−4.06	−0.41	0.20	0.98	6.30	0.08
delta_flesch	50.00	−0.79	5.69	−16.15	−3.50	−0.78	1.40	17.84	0.30
Complexity measures									
delta_doc_length	50.00	636.76	2347.45	−1893.00	−184.00	38.50	987.75	15824.00	0.12
delta_file_size	50.00	6.18	21.61	−16.92	−2.38	−0.11	8.98	140.66	0.13
delta_entropy	50.00	0.18	0.28	−0.26	0.02	0.08	0.33	0.85	0.00
delta_relbigrams	50.00	− 0.01	0.07	−0.22	−0.05	0.01	0.02	0.07	0.56
Informativeness									
cosine similarity (tf-idf)	50.00	0.84	0.07	0.70	0.80	0.83	0.89	0.95	–
cosine similarity (doc2vec)	50.00	0.90	0.04	0.79	0.88	0.91	0.93	0.97	–
delta_boilerplate	50.00	−0.03	0.12	−0.36	−0.09	−0.02	0.03	0.17	0.13
delta_specificity	50.00	0.00	0.03	−0.11	−0.01	0.01	0.02	0.07	0.08

Readability and complexity measures based on the pre-MiFID II policy are subtracted from those of the same institution's post-MiFID II policy. Cosine similarity compares the pre- and post-MiFID II policies of the same institution

comparable across almost all measures in both years. One exception is the document length of policies issued by cooperative banks, which—in 2009—is significantly shorter compared to the policies issued by other banks. However, the document length of cooperative banks' policies adjusts to similar values as the other banks' policies in 2020.

To validate the descriptive findings, we conduct the following pooled regression analysis to identify the impact of MiFID II on informativeness and comprehensibility of best execution policies:

$$Y_{i,t} = \alpha + \beta_1 \cdot MiFID\ II_t + \beta_2 \cdot log\_Employees_{i,t} + \beta_3 \cdot log\_TotalAssets_{i,t} + \sum_{k=4}^{6} \beta_k \cdot BankType_i + \varepsilon_{i,t} \quad (1)$$

Thereby, $Y_{i,t}$ accounts for each readability, complexity, and informativeness[25] measure of bank i's best execution policy in year t (2009 and 2020). *MiFID II*, our main variable of interest, is a dummy variable that equals one if the respective policy is from the year 2020, i.e., after MiFID II had to be applied. Further, we control for bank specific characteristics by taking the natural logarithm of the number of employees (*log_Employees*) and the natural logarithm of total assets (*log_TotalAssets*) into account. We also investigate whether the type of the bank that issued a specific policy has an effect on changes in readability, complexity, and informativeness. In Eq. (1), *BankType* represents dummy variables for the different bank types.[26]

Table 5 reports the results of this analysis. The regression analysis strongly supports the descriptive findings, i.e., all analyzed measures did not change or even worsened after MiFID II. Specifically, we observe a significant increase in the number of words per sentence (plus 1.39 words) and the modified Fog index (plus 0.64) of the best execution policies that were issued after MiFID II, which indicates that these documents are harder to read. We also find a significant increase of the policies' complexity according to file size (plus 5.51 kilobyte) and entropy (plus 0.15). Furthermore, the regression analysis confirms that banks' best execution policies became more similar after MiFID II with an increase in the tf-idf (doc2vec) cosine similarity of 0.10 (0.03) due to MiFID II relative to a pre-MiFID II average of 0.60 (0.86), while the share of boilerplate and specific information remained unchanged. Concerning a potential influence of the bank type, we see hardly any difference in changes of the policies' complexity between the different bank types. Regarding informativeness, we find evidence that particularly policies issued by savings banks lost informativeness relative to the policies of other bank types as three of the four informativeness measures are positive and significant indicating higher similarity and more boilerplate content. Concerning readability, our results show that policies issued by cooperative banks and to a lesser extent policies issued by savings banks are even harder to read after MiFID II relative to private universal banks and state-owned banks (German Landesbanken). In summary, the results of the regression

[25] Within this regression analysis, we measure similarity as the average cosine similarity of a bank's execution policy with the policies of the other 49 banks in the same year.

[26] Due to multicollinearity, only three *BankType*-dummies (i.e., cooperative bank, state-owned bank, and savings bank) are included in the regression, so that their coefficients need to be interpreted against private universal banks.

Table 5 Results of the regression analysis based on Eq. (1) regarding the impact of MiFID II on readability, complexity, and informativeness of best execution policies

	Readability measures				Complexity measures				Informativeness measures			
	Avg_word_len	Wps	Modified fog	Flesch	Doc_length	File_size	Entropy	Bigrams/doc_length	Similarity (tf-idf)	Similarity (doc2vec)	Boilerplate	Specificity
Intercept	6.886***	27.101***	16.91***	23.029***	173.692	0.929	7.267***	0.702***	0.388***	0.758***	0.395***	0.045**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.889)	(0.935)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.017)
MiFID	−0.023	1.391**	0.642***	−0.758	566.711	5.514*	0.151***	−0.01	0.099***	0.031***	−0.029	0.003
	(0.627)	(0.012)	(0.008)	(0.362)	(0.113)	(0.091)	(0.001)	(0.4)	(0.0)	(0.0)	(0.245)	(0.611)
Cooperative bank	0.321***	2.579**	1.813***	−5.861***	−669.158	1.056	−0.467***	−0.012	0.029	− 0.002	0.093**	0.047***
	(0.000)	(0.010)	(0.000)	(0.000)	(0.302)	(0.858)	(0.0)	(0.565)	(0.365)	(0.875)	(0.042)	(0.000)
State-owend bank	0.23***	−1.931**	−0.086	−1.754	−340.882	−1.007	0.071	−0.003	0.027	0.023**	0.138***	− 0.008
	(0.003)	(0.028)	(0.818)	(0.187)	(0.548)	(0.845)	(0.316)	(0.88)	(0.325)	(0.037)	(0.001)	(0.323)
Savings bank	0.516***	−3.472***	0.019	−4.750***	−713.267	−4.821	0.116*	0.03*	0.169***	0.064***	0.253***	− 0.012
	(0.000)	(0.000)	(0.955)	(0.000)	(0.154)	(0.289)	(0.065)	(0.07)	(0.0)	(0.0)	(0.0)	(0.107)
log_Employees	−0.027	0.188	0.112	−0.022	373.506	3.537	0.078*	−0.003	−0.009	−0.003	−0.031	−0.001
	(0.564)	(0.718)	(0.622)	(0.978)	(0.276)	(0.258)	(0.071)	(0.785)	(0.586)	(0.594)	(0.205)	(0.885)
log_TotalAssets	0.055	−0.530	−0.138	−0.041	−126.969	−1.196	−0.011	−0.004	0.022	0.01	0.018	0.00
	(0.212)	(0.284)	(0.523)	(0.957)	(0.694)	(0.685)	(0.777)	(0.71)	(0.172)	(0.119)	(0.419)	(0.972)
Observations	100	100	100	100	100	100	100	100	100	100	100	100
R2	0.425	0.376	0.251	0.259	0.159	0.144	0.547	0.188	0.523	0.491	0.501	0.326

MiFID II is a dummy variable being one for policies issued after MiFID II; *Cooperative Bank*, *State-owned Bank* (German "Landesbank"), and *Savings Bank* are dummy variables indicating the respective type of the bank; *log_Employees* and *log_TotalAssets* are the natural logarithm of each bank's number of employees and total assets in the year corresponding to the analyzed best execution policy

analyses show that best execution policies became harder to read, more complex, and more similar after MiFID II.

Case 2: cross-sectional analysis based on benchmark approach

In order to comprehensively demonstrate the applicability of the RIA-framework, we also perform a benchmark analysis to evaluate the impact of the best execution requirements in MiFID II as if it was a new regulation (see Section "Data Acquisition (Step 3)"). For this purpose, we make use of the 124 European best execution policies addressing retail investors and the 167 European best execution policies addressing professional investors (all in English). We differentiate between policies addressing retail and professional investors because the regulation explicitly requires investment firms to differentiate between these two types of clients (MiFID II Art. 27(9a)).

Informativeness—case 2

We investigate the informativeness of European best execution policies based on their textual similarity, and the share of boilerplate and specific information. Table 6 shows the descriptive statistics of the similarity, boilerplate, and specificity scores for all analyzed European best execution policies as well as for policies separated by type, i.e., retail and professional clients. Overall, we find that also the sample of European best execution policies shows relatively high similarity scores although being lower than in the German sample. Specifically, the cosine similarity based on doc2vec on average amounts to 0.71 indicating a considerable homogeneity between banks' best execution policies while the similarity based on tf-idf is slightly lower with an average of 0.42. These results also hold when differentiating between policies aimed at retail and professional clients although policies addressing retail clients are slightly more similar as suggested by the WRS-test. The best execution policies do have a large amount of boilerplate information with an overall average of 0.41, while their specificity is relatively low (overall average of 0.07). As the WRS-test indicates, there is no significant difference between the boilerplate information and the specificity between best execution policies addressed to retail and professional clients. Consequently, the relatively high similarity of the documents together with the high share of boilerplate information and the low specificity show that European best execution policies valid after MiFID II are not very informative.

