

Research Report

Regulatory Impact Analysis in Case of Unstructured Data

REGULATORY IMPACT ANALYSIS (RIA) SERVES 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 COMMUNICATING TEXTUAL DATA TO CONSUMERS AND SUPERVISORS. THEREFORE, WE DEVELOP A METHODOLOGICAL FRAMEWORK FOR RIA IN CASE OF UNSTRUCTURED DATA BASED ON TEXTUAL ANALYSIS AND APPLY IT TO A RECENT FINANCIAL MARKET REGULATION: MIFID II.

Benjamin Clapham

Jens Lausen

Micha Bender

Peter Gomber

Introduction

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

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 (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. To enable regulators and researchers to assess the impact of regulatory actions aimed

at unstructured data and to improve evidence-based policy making with the help of regulatory intelligence and RegTech-solutions, our paper (Clapham et al., 2023) develops a methodological framework for RIA in case of unstructured data. It builds on methods from textual analysis (TA) and natural language processing (NLP). We evaluate the RIA-framework and make use of it to assess the impact of a recent regulation: the changes in best execution requirements of the Markets in Financial Instruments Directive II (MiFID II) in Europe that are applicable since January 2018. These rule changes require investment firms, i.e., banks and brokers, 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 order routing to achieve the best possible result for their clients. Thus, these policies should enhance transparency for investors and protect them from potential downsides of the existing stock market fragmentation in Europe.

Development of the Framework

Following the design science research paradigm (Hevner et al., 2004), we develop a framework for the analysis and evaluation of regulatory actions that result in unstructured data such as text documents. To create the framework, we build on existing RIA guidelines and on methods from TA and NLP. The RIA-framework provides detailed guidance as well as 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. The framework (shown in Figure 1) can be applied as follows:

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

Step 2: Identify the specific dimensions that are affected by the regulatory action. Dimensions in this context refer to the means (e.g., informativeness) that are targeted by the regulation in order to achieve the identified regulatory goal.

Step 3: Acquire the necessary data. 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, i.e., text documents, before as well as after the introduction of the regulatory change have to be collected. In case of the introduction of a new regulation, the collection of data before the regulatory change is mostly impossible since it often requires firms to publish new textual documents. Therefore, we propose a benchmark approach to assess the impact of new regulations. To collect appropriate benchmarks (e.g., textual data generated in similar regulatory areas), it is important to ensure comparability between affected objects and chosen benchmarks.