In order to derive substantive conclusions on the similarity of best execution policies, however, we need to compare the level of similarity with a suitable reference. Therefore, we cluster the execution policies according to context-specific characteristics. These characteristics are derived from the regulation itself as it requires banks and brokers to differentiate between asset classes (MiFID II Art. 27(5)) in their best execution policies. If banks and brokers provide relevant information for clients in their policies, the similarity of policies within the same asset class cluster should increase because similar information should be included. Put differently, policies with the same context-specific focus and which address the same client group (e.g., retail clients with the intention to trade equities) should be more similar than policies of brokers with a different focus if the policies provide informational value as

requested by the new regulation. To investigate this, we perform a cluster analysis, whose results based on the tf-idf similarity are reported in Table 7.

To derive clusters according to covered asset classes, we first determine asset class weightings for each policy based on a TA approach, which are then used to cluster the policies using a K-Means algorithm (MacQueen et al. 1967).[27] This results in four asset class clusters for retail policies (labels: generalists, equity & debt focus, commodity focus, and foreign exchange (FX) focus) and three clusters for professional policies (labels: generalists, commodity focus, and FX focus). The results show that the similarity within five of the six clusters indeed slightly but significantly rises compared to the similarity of all retail (professional) policies, which is confirmed by a WRS-test. For the cluster "equity & debt focus" aimed at retail clients, the similarity of the policies and thus their informativeness slightly decreases while we observe no change for the cluster "generalists" aimed at professional clients. Yet, these two clusters represent the largest number of policies and corresponding firms. Similar results are obtained when calculating the similarity within clusters based on doc2vec (see Table 11 in the Appendix). Consequently, this analysis provides mixed results on the informational value of European best execution policies. In order to rule out that our results are driven by country-specific differences concerning the implementation of the European legislation into national law and its interpretation, we conduct a robustness test by dividing the sample according to brokers' geographical orientation (i.e., those being active in only one member state and those being active in multiple countries, who thus need to comply with potentially different national laws). The analysis shows that our results are not driven by such an effect since the similarity within these clusters only marginally changes compared to the full retail and professional samples (see Tables 12 and 13 in the Appendix).

Comprehensibility—case 2

We now analyze whether the regulation achieved its goal to increase the ease of understanding of best execution policies after the introduction of MiFID II. As derived in Step 4 of the framework, we apply different readability and textual complexity measures in order to assess the comprehensibility of policies and benchmark these to other documents which cover different ranges of readability and textual complexity.[28] Figure 3 shows the distributions of the different textual complexity measures for the retail and professional best execution policies as well as the benchmarks as described in Step 3.

[27] To determine a sound basis of the asset classes covered in a policy, a word frequency analysis of the policy corpus is conducted. Thereby, single words as well as bigrams are extracted and compared to asset classes specified by Saunders and Cornett (2012) to receive a complete list of asset classes including synonyms. Afterwards, each policy is matched with the constructed list of asset classes and the occurrence of each asset class per policy is stored. To standardize the occurrence of an asset class per policy, the frequency of an asset class is put in relation to the policy's length. Furthermore, this relative value is again divided by the mean of all policies' relative asset class occurrences. Finally, these standardized asset class occurrences are clustered using a K-Means algorithm (MacQueen et al. 1967) and the optimal number of clusters is determined by the elbow method (Tibshirani et al. 2001). The clustering is implemented with the *scikit-learn* library in Python.

[28] Due to outliers in 10-K filings, regulatory documents, and Wikipeda articles, we remove the upper and lower 5% of the observations for each readability and comprehensibility measure to derive reliable distributions for these benchmarks.

Table 6 Descriptive statistics of similarity (tf-idf, doc2vec), boilerplate and specificity scores for the 2020 retail and professional best execution policies

	Category	Count	Mean	SD	Min	25%	50%	75%	Max	WRS
Cosine similarity	Overall	34782.00	0.42	0.18	0.08	0.27	0.39	0.57	1.00	–
(based tf-idf)	Retail	15252.00	0.46	0.17	0.09	0.31	0.46	0.60	1.00	0.00
	Professional	27722.00	0.42	0.18	0.08	0.27	0.38	0.57	1.00	
Cosine similarity	Overall	34782.00	0.71	0.11	0.27	0.64	0.72	0.79	1.00	–
(based on doc2vec)	Retail	15252.00	0.72	0.11	0.30	0.64	0.73	0.80	1.00	0.00
	Professional	27722.00	0.71	0.11	0.27	0.63	0.72	0.79	1.00	
Boilerplate	Overall	187.00	0.41	0.12	0.06	0.34	0.40	0.50	0.75	–
	Retail	124.00	0.41	0.13	0.06	0.32	0.40	0.48	0.67	0.49
	Professional	167.00	0.42	0.12	0.13	0.34	0.41	0.50	0.75	
Specificity	Overall	187.00	0.07	0.03	0.02	0.05	0.06	0.08	0.22	–
	Retail	124.00	0.06	0.03	0.02	0.05	0.06	0.07	0.22	0.21
	Professional	167.00	0.07	0.03	0.02	0.05	0.06	0.08	0.22	

Count represents the number of observations in the similarity analysis, which is determined by $N \times (N-1)$, where N is the number of policies

Table 7 Asset class cluster-wise descriptive statistics of similarity scores (cosine similarity based on tf-idf) for the 2020 retail and professional best execution policies and WRS-test comparing each cluster with all policies of the respective category

Category	Asset class cluster	Count	Mean	SD	Min	25%	50%	75%	Max	WRS
Retail	Generalists	552.00	0.64	0.14	0.21	0.60	0.68	0.74	0.97	0.00
	Equity and debt focus	6642.00	0.43	0.16	0.11	0.31	0.41	0.55	1.00	0.00
	Commodity focus	6.00	0.65	0.08	0.58	0.59	0.62	0.72	0.75	0.00
	FX focus	210.00	0.62	0.15	0.27	0.51	0.65	0.72	0.87	0.00
Professional	Generalists	12656.00	0.42	0.18	0.09	0.28	0.38	0.57	1.00	0.26
	Commodity focus	870.00	0.51	0.16	0.15	0.38	0.50	0.64	1.00	0.00
	FX focus	552.00	0.57	0.14	0.23	0.46	0.57	0.68	0.97	0.00

Count represents the number of observations in the similarity analysis, which is determined by $N \times (N-1)$, where N is the number of policies

Although investment firms are obliged to differentiate between retail and professional clients in their policies, the distributions across the different readability and complexity measures almost completely overlap. Policies aimed at retail clients are as hard to read as policies aimed at professional clients. Consequently, we do not find any significant difference in ease of understanding of these policies, which—as a first finding—casts doubt on whether the regulation's goal of comprehensible best execution policies for both retail and professional clients is achieved. This result is

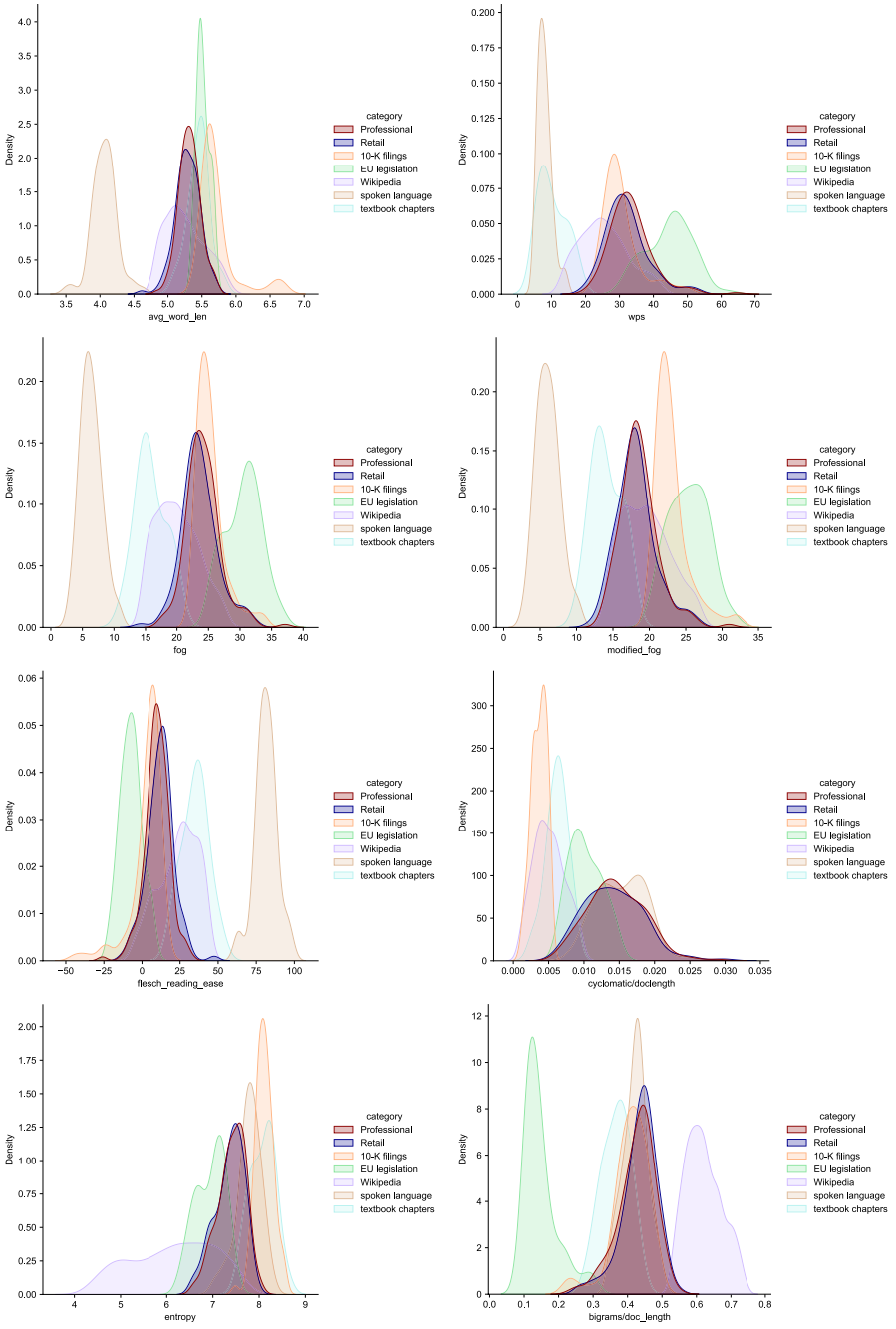

Fig. 3 Comparison of the distribution of the readability and textual complexity measures with benchmarks

also confirmed by the descriptive statistics and the corresponding WRS-tests provided in Table 10 in the Appendix.[29]