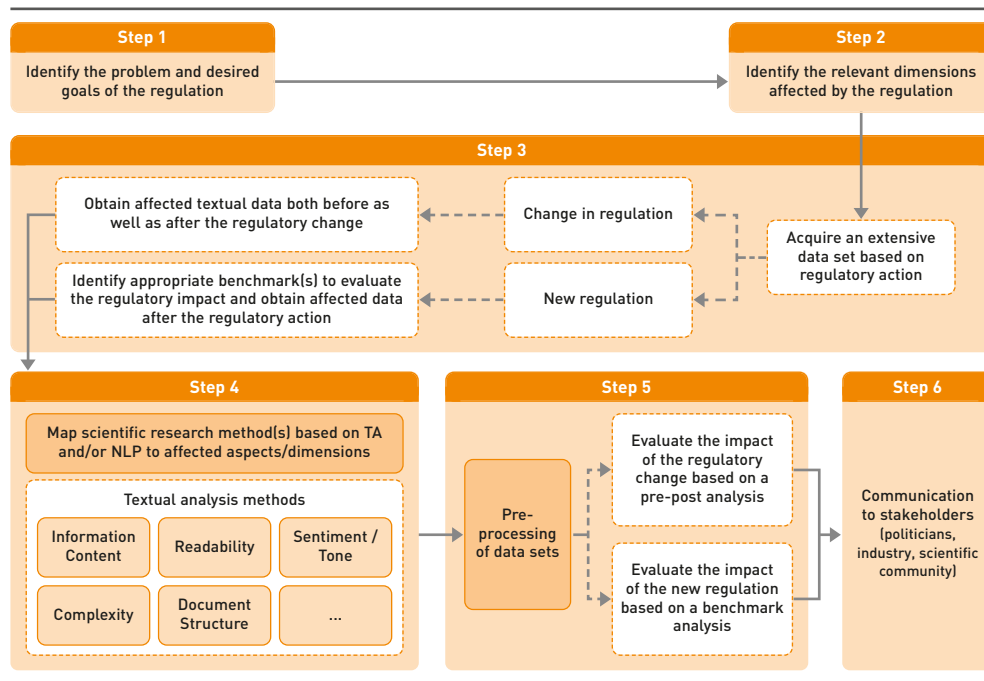


Figure 1: RIA-Framework for Unstructured Data

Step 4: Map appropriate scientific research method(s) to affected regulatory dimensions. The mapping of TA and NLP methodologies (see, e.g., Loughran and McDonald, 2016) 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.

Step 5: Preprocess data and evaluate the impact of the regulatory change or of a new regulation. This step 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. Relevant textual information needs to be extracted and the handling of textual information from figures, tables, and lists needs to be determined. Depending on the methodology used for the analysis, further text cleaning steps such as removing stopwords or stemming need to be performed. For the assessment of the regulatory actions, the TA and NLP methods determined in

Step 4 are applied. Once these analyses are conducted, the results need to be compared, e.g., by using data visualization and statistical tests to evaluate whether changes in the analyzed measures can be observed and whether these changes correspond to the regulatory objectives.

Step 6: Communicate results. The final step of the framework represents the communication of the results of the RIA to relevant stakeholders, e.g., policy makers, regulators, consumer protection associations, and the scientific community, by publishing a policy white paper or a research report.

Empirical Evaluation: Best Execution Policies in MiFID II

To demonstrate its applicability, we use the RIA-framework to perform a regulatory impact analysis of the best execution requirements outlined in MiFID II. With more than 300 trading venues (as of September 2022), the European securities market is highly fragmented. Consequently, there is a large choice of venues for the execution of an order in a financial instrument. The selection of the appropriate trading venue by the investment firm for the customer order shall ensure the best possible result for the client ("best execution"). In this process, investment firms have to take into account a range of factors such as price, costs, speed, likelihood of execution and have to document their choices in best execution policies that are accessible to their clients. With MiFID II, which has to be applied since January 2018,

European authorities intend to improve best execution policies by requiring firms to specify and publish policies that are more *informative* and *comprehensible* in order to provide value to clients and to foster competition between investment firms.

We collect best execution policies of 50 German banks and brokers before as well as after the application of MiFID II. For a sample of 187 European best execution policies, we also perform a benchmark analysis and compare it to textual data generated in similar areas. In order to assess the *informational content* of the policies, we rely on three different measures: textual similarity, the percentage of boilerplate information, and the share of specific information in the documents. For the analysis of the policies' *comprehensibility*, we rely on two sets of measures: (1) readability measures such as the average number of words per sentence or the Fog index, which estimates the years of formal education a person needs to understand a text on the first reading; (2) textual complexity measures such as document length or the number of conditional statements in a document.

In the pre-post-analysis of German best execution policies, we find that all analyzed measures did not change or even worsened after MiFID II went live. Our results show that best execution policies became harder to read and more complex after MiFID II. Also, we find that the policies became more similar, which indicates that the information provided in these documents is less

useful for investors to differentiate between the providers of order execution services. Besides, the policies remain relatively unspecific and include a large share of boilerplate information. Most policies recite large parts of the regulation, which is neither informative for clients nor does it foster competition between brokers based on how they handle client orders to achieve best execution.