Yet, comparing the readability of retail and professional best execution policies does not allow to draw conclusions how difficult these policies are actually to read. For this purpose, the benchmark analysis is conducted. Considering the distributions of the benchmarks, we observe that the value range of readability and complexity of best execution policies is limited and can be clearly differentiated from the benchmarks. According to the readability measures (average word length, wps, (modified) Fog index,[30] and Flesch reading ease score), best execution polices are as difficult to read as companies' annual reports in 10-K filings and almost as difficult to read as European regulatory documents themselves. Spoken language, Wikipedia articles, and textbook chapters are noticeably easier to read than best execution policies. Similar observations can be made concerning textual complexity. Based on the relative number of cyclomatic statements, the execution policies again belong to the most complex documents which are in this case spoken language and regulatory documents. Most of the other benchmarks reveal noticeable smaller amounts of cyclomatic statements. This can be explained by the character of these documents. While regulatory documents and spoken language use many conditional terms (Li et al. 2015; Auer 2009), e.g., to explain under what circumstances a certain regulation is applicable or under what conditions a statement holds, the nature of textbook chapters, Wikipedia articles, and 10-K filings is descriptive and explanatory and, thus, conditional statements are avoided to not confuse the reader. Best execution policies also include a high share of cyclomatic statements, which are for example used to state under which condition an order is routed to a specific trading venue. Also regarding complexity measured by the number of unique bigrams, best execution policies belong to the most complex documents and are comparable to the MD&A section of 10-K filings. Yet, best execution policies use a slightly smaller, i.e., more limited, vocabulary than most of the benchmarks as shown by the distribution of entropy. Best execution policies show an overall lower entropy with the distribution mainly ranging from 6.5 to 8 centered at approximately 7.5, while the entropy of spoken language, textbook chapters, and 10-K filings is higher with the center of the distribution being close to or larger than 8. Only regulatory texts and Wikipedia use an even smaller variation of different terms. Yet, the meaningfulness of entropy as a measure for comprehensibility is less convincing than the other readability and complexity measures when comparing texts from different contexts. Due to the topical focus and the regular repetition of certain terms and definitions, words in legal texts and also best execution policies are more predictable than in other documents.

In summary, the analysis shows that best execution policies are not easy to understand as intended by the regulatory requirements in MiFID II but are highly complex

[29] Our results also hold when separating best execution policies into three groups, i.e., policies that apply to retail investors only, professional investors only, and both retail and professional investors (see Table 14, and Figs. 7 and 8 in the Appendix) as the distribution of readability and complexity is highly similar in all three groups.

[30] For the benchmark approach, the standard Fog index is more meaningful since the benchmarks are not related to best execution and the exclusion of best execution related common complex words would thus bias the results. Nevertheless, we still report the distributions based on the modified Fog index in Fig. 3.

documents which are difficult to read. According to the analysis, comprehensibility of best execution policies is similar to regulatory documents and companies' annual reports, which is way above what can be expected from retail clients.

4.6 Communication to stakeholders (Step 6)

The results of the analysis show that the regulatory change in MiFID II regarding best execution policies did not achieve the desired results. The *informativeness* of the policies did not increase as intended but rather decreased as shown by the pre-post analysis of German best execution policies. Also for the cross-sectional analysis of best execution policies, we find that these documents are relatively homogeneous and contain high levels of boilerplate information and few specifics. Concerning the second dimension, which is the *comprehensibility* of best execution policies, the analysis shows that MiFID II did not lead to easier understandable policies but to policies that are actually more complex and harder to read (longitudinal pre-post analysis). Also, it shows that policies are still too difficult to read compared with the applied benchmarks. Consequently, the analysis of the new regulatory requirements for best execution in MiFID II suggests that the goals of increased investor protection and competition between brokers due to more informative and easily understandable best execution policies were not achieved. This finding is important for regulators, investment firms, and investors alike and calls for further regulatory action.

5 Evaluation of the framework

We now evaluate our proposed RIA-framework in case of unstructured data based on its application to the changed best execution requirements in MiFID II. Rigorous evaluation of the developed artifact is an essential step in design science research (Hevner et al. 2004). With the application of the RIA-framework to a real-world setting, we follow Peffers et al. (2007) to demonstrate the applicability and usefulness of the developed artifact. Case studies are frequently used to evaluate conceptual, actionable instructions such as our framework in design science research (Peffers et al. 2012).

The objective of the framework is to evaluate the quality and effectiveness of regulatory actions that result in unstructured data by providing the necessary process steps, decisions, and data requirements, as well as the suitable methodologies. The developed framework enables regulatory authorities and researchers to assess the impact of regulatory actions aimed at or resulting in unstructured data in a clear and structured manner, and thus provides a solution for this class of real-world problems. The application of the framework to the best execution requirements in MiFID II shows that it fulfills these objectives. The impact of the change in regulation could clearly and objectively be identified. As the demonstration in the previous section shows, each step of the proposed framework was successfully applied to the regulatory change in best execution requirements in MiFID II. In particular, the two most

important steps for RIA in case of unstructured data, i.e., the selection of appropriate data and methodology, provided clear guidance for the assessment of the regulatory impact. Specifically, with the help of the benchmark approach, the framework does not only enable to assess changes in regulation by comparing documents before and after the regulatory change, but also enables to analyze new regulations resulting in new documents based on appropriate benchmarks. The RIA-framework also provides the necessary guidance to map TA and NLP methods to the dimensions affected by a regulation. After completion of all six steps of the proposed framework, we were able to show how the regulatory action impacted the dimensions targeted by the regulatory change, whether the regulation achieved the desired goals, and whether further regulatory action might be necessary.

6 Discussion

More and more regulations aim at or result in huge numbers of textual documents. In order to assess whether regulatory actions have met the desired goals, RIA needs to be conducted. Yet, and in contrast to regulatory actions that can be measured with structured data, no framework or guideline exist as to conduct RIA in case of unstructured data. To solve this problem and to contribute to research on RegTech supporting regulators in improving regulatory intelligence, we develop a framework for RIA in case of unstructured data that is based on existing RIA guidelines for structured data and methods from TA and NLP. The framework is mainly intended as a methodology for researchers and regulatory authorities, but can also be applied by different stakeholders affected by a regulatory action. With this study, we pave the way for a largely untapped field of research within the RegTech literature, i.e., RegTech and decision support for regulators and policy makers in addition to the current main fields: compliance by firms and supervision by competent authorities.

Our research approach is based on the design science paradigm (Hevner et al. 2004) and follows the established design science research methodology by Peffers et al. (2007). Most relevant contributions in design science research represent either an improvement of an existing process or the extension of existing methods to a yet unsolved problem in another field (Gregor and Hevner 2013). Our framework belongs to the second group because it applies established TA and NLP methods in the context of RIA so that also the impact of regulations that result in or aim at unstructured data can be assessed. This process is nontrivial as one crucial step is the mapping of appropriate methods to the dimensions affected by the regulation in light of the regulation's overall goals. Our framework provides guidance in this respect and is extensible based on the investigation of further use cases.

The evaluation of the artifact via the case study of best execution requirements in MiFID II shows that the RIA-framework is valid, useful, and provides a solution to a previously unsolved problem. The framework gives guidance to regulatory authorities and researchers to assess both the effects of regulatory actions resulting in unstructured data and the compliance of firms that have to provide such textual data. Thereby, the developed RIA-framework can support evidence-based policy making

and improve the quality of regulatory actions aimed at or resulting in unstructured data.

Existing approaches to evaluate such regulatory actions are based on manual inspection of the relevant documents and qualitative assessments of interviews and peer review procedures, which is highly burdensome, resource-intensive, and often does not lead to objective and clear results (European Securities and Markets Authority 2015, 2017). In contrast, our proposed approach is largely automated, follows established research methods, and leads to objective results. Based on TA, NLP and associated research methods, the affected documents are parsed, preprocessed, and automatically analyzed so that no or few manual inspection of the documents is necessary.

Our framework is not unique to the finance discipline or the financial industry—neither the problem at hand (improving informativeness and comprehensibility of texts for customers) nor any step within our proposed framework. Researchers, regulators, policy makers, and other stakeholders can, therefore, use our method to assess the impact of regulations aimed at unstructured data in different contexts and different domains. Examples for application areas in other domains are quality requirements for patient information leaflets in the pharmaceutical sector[31], companies' corporate social responsibility (CSR) reportings[32], or cybersecurity-related disclosure requirements[33]. We believe that the framework is universally applicable and that it represents a general principle to solve a class of real-world problems (i.e., RIA in case of unstructured data), rather than describing a unique set of steps and methods to solve a unique problem (i.e., impact of best execution requirements in MiFID II) consistent with design theory (Gregor and Hevner 2013). Future research can apply the framework to other domains to confirm this assumption.

There are some limitations to our study: Although the proposed framework uses research methods from TA and NLP and, thus, facilitates the resource-efficient analysis of documents, not every step of the framework can be automated. In particular, the identification of the impacted dimensions and the mapping of appropriate analysis methods is of high importance and requires human intervention and substantial background knowledge in the respective field. Also, the dimensions impacted by a regulation and corresponding suitable research methods are case specific to a certain degree. Yet, our RIA-framework provides guidance on generally applicable mappings of frequently occurring regulatory dimensions and corresponding research

[31] For example, the EU healthcare initiative (https://health.ec.europa.eu/other-pages/basic-page/information-patients-legislative-approach_en) and the best practice guidance on patient information leaflets in the UK (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/946602/Best_practice_guidance_on_patient_information_leaflets.pdf) demand patient information leaflets to contain relevant information in a comprehensible language that can be easily understood by lay people.

[32] For example, the proposal of the European Commission for corporate sustainability reporting https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52021PC0189&from=EN requires the reported information in CSR reports to be "understandable, relevant, representative, verifiable, comparable, and [...] represented in a faithful manner" (Article 19b (2)).

[33] For example, the SEC considers a proposal to mandate cybersecurity disclosures by public companies (www.sec.gov/news/statement/gensler-cybersecurity-20220309), which would require ongoing disclosures on companies' governance, risk management, and strategy with respect to cybersecurity risks.

methods, which can be extended in future research. Moreover, the framework supports researchers and regulators to select appropriate TA and NLP methods for their specific dimensions and cases. Finally, the application of the framework can be challenging in case of data limitations as it is the case with the analysis of German best execution policies where, due to data availability, the time span between the pre- and post-MiFID II documents is relatively long. Yet, such data problems regularly do not exist for the most important potential users of the framework, i.e., regulators and supervisory authorities, who can request the relevant documents for the analysis from investment firms or other impacted entities.

The framework can be extended in several ways: Because cost-benefit analyses prior to enacting a regulation are common to estimate the potential benefits of a regulation (Boardman et al. 2017), the framework can be extended to also cover ex-ante RIA. For such an ex-ante analysis, the framework could follow the benchmark approach proposed for new regulations and assess the impact of the intended regulatory action based on already existing documents that approximate those of the proposed regulation. Moreover, the framework can be extended by considering not only textual data but also other types of unstructured data such as images, videos clips, or speeches. Especially for regulations regarding social media platforms and the content which is provided there[34], images and video clips represent important sources of information.

7 Conclusion

RIA is highly important for evidence-based policy making and to achieve "better regulation"[35] given the complexity and interdependencies in today's regulatory environment. At the same time, more unstructured data is generated and more regulations aim at such documents or result in unstructured data themselves. While RIA and its associated frameworks and guidelines (e.g., OECD 1997) are common for regulatory actions that are measurable by quantitative data (Radaelli 2004), RIA for regulatory actions aimed at unstructured data is scarce and no suitable framework exists. New IT-enabled processes and frameworks are necessary to solve this problem and to improve regulatory intelligence.

Our study provides a solution to this previously unsolved real-world problem by developing a framework for RIA in case of unstructured data. Thereby, we help to close this research gap and contribute to the sparse strand of literature on RegTech that provides innovative IT-solutions for regulators and policy makers. Following the design science research paradigm, the framework is developed based on existing research related to RIA and methods from the fields of TA and NLP. The framework

[34] The regulation of social media platforms and the content provided there is an area which is more and more debated among policy makers worldwide against the backdrop of fake news and hate speech. Also, several jurisdictions have already enacted regulations for the content on social media platforms (see, e.g., The Law Library of Congress, Global Legal Research Directorate (2019) for an overview).

[35] See, e.g., the "better regulation" strategy of the European Commission (Radaelli 2018; European Commission 2019).

provides clear guidance together with the necessary process steps, decisions, and data requirements, as well as suitable methodologies to assess the impact of a regulation aimed at or resulting in unstructured data in an organized, largely automated, and objective manner. There are three new contributions of this paper to RegTech literature: First, the RIA-framework itself is a contribution to regulatory intelligence. Second, the framework explains how to map appropriate methods to the dimensions affected by the regulation in light of the desired regulatory goals. Third, the framework includes a benchmark approach to assess the impact of regulations resulting in new documents whereby the benchmarks represent reference documents to evaluate the regulatory impact.

In line with design science principles, we demonstrate and evaluate the applicability of the framework and its usefulness based on a use case. Specifically, we apply the framework to assess the impact of the regulations for banks and brokers' best execution policies in MiFID II. Thereby, we conduct a longitudinal analysis comparing best execution policies pre- and post MiFID II of the largest German investment firms as well as a cross-sectional benchmark analysis of the main European investment firms. By using our framework in this case study, we find that the requirements in MiFID II regarding informativeness and comprehensibility of best execution policies did not reach the desired goals and, thus, investor protection was not strengthened, showing that further regulatory action is necessary. Although we have proposed the RIA-framework for financial regulations and demonstrated its usefulness with an example from this domain, the framework is not unique to the financial industry. It can be applied in other domains and to serve researchers and regulators as a toolbox to assess the impact of a regulatory action aimed at or resulting in unstructured data. Furthermore, our study provides one of the first steps into a research field "RegTech for regulators", supporting regulatory intelligence and evidence-based policy making, which may open new paths for future research.

Appendix

Appendix A. Additional tables and figures pre-post analysis

See Tables 8, 9 and Figs. 4, 5, 6.

Table 8 Descriptive statistics of readability and complexity measures for the 2009 and 2020 German best execution policies

	Year	Count	Mean	SD	Min	25%	50%	75%	Max	WRS
Readability measures										
Avg_word_len	2009	50.00	7.50	0.30	6.88	7.24	7.61	7.74	8.01	0.75
	2020	50.00	7.51	0.29	6.76	7.27	7.65	7.74	7.76	
Wps	2009	50.00	21.78	3.01	17.17	19.54	20.34	24.00	28.70	0.03
	2020	50.00	22.89	3.32	19.22	20.63	20.63	24.16	30.37	
Modified_fog	2009	50.00	16.57	1.03	14.50	16.14	16.58	17.32	19.25	0.12
	2020	50.00	17.17	1.43	13.63	16.75	16.75	16.98	20.99	
Flesch	2009	50.00	19.68	4.73	9.98	16.51	18.88	23.38	30.48	0.78
	2020	50.00	18.88	4.27	11.40	17.46	17.46	21.54	34.21	
Complexity measures										
Doc_length	2009	50.00	1420.30	742.56	292.00	1198.25	1423.50	1504.00	5037.00	0.04
	2020	50.00	2057.06	2417.09	594.00	1240.00	1240.00	2040.75	17547.00	
File_size	2009	50.00	14.66	6.40	4.38	12.91	15.11	15.69	43.73	0.04
	2020	50.00	20.85	21.86	4.80	12.74	12.74	21.52	158.50	
Entropy	2009	50.00	7.78	0.35	6.95	7.74	7.88	7.94	8.66	0.00
	2020	50.00	7.95	0.24	7.35	7.97	7.97	8.06	8.62	
Bigrams/doc_length	2009	50.00	0.65	0.04	0.53	0.63	0.66	0.67	0.77	0.73
	2020	50.00	0.63	0.07	0.39	0.59	0.68	0.68	0.72	

Table 9 Descriptive statistics of the readability, complexity, and informativeness measures per bank type. Panel A reports the mean values for 2009 separately for policies issued by private universal banks, cooperative banks, state-owned banks (German Landesbanken), and savings banks. Panel B shows the mean values for 2020

	Readability measures					Complexity measures				Informativeness measures			
	Count	Avg_wod_len	Wps	Modified_fog	Flesch	Doc_length	File_size	Entropy	Bigrams/doc_length	Similarity (tf-idf)	Similarity(doc2vec)	Boilerplate	Specificity
Panel A: 2009													
Private universal banks	17.00	7.28	22.96	16.53	22.12	1704.94	15.82	7.80	0.63	0.55	0.83	0.34	0.04
Cooperative banks	5.00	7.52	22.93	16.12	20.66	577.80	9.74	7.08	0.68	0.51	0.81	0.48	0.08
State-owned bank	8.00	7.50	21.33	16.35	20.34	1401.75	14.41	7.83	0.64	0.61	0.87	0.51	0.03
Savings bank	20.00	7.69	20.66	16.81	17.09	1396.40	15.02	7.91	0.66	0.69	0.89	0.57	0.02
Panel B: 2020													
Private universal banks	17.00	7.26	23.65	16.76	21.82	3074.76	28.31	8.05	0.61	0.67	0.88	0.29	0.03
Cooperative banks	5.00	7.56	30.08	20.88	11.80	1931.60	27.65	7.58	0.57	0.72	0.87	0.38	0.09
State-owned bank	8.00	7.52	21.43	16.64	20.17	2055.38	21.63	8.01	0.61	0.67	0.89	0.45	0.03
Savings bank	20.00	7.70	21.03	16.80	17.64	1224.05	12.49	7.94	0.68	0.79	0.91	0.59	0.03