Looking at the European sample of best execution policies, we find additional evidence that MiFID II failed to achieve the desired goals. 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 (see Figure 2). Policies aimed at retail clients are as hard to read as policies aimed at professional clients. Based on the benchmark analysis, we find that best execution policies are among the most difficult and complex texts among the benchmarks. Figure 2 exemplarily shows that best execution policies 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 results exist regarding textual complexity. Based on the number of conditional statements, best execution policies belong to the most complex documents. Moreover, best execution policies use a large variety of vocabulary as measured by the number of unique bigrams. They are comparable to the management discussion section of annual reports. In summary, the analysis shows that best execution policies are not easy to understand as intended by the regulator but are highly complex documents that are difficult to read. The 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.

Discussion of the Results

More and more regulations aim at or result in huge numbers of textual documents. To enable researchers and regulators to assess whether regulatory actions have met the desired goals, this study develops a framework for RIA in case of unstructured data and applies it to the MiFID II requirements on best execution policies. We show that the regulation has not achieved its goal to increase informativeness and comprehensibility for customers. 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. Although the framework is based on a use case from finan-

cial regulation, the framework should be applicable also to regulatory initiatives of other economic sectors as no step of the framework is unique to the financial industry. Rather, the framework represents a general principle to solve a class of real-world problems, i.e., conducting RIA in case of unstructured data.

References

Clapham, B.; Bender, M.; Lausen, J.; Gomber, P.: Policy Making in the Financial Industry: A Framework for Regulatory Impact Analysis Using Textual Analysis. Forthcoming in: Journal of Business Economics.

Hevner, A. R.; March, S. T.; Park, J.; Ram, S.: Design Science in Information Systems Research. In: MIS Quarterly, 28 (2004) 1, pp. 75–105.

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

OECD: Regulatory Impact Analysis: Best Practices in OECD Countries, <http://www.oecd.org/regreform/regulatory-policy/35258828.pdf>, 1997.

Radaelli, C. M.: The Diffusion of Regulatory Impact Analysis – Best Practice or Lesson-Drawing? In: European Journal of Political Research, 43 (2004) 5, pp. 723–747.

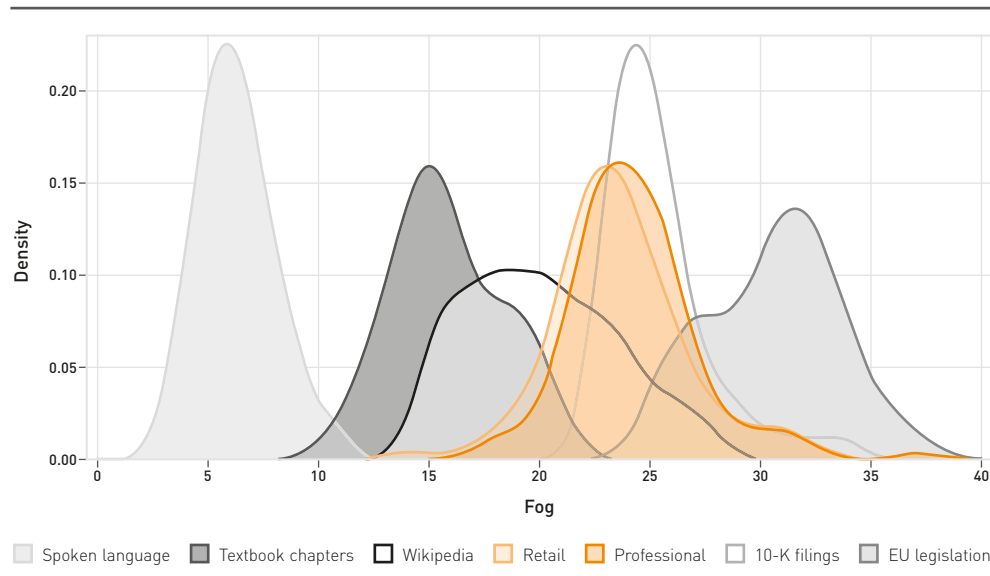


Figure 2: Comparison of the Readability of Best Execution Policies and Benchmark Documents Based on the Fog Index