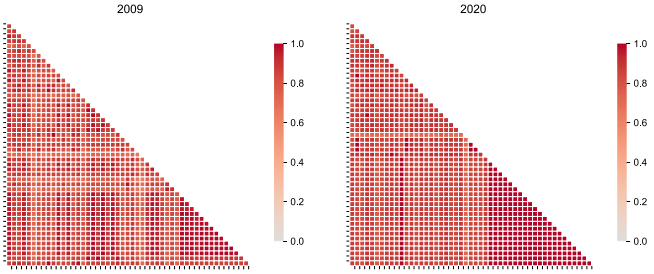

Fig. 4 Similarities (based on doc2vec) of German best execution policies of 2009 and 2020

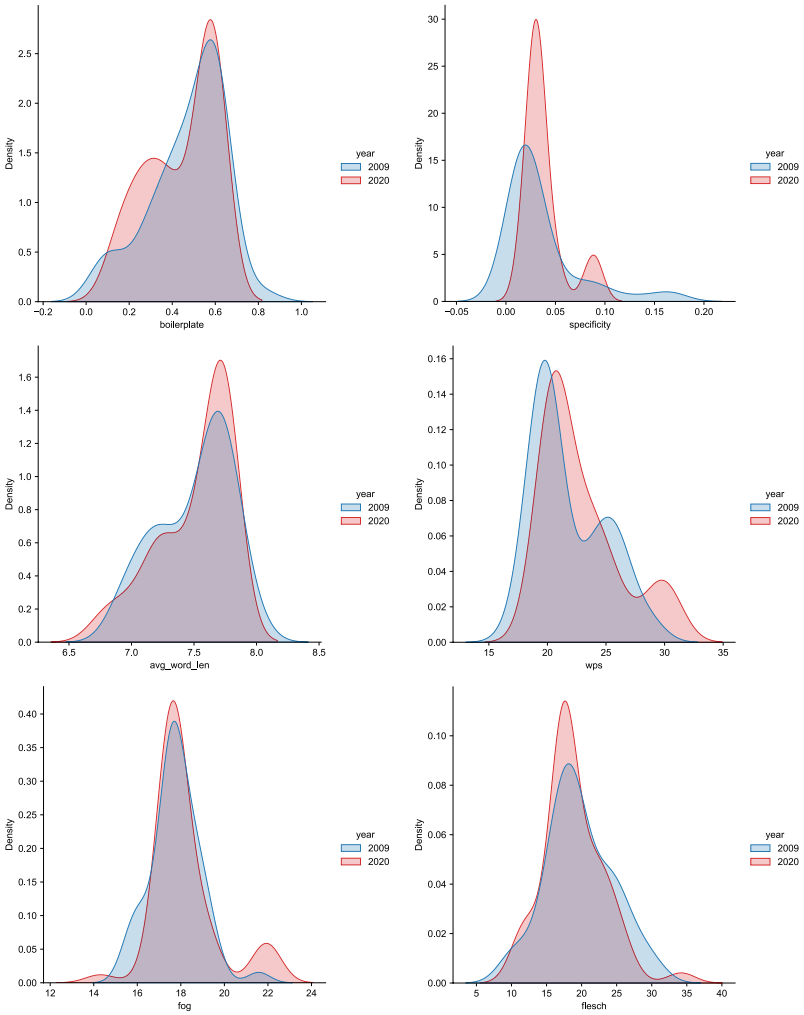

Fig. 5 Comparison of the distribution of the informativeness and readability measures for the 2009 and 2020 German best execution policies

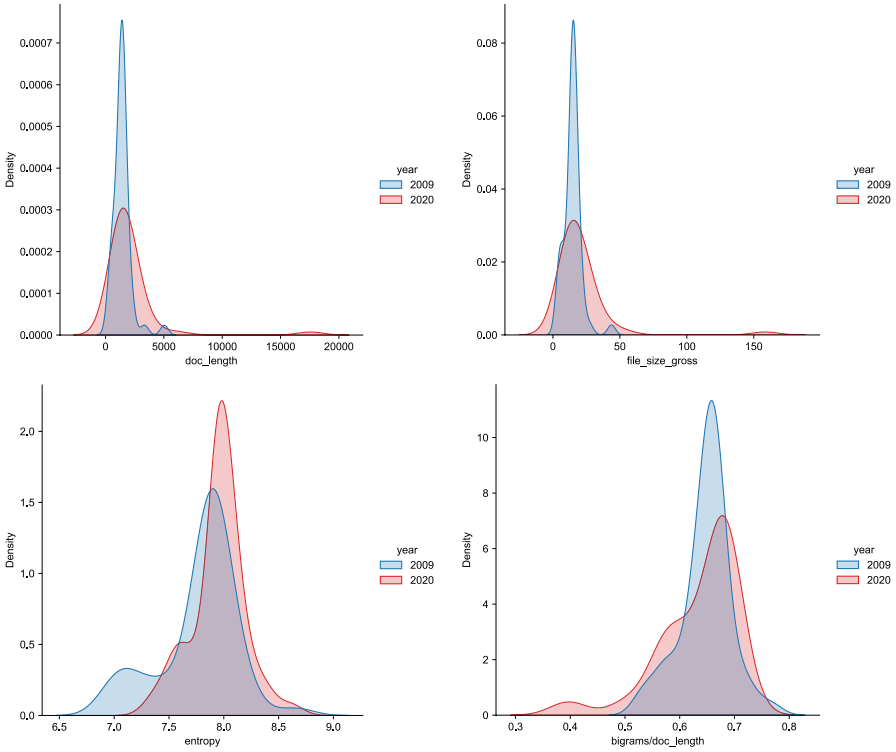

Fig. 6 Comparison of the distribution of the complexity measures for the 2009 and 2020 German best execution policies

Appendix B. Additional tables and figures benchmark analysis

See Tables 10, 11, 12, 13 and 14 and Figs. 7 and 8.

Table 10 Descriptive statistics of readability and complexity measures for the 2020 retail and professional best execution policies

	Category	Count	Mean	SD	Min	25%	50%	75%	Max	WRS
Readability measures										
Avg_word_len	Retail	124.00	5.29	0.18	4.61	5.18	5.28	5.41	5.72	0.32
	Professional	167.00	5.32	0.16	4.83	5.22	5.31	5.41	5.72	
Wps	Retail	124.00	32.07	6.30	19.88	27.92	31.55	34.69	52.81	0.05
	Professional	167.00	33.26	6.32	20.03	29.43	32.49	35.87	64.43	
Modified_fog	retail	124.00	18.16	2.68	12.04	16.44	17.97	19.43	26.62	0.13
	Professional	167.00	18.59	2.63	13.34	17.04	18.27	19.79	30.91	
Flesch	Retail	124.00	12.15	8.88	− 12.95	7.19	13.06	17.19	47.35	0.02
	Professional	167.00	9.89	8.05	− 25.97	5.85	9.62	14.65	32.42	
Complexity measures										
Doc_length	Retail	124.00	3634.92	2660.64	606.00	1815.50	2866.00	4567.50	17802.00	0.08
	Professional	167.00	4086.40	2777.21	606.00	2048.50	3430.00	5199.50	17802.00	
File_size	Retail	124.00	27.36	20.04	4.12	12.35	22.68	35.25	134.93	0.06
	Professional	167.00	31.57	21.61	4.12	16.04	25.42	39.94	134.93	
Cyclomatic/doclength	retail	124.00	0.01	0.00	0.01	0.01	0.01	0.02	0.03	0.36
	Professional	167.00	0.01	0.00	0.01	0.01	0.01	0.02	0.03	
Entropy	Retail	124.00	7.36	0.31	6.58	7.18	7.41	7.57	7.95	0.22
	Professional	167.00	7.41	0.30	6.60	7.24	7.43	7.64	8.04	
Bigrams/doc_length	Retail	167.00	0.43	0.05	0.24	0.40	0.44	0.46	0.55	0.17
	Professional	167.00	0.43	0.05	0.24	0.40	0.44	0.46	0.55	

Table 11 Asset class cluster-wise descriptive statistics of similarity scores (cosine similarity based on doc2vec) for the 2020 retail and professional best execution policies and WRS-test comparing each cluster with all policies of the respective category. Count represents the number of observations in the similarity analysis, which is determined by $N \times (N-1)$, where N is the number of policies

Category	Asset class cluster	Count	Mean	SD	Min	25%	50%	75%	Max	WRS
Retail	Generalists	552.00	0.74	0.12	0.39	0.68	0.76	0.84	0.99	0.00
	Equity & debt focus	6642.00	0.71	0.11	0.30	0.63	0.72	0.79	1.00	0.00
	Commodity focus	6.00	0.70	0.05	0.66	0.66	0.67	0.74	0.77	0.51
	FX focus	210.00	0.77	0.08	0.55	0.72	0.78	0.84	0.94	0.00
Professional	Generalists	12656.00	0.71	0.11	0.27	0.64	0.72	0.80	1.00	0.00
	Commodity focus	870.00	0.68	0.10	0.33	0.62	0.68	0.75	1.00	0.00
	FX focus	552.00	0.75	0.09	0.49	0.71	0.77	0.82	0.99	0.00

Table 12 Descriptive statistics of similarity scores (cosine similarity based on tf-idf) for the 2020 retail and professional best execution policies clustered according to a regional or multinational focus of the respective bank or broker and WRS-test comparing the similarity distribution in each cluster with all policies of the respective category. Count represents the number of observations in the similarity analysis, which is determined by $N \times (N-1)$, where N is the number of policies

Category	Cluster	Count	Mean	SD	Min	25%	50%	75%	Max	WRS
Retail	All policies	15252.00	0.46	0.17	0.09	0.31	0.46	0.60	1.00	
Professional	All policies	27722.00	0.42	0.18	0.08	0.27	0.38	0.57	1.00	
Retail	Regional	4556.00	0.48	0.15	0.14	0.36	0.49	0.59	1.00	0.00
	Multinational	3080.00	0.48	0.18	0.12	0.33	0.45	0.64	0.96	0.00
Professional	Regional	4692.00	0.45	0.15	0.14	0.33	0.44	0.57	1.00	0.00
	Multinational	9506.00	0.44	0.19	0.09	0.28	0.40	0.61	1.00	0.00

Table 13 Descriptive statistics of similarity scores (cosine similarity based on doc2vec) for the 2020 retail and professional best execution policies clustered according to a regional or multinational focus of the respective bank or broker and WRS-test comparing the similarity distribution in each cluster with all policies of the respective category. Count represents the number of observations in the similarity analysis, which is determined by $N \times (N-1)$, where N is the number of policies

Category	Cluster	Count	Mean	SD	Min	25%	50%	75%	Max	WRS
Retail	All policies	15252.00	0.72	0.11	0.30	0.64	0.73	0.80	1.00	
Professional	All policies	27722.00	0.71	0.11	0.27	0.63	0.72	0.79	1.00	
Retail	Regional	4556.00	0.71	0.12	0.34	0.63	0.72	0.80	1.00	0.01
	Multinational	3080.00	0.73	0.11	0.35	0.66	0.74	0.81	0.99	0.00
Professional	Regional	4692.00	0.71	0.11	0.34	0.64	0.72	0.79	1.00	0.10
	Multinational	9506.00	0.71	0.11	0.33	0.63	0.72	0.79	1.00	0.49

Table 14 Descriptive statistics of readability and complexity measures for the 2020 best execution policies separated by policies that apply to retail investors only, professional investors only, and to both retail & professional investors

	Category	Count	Mean	SD	Min	25%	50%	75%	Max
Readability measures									
Avg_word_len	Retail	20.00	5.17	0.20	4.61	5.07	5.21	5.27	5.44
	Professional	63.00	5.32	0.14	5.00	5.24	5.32	5.39	5.72
	Retail & professional	104.00	5.32	0.17	4.83	5.20	5.31	5.42	5.72
Wps	Retail	20.00	29.87	5.62	19.88	26.67	28.41	32.28	44.77
	Professional	63.00	34.54	6.11	25.04	30.49	33.95	36.46	64.43
	Retail & professional	104.00	32.49	6.35	20.03	28.13	31.74	35.00	52.81
Modified_fog	retail	20.00	16.81	2.42	12.04	15.56	16.90	17.76	21.84
	Professional	63.00	18.88	2.58	14.41	17.25	18.52	20.16	30.91
	Retail & professional	104.00	18.42	2.66	13.34	16.87	18.14	19.57	26.62
Flesch	Retail	20.00	17.47	10.12	− 3.92	13.96	16.14	21.60	47.35
	Professional	63.00	7.86	7.26	− 25.97	5.03	8.01	12.29	21.90
	Retail & professional	104.00	11.12	8.29	− 12.95	6.70	11.25	16.15	32.42
Complexity measures									
Doc_length	Retail	20.00	2594.65	2411.25	728.00	1446.50	1846.50	2907.00	11927.00
	Professional	63.00	4501.44	2919.55	897.00	2502.00	3765.00	5700.50	13794.00
	Retail & professional	104.00	3834.97	2670.31	606.00	1934.00	3245.00	4901.50	17802.00
File_size	Retail	20.00	18.86	17.93	4.40	9.82	12.58	23.81	86.38
	Professional	63.00	35.80	23.46	6.58	18.77	31.94	42.21	102.93
	Retail & professional	104.00	29.00	20.09	4.12	13.98	23.87	37.63	134.93
Cyclomatic/doclength	Retail	20.00	0.02	0.00	0.01	0.01	0.01	0.02	0.03
	Professional	63.00	0.02	0.00	0.01	0.01	0.01	0.02	0.03
	Retail & professional	104.00	0.01	0.00	0.01	0.01	0.01	0.02	0.02
Entropy	Retail	20.00	7.31	0.34	6.58	7.18	7.42	7.51	7.82
	Professional	63.00	7.48	0.28	6.62	7.30	7.50	7.67	8.04

Table 14 (continued)

	Category	Count	Mean	SD	Min	25%	50%	75%	Max
	Retail & professional	104.00	7.37	0.31	6.60	7.18	7.41	7.58	7.95
Bigrams/doc_length	Retail	20.00	0.45	0.04	0.33	0.43	0.45	0.47	0.51
	Professional	63.00	0.42	0.06	0.24	0.38	0.43	0.46	0.52
	Retail & professional	104.00	0.43	0.05	0.26	0.41	0.44	0.46	0.55
Informativeness measures									
Boilerplate	Retail	20.00	0.40	0.16	0.06	0.31	0.38	0.50	0.67
	Professional	63.00	0.43	0.12	0.15	0.35	0.41	0.52	0.75
	Retail & professional	104.00	0.41	0.12	0.13	0.33	0.40	0.47	0.66
Specificity	Retail	20.00	0.06	0.02	0.02	0.04	0.05	0.07	0.12
	Professional	63.00	0.07	0.02	0.02	0.06	0.07	0.08	0.13
	Retail & professional	104.00	0.07	0.03	0.03	0.05	0.06	0.07	0.22

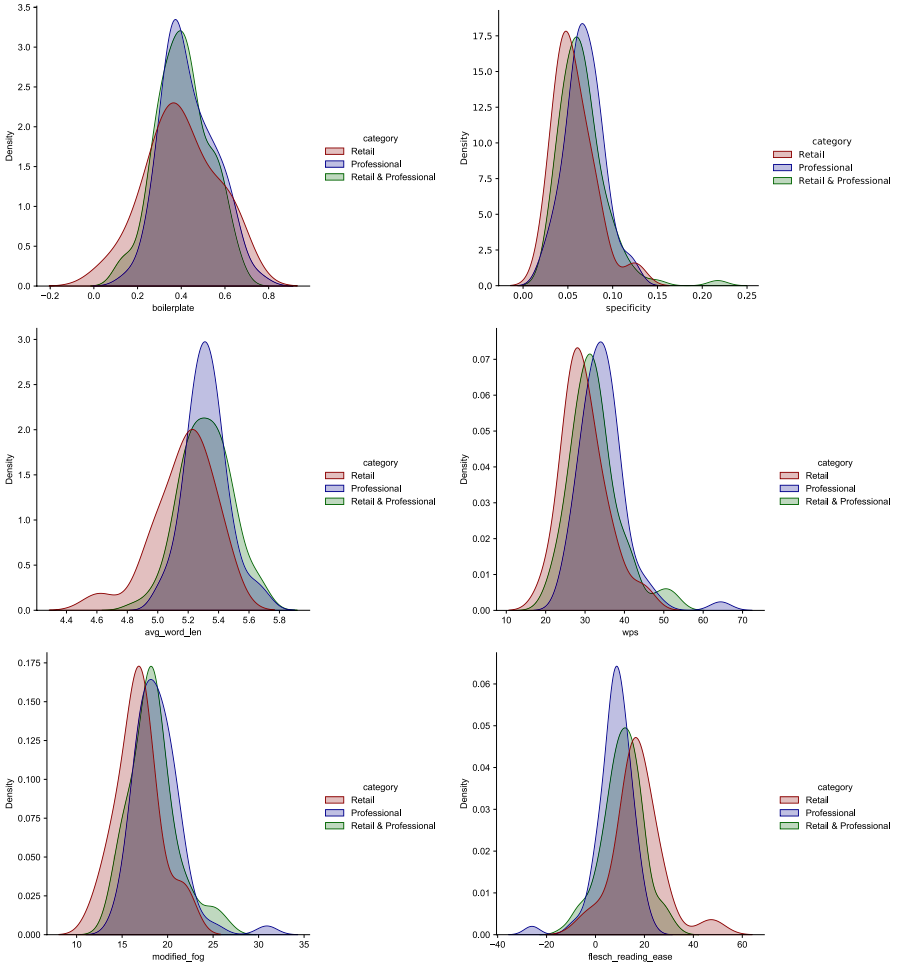

Fig. 7 The distribution of informativeness and readability measures is highly similar for all three groups of policies, i.e., policies applying to retail customers only, professional clients only, and to retail & professional clients

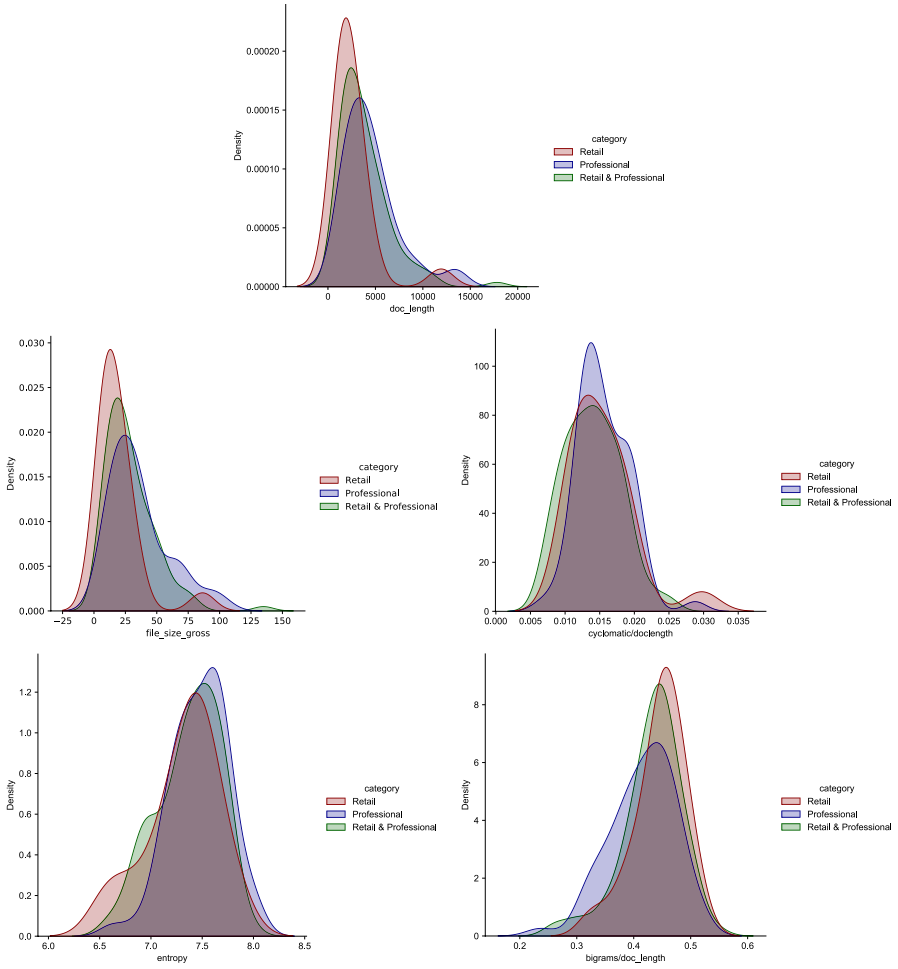

Fig. 8 The distribution of complexity measures is highly similar for all three groups of policies, i.e., policies applying to retail customers only, professional clients only, and to retail & professional clients

Funding Open Access funding enabled and organized by Projekt DEAL. We gratefully acknowledge financial support for this research project from efl—the Data Science Institute.

Data availibility The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

Aggarwal CC, Zhai C (2012) Mining text data. Springer Science & Business Media, New York, NY
Al-Ubaydli O, McLaughlin PA (2017) RegData: a numerical database on industry-specific regulations for all united states industries and federal regulations, 1997–2012. Regul Govern 11(1):109–123
Angelov D (2020) Top2Vec: distributed representations of topics. arXiv preprint arXiv:2008.09470
Arner DW, Barberis JN, Buckley RP (2016) The emergence of RegTech 2.0: from know your customer to know your data. J Finance Transform 44:79–86
Arner DW, Barberis J, Buckey RP (2017) FinTech, RegTech, and the reconceptualization of financial regulation. Northwestern J Int Law Bus 37:371–413
Auer P (2009) On-line syntax: thoughts on the temporality of spoken language. Lang Sci 31(1):1–13
Bag S, Kumar SK, Tiwari MK (2019) An efficient recommendation generation using relevant Jaccard similarity. Inf Sci 483:53–64
Bannier C, Pauls T, Walter A (2019) Content analysis of business communication: introducing a German dictionary. J Bus Econ 89(1):79–123
Barrón-Cedeño A, Basile C, Degli Esposti M, Rosso P (2010) Word length n-grams for text re-use detection. International conference on intelligent text processing and computational linguistics. Springer, Berlin, Heidelberg, pp 687–699
Bhattacharyya S, Jha S, Tharakunnel K, Westland JC (2011) Data mining for credit card fraud: a comparative study. Decis Support Syst 50(3):602–613
Bird S, Klein E, Loper E (2009) Natural language processing with Python: analyzing text with the natural language toolkit. O'Reilly Media Inc, Sebastopol, CA
Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation. J Mach Learn Res 3:993–1022
Boardman AE, Greenberg DH, Vining AR, Weimer DL (2017) Cost–benefit analysis: concepts and practice. Cambridge University Press, New York, NY
Bommarito MJ, Katz DM (2010) A mathematical approach to the study of the United States code. Physica A 389(19):4195–4200
Bonsall SB IV, Leone AJ, Miller BP, Rennekamp K (2017) A plain English measure of financial reporting readability. J Account Econ 63(2–3):329–357
Burke M, Fry J (2019) How easy is it to understand consumer finance? Econ Lett 177:1–4
Butler T, O'Brien L (2019) Understanding RegTech for digital regulatory compliance. Disrupting finance. Palgrave Pivot, Cham, pp 85–102
Carley K (1993) Coding choices for textual analysis: a comparison of content analysis and map analysis. Sociol Methodol 23:75–126
Cohen L, Malloy C, Nguyen Q (2020) Lazy prices. J Finance 75(3):1371–1415
Committee of European Securities Regulators (2007) Best execution under MiFID. https://www.esma.europa.eu/sites/default/files/library/2015/11/07_050b.pdf
Crossley SA, Allen DB, McNamara DS (2011) Text readability and intuitive simplification: a comparison of readability formulas. Read Foreign Lang 23(1):84–101
Debortoli S, Müller O, Junglas I, vom Brocke J (2016) Text mining for information systems researchers: an annotated topic modeling tutorial. Commun Assoc Inf Syst 39(1):110–135
Devlin J, Chang M-W, Lee K, Toutanova K (2018) BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805
Dong W, Liao S, Zhang Z (2018) Leveraging financial social media data for corporate fraud detection. J Manage Inf Syst 35(2):461–487
Du Bois JW, Chafe WL, Meyer C, Thompson SA, Martey N (2000) Santa Barbara corpus of spoken American English. CD-ROM. Linguistic Data Consortium, Philadelphia
Dyer T, Lang M, Stice-Lawrence L (2017) The evolution of 10-K textual disclosure: evidence from latent dirichlet allocation. J Account Econ 64(2–3):221–245

Esuli A, Sebastiani F (2006) Determining term subjectivity and term orientation for opinion mining. In: 11th Conference of the European chapter of the association for computational linguistics, pp 193–200

European Commission (2005) Impact assessment guidelines. https://ec.europa.eu/transparency/regdoc/rep/2/2005/EN/SEC-2005-791-2-EN-MAIN-PART-1.PDF

European Commission (2019) Better regulation: taking stock and sustaining our commitment. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52019DC0178 &from=EN

European Parliament (2020) Financial services policy. https://www.europarl.europa.eu/factsheets/en/sheet/83/financial-services-policy

European Parliament and Council (2014a) Directive 2014/65/EU on markets in financial instruments: MiFID II. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32014L0065 & from=EN

European Parliament and Council (2014b) Regulation (EU) no 1286/2014 on key information documents for packaged retail and insurance-based investment products (PRIIPs). https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32014R1286

European Securities and Markets Authority (2015) Best execution under MiFID—peer review report. https://www.esma.europa.eu/sites/default/files/library/2015/11/2015-494_peer_review_report_on_best_execution_under_mifid_0.pdf

European Securities and Markets Authority (2017) Follow-up report to the peer review on best execution. https://www.esma.europa.eu/sites/default/files/library/esma42-1643088512-2962_follow-up_best_execution_peer_review_report.pdf

Finkel JR, Grenager T, Manning CD (2005). Incorporating non-local information into information extraction systems by Gibbs sampling. In: Proceedings of the 43rd annual meeting of the association for computational linguistics (ACL'05), pp 363–370

Gai P, Kemp M, Sánchez Serrano A, Schnabel I (2019) Regulatory complexity and the quest for robust regulation. Report of the Advisory Scientific Committee 8, European Systemic Risk Board. https://ideas.repec.org/p/srk/srkasc/20198.html

Geerts GL (2011) A design science research methodology and its application to accounting information systems research. Int J Account Inf Syst 12(2):142–151

Glancy FH, Yadav SB (2011) A computational model for financial reporting fraud detection. Decis Support Syst 50(3):595–601

Gomber P, Pujol G, Wranik A (2012) Best execution implementation and broker policies in fragmented European equity markets. Int Rev Bus Res Pap 8(2):144–162

Gozman D, Liebenau J, Aste T (2020) A case study of using blockchain technology in regulatory technology. MIS Q Exec 19(1):19–37

Gregor S, Hevner AR (2013) Positioning and presenting design science research for maximum impact. MIS Q 37(2):337–355

Grootendorst M (2022) BERTopic: neural topic modeling with a class-based TF-IDF procedure. arXiv preprint arXiv:2203.05794

Gunning R (1969) The Fog index after twenty years. J Bus Commun 6(2):3–13

Hanley KW, Hoberg G (2010) The information content of IPO prospectuses. Rev Finan Stud 23(7):2821–2864

Hassan TA, Hollander S, van Lent L, Tahoun A (2019) Firm-level political risk: measurement and effects. Q J Econ 134(4):2135–2202

Hevner AR, March ST, Park J, Ram S (2004) Design science in information systems research. MIS Q 28(1):75–105

Hirschberg J, Manning CD (2015) Advances in natural language processing. Science 349(6245):261–266

Hoberg G, Phillips G (2016) Text-based network industries and endogenous product differentiation. J Polit Econ 124(5):1423–1465

Hope O-K, Hu D, Lu H (2016) The benefits of specific risk-factor disclosures. Rev Acc Stud 21(4):1005–1045

Hu N, Bose I, Koh NS, Liu L (2012) Manipulation of online reviews: an analysis of ratings, readability, and sentiments. Decis Support Syst 52(3):674–684

Huang AH, Lehavy R, Zang AY, Zheng R (2018) Analyst information discovery and interpretation roles: a topic modeling approach. Manage Sci 64(6):2833–2855

Humpherys SL, Moffitt KC, Burns MB, Burgoon JK, Felix WF (2011) Identification of fraudulent financial statements using linguistic credibility analysis. Decis Support Syst 50(3):585–594

Jiang JJ, Conrath DW (1997). Semantic similarity based on corpus statistics and lexical taxonomy. In: Proceedings of research in computational linguistics (ROCLING X), Taiwan, pp 1–15
Kang Y, Cai Z, Tan C-W, Huang Q, Liu H (2020) Natural language processing (NLP) in management research: a literature review. J Manage Analyt 7(2):139–172
Kannan S, Gurusamy V, Vijayarani S, Ilamathi J, Nithya M (2014) Preprocessing techniques for text mining. Int J Comp Sci Commun Netw 5(1):7–16
Kathuria A, Gupta A, Singla RK (2021) A review of tools and techniques for preprocessing of textual data. Computational methods and data engineering. Springer Singapore, Singapore, pp 407–422
Katz DM, Bommarito MJ (2014) Measuring the complexity of the law: the United States Code. Artif Intell Law 22(4):337–374
Kelly B, Papanikolaou D, Seru A, Taddy M (2018) Measuring technological innovation over the long run. Tech. rep, National Bureau of Economic Research
Kim C, Wang K, Zhang L (2019) Readability of 10-K reports and stock price crash risk. Contemp Account Res 36(2):1184–1216
Kirkos E, Spathis C, Manolopoulos Y (2007) Data mining techniques for the detection of fraudulent financial statements. Expert Syst Appl 32(4):995–1003
Kirkpatrick C, Parker D (2004) Editorial: regulatory impact assessment—an overview. Public Money Manage 24(5):267–270
Korenius T, Laurikkala J, Juhola M (2007) On principal component analysis, cosine and Euclidean measures in information retrieval. Inf Sci 177(22):4893–4905
Lacity MC, Janson MA (1994) Understanding qualitative data: a framework of text analysis methods. J Manage Inf Syst 11(2):137–155
Landauer TK, Foltz PW, Laham D (1998) An introduction to latent semantic analysis. Discourse Process 25(2–3):259–284
Lang M, Stice-Lawrence L (2015) Textual analysis and international financial reporting: large sample evidence. J Account Econ 60(2–3):110–135
Laruelle S, Lehalle C-A (2018) Market microstructure in practice. World Scientific Publishing, Danvers, MA
Lau JH, Baldwin T (2016) An empirical evaluation of doc2vec with practical insights into document embedding generation. In: Proceedings of the 1st workshop on representation learning for NLP, pp 78–86
Lausen J, Clapham B, Siering M, Gomber P (2020) Who is the next Wolf of Wall Street? Detection of financial intermediary misconduct. J Assoc Inf Syst 21(5):1153–1190
Le Q, Mikolov T (2014) Distributed representations of sentences and documents. In: Proceedings of the 31st international conference on machine learning, pp 1188–1196
Lee LH, Wan CH, Rajkumar R, Isa D (2012) An enhanced support vector machine classification framework by using Euclidean distance function for text document categorization. Appl Intell 37(1):80–99
Li F (2008) Annual report readability, current earnings, and earnings persistence. J Account Econ 45(2–3):221–247
Li W, Azar P, Larochelle D, Hill P, Lo AW (2015) Law is code: a software engineering approach to analyzing the united states code. J Bus Technol Law 10:297–374
Libby R, Libby PA, Short DG, Kanaan G, Gowing M (2004) Financial accounting. McGraw-Hill/Irwin, Boston, MA
Loughran T, McDonald B (2014) Measuring readability in financial disclosures. J Financ 69(4):1643–1671
Loughran T, McDonald B (2016) Textual analysis in accounting and finance: a survey. J Account Res 54(4):1187–1230
Lundholm RJ, Rogo R, Zhang JL (2014) Restoring the tower of Babel: how foreign firms communicate with US investors. Account Rev 89(4):1453–1485
MacQueen J et al (1967) Some methods for classification and analysis of multivariate observations. In: Proceedings of the 5th Berkeley symposium on mathematical statistics and probability, no. 14, Oakland, CA, pp 281–297
Mikolov T, Chen K, Corrado G, Dean J (2013) Efficient estimation of word representations in vector space. arXiv preprint arXiv: 1301.3781

Moyano JP, Ross O (2017) KYC optimization using distributed ledger technology. Bus Inf Syst Eng 59(6):411–423
OECD (1995) Recommendation of the council on improving the quality of government regulation. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0278
OECD (1997) Regulatory impact analysis: best practices in OECD countries. http://www.oecd.org/regreform/regulatory-policy/35258828.pdf
OECD (2004) Regulatory performance: ex-post evaluation of regulatory tools and institutions. http://www.oecd.org/regreform/regulatory-policy/34227774.pdf
OECD (2008) Building an institutional framework for regulatory impact analysis (RIA)—guidance for policy makers. http://www.oecd.org/regreform/regulatory-policy/40984990.pdf
OECD (2018) OECD regulatory policy outlook 2018. OECD Publishing, Paris. https://www.oecd.org/governance/oecd-regulatory-policy-outlook-2018-9789264303072-en.htm
OECD (2020) Regulatory impact assessment—OECD best practice principles for regulatory policy. https://www.oecd-ilibrary.org/governance/regulatory-impact-assessment_7a9638cb-en
Pak A, Paroubek P (2010). Twitter as a corpus for sentiment analysis and opinion mining. In: Proceedings of the 7th conference on international language resources and evaluation (LREC), vol 10. Valletta, Malta, pp 1320–1326
Peffers K, Tuunanen T, Rothenberger MA, Chatterjee S (2007) A design science research methodology for information systems research. J Manage Inf Syst 24(3):45–77
Peffers K, Rothenberger M, Tuunanen T, Vaezi R (2012) Design science research evaluation. International conference on design science research in information systems. Springer, Berlin, Heidelberg, pp 398–410
Pierrehumbert JB (2001) Exemplar dynamics: word frequency. Freq Emerg Linguist Struct 45:137–157
Radaelli CM (2004) The diffusion of regulatory impact analysis—best practice or lesson-drawing? Eur J Polit Res 43(5):723–747
Radaelli CM (2018) Halfway through the better regulation strategy of the Juncker Commission: what does the evidence say? J Common Market Stud 56:85–95
Raghuveer K (2012) Legal documents clustering using latent dirichlet allocation. IAES Int J Artif Intell 2(1):34–37
Ramos J et al (2003) Using TF-IDF to determine word relevance in document queries. In: Proceedings of the first instructional conference on machine learning, vol 242. New Jersey, USA, pp 29–48
Reichmann D, Möller R, Hertel T (2022) Nothing but good intentions: the search for equity and stock price crash risk. J Bus Econ 1–35
Reshamwala A, Mishra D, Pawar P (2013) Review on natural language processing. IRACST Eng Sci Technol Int J 3(1):113–116
Salton G, Buckley C (1988) Term-weighting approaches in automatic text retrieval. Inf Process Manage 24(5):513–523
Saunders A, Cornett MM (2012) Financial markets and institutions. McGraw-Hill/Irwin, Boston
Siering M, Clapham B, Engel O, Gomber P (2017) A taxonomy of financial market manipulations: establishing trust and market integrity in the financialized economy through automated fraud detection. J Inf Technol 32(3):251–269
Simon HA (1996) The sciences of the artificial. MIT Press, Cambridge, MA
Singhal A, Salton G, Mitra M, Buckley C (1996) Document length normalization. Inf Process Manage 32(5):619–633
Tan C-M, Wang Y-F, Lee C-D (2002) The use of bigrams to enhance text categorization. Inf Process Manage 38(4):529–546
The Law Library of Congress, Global Legal Research Directorate (2019) Government responses to disinformation on social media platforms, report for congress LL file no. 2019-017919. https://www.hsdl.org/?view &did=835597
Tibshirani R, Walther G, Hastie T (2001) Estimating the number of clusters in a data set via the gap statistic. J Roy Stat Soc Ser B (Stat Methodol) 63(2):411–423
U.S. Securities and Exchange Commission (2009) Enhanced disclosure and new prospectus delivery option for registered open-end management investment companies. https://www.sec.gov/rules/final/2009/33-8998.pdf

Vijayarani S, Ilamathi MJ, Nithya M et al (2015) Preprocessing techniques for text mining—an overview. Int J Comp Sci Commun Netw 5(1):7–16

Williams JW (2013) Regulatory technologies, risky subjects, and financial boundaries: governing "fraud'' in the financial markets. Acc Organ Soc 38(6–7):544–558