FROM PROPOSAL TO POLICY: ASSESSING THE RELATIONSHIP BETWEEN PUBLIC COMMENTS AND US FEDERAL REGULATIONS USING NATURAL LANGUAGE PROCESSING.

By Vladislav Shatilov

Submitted to Central European University – Private University
Department of Public Policy

In partial fulfilment of the requirements for the degree of Master of Public Administration

Supervisor: Associate Professor Mihály Fazekas

Vienna, Austria
2024

Author's Declaration

I, the undersigned, Vladislav Shatilov, candidate for Master of Public Administration, declare herewith that the present thesis is exclusively my own work, based on my research.

All sources have been properly credited in the text, notes, and the bibliography. I declare that no unidentified and illegitimate use was made of the work of others, and no part of the thesis infringes on any person's or institution's copyright. Furthermore, I declare that no part of this thesis has been generated using Large Language Models.

I also declare that no part of the thesis has been submitted in this form as coursework for credits or to another institution of higher education for an academic degree.

Date: 03.06.2024

Name: Vladislav Shatilov

Signature: ____________________

Abstract

Public commenting has become very important in regulatory policy globally, aiming to enhance regulation quality and transparency. Despite widespread adoption and substantial engagement from both the public and governmental bodies, the association between public commenting and regulation revisions remains underexplored. Existing research offer limited generalizability by focusing narrowly on specific domains or using small datasets. To address these gaps, this research introduces a novel approach that evaluates the relationship between the content of public comments and changes in regulatory texts in the United States. Multiple text similarity techniques are utilized to compare the initial and final versions of regulations. The analysis employs multiple regression to analyse the relationship and a machine learning model to detect non-linear patterns of it. Key variables examined include the share of words by which differ two versions of regulations in comments, the volume of comments, and their emotional tone. The key finding is that larger revisions of regulatory texts are strongly associated with greater stakeholders' attention to the subsequently revised parts, while the effect of other factors is unclear or minimal.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

APA	Administrative Procedures Act
API	Application Programming Interface
CFR	Code of Federal Regulations
GSA	General Services Administration
OECD	Organization for Economic Co-operation and Development
PDP	Partial Dependency Plot
VIF	Variance Inflation Factor

Introduction

Public consultations have become a popular approach in regulatory policy across many countries, with primary goals of improving the quality of regulations and ensuring transparency for the public. Between 2017 and 2018, 135 out of 186 countries had their developed regulations available to the public, facilitating essential information for regulatory development, and many of these countries conduct public consultations.[1] Often, this is done using electronic tools. For instance, 65% of European and Central Asian countries held consultations through unified websites,[2] a process known as public commenting. The popularity of this method is well-founded: regulatory institutions should be accountable to the public, allowing individuals to influence regulatory policy and challenge unnecessary or unjust regulations.

However, the question remains: do public comments actually affect the regulations on which they are posted? The level of engagement in this procedure is indeed high, with many comments posted on regulations by electronic platforms, creating significant workloads for agencies.[3] The attention of governments to the procedure is also significant.[4] Given the attention and costs associated with this process, there should be clear evidence that it produces positive outcomes for both governments and the public. This is especially relevant for the United States, which was the first country to adopt this procedure and where the process is notably open and inclusive.[5] Demonstrating its effectiveness could encourage other countries to implement or improve similar procedures.

Fortunately, evidence does show that public comments are positively associated with changes in regulations. However, the generalizability of these results is limited, as they often rely on narrow data sets or focus on specific domains. This fragmented evidence highlights the complex nature of the relationship but makes it difficult to apply findings across different domains and regulations. Additionally, existing research tends to focus on the commenters rather than the aspects related to the comments themselves. It is essential to understand which

[1] "Global Indicators of Regulatory Governance", World Bank, accessed May 25, 2024, https://rulemaking.worldbank.org/en/key-findings
[2] Ibid.
[3] Ari-Veikko Anttiroiko, Gloria T. Lau, Kincho H. Law, "A Prototype Study on Electronic Rulemaking", (2008): 10.
[4] "Consultations - what's new and why they are so important", GOV.UK, published January 15, 2016, https://civilservice.blog.gov.uk/2016/01/15/consultations-whats-new-and-why-they-are-so-important/
[5] OECD, *Background Document on Public Consultation*, (Paris: OECD, n.d.), 3.

attributes of comment content make them more or less influential, as this knowledge is valuable both academically and practically for stakeholders and agencies.

To achieve high external validity, more extensive data collection is necessary. Analyzing large datasets presents methodological challenges that cannot be addressed with the qualitative methods typically used in past studies in the field. This task requires an enhanced methodology capable of processing vast amounts of data while addressing specific questions.

In this paper, I aim to provide robust evidence supporting the relationship between public commenting and regulatory development, ensuring generalizability across all domains. I focus on the US context for the reasons mentioned above. I introduce a novel tool to determine whether content of public comments has association with revisions of regulatory texts and identify which attributes of the comments are the most prominent in this association. By comparing the initial proposed versions of regulations, which are subject to public comments, with the final enacted versions, I can assess the effect of commenting. I use diverse text similarity techniques to estimate differences between versions in superficial, textual, and semantic terms. This approach links comments to differences between rules by identifying the words by which differ two versions of regulations and calculating their presence in the comments. Additionally, I consider the number of comments and their emotional tone as variables. The primary analytical tool is a Multiple Regression Model, which effectively captures the linear effects of several variables on one outcome. To account for potential non-linear effects, I also employ a machine learning model, providing valuable insights into the true form of these relationships.

Literature review

Rules and comments

There is no universally accepted definition of "regulation" and "rule" but to have one, a comprehensive definition should apply universally across all contexts, as they can vary in detail. OECD provide such a definition and describes regulation as a “diverse set of instruments by which governments set requirements on enterprises and citizens”, where regulation can be in various formats and “include laws, formal and informal orders and subordinate rules issued by all levels of government, and rules issued by non-governmental or self-regulatory bodies to whom governments have delegated regulatory powers”, and can be categorized into three groups—economic, social, or administrative—depending on their focus.[6] Another definition posits that regulation refers to public policies designed to control economic activities and their impacts at the levels of industries, companies, or individual units, while regulatory policy pertains to the development of these regulations.[7] Although primary legislation adopted by parliaments, or simply "laws," are often included in these definitions, as in the first it is even stated explicitly, the term "regulation" is frequently used to describe secondary legislation, which is adopted not by parliament but by executives, especially in the UK and European Union.[8] The term "rule" is narrower and typically applies only to legal acts adopted by agencies, and this term used predominantly in the American context. Administrative Procedures Act (APA) of USA provides extensive definition where the rule is “the whole or part of an agency statement of general or particular applicability and future effect designed to implement, interpret, or prescribe law or policy or describing the organization, procedure or practice requirements of an agency” and any action aimed in developing those rules can be called rulemaking.[9] This definition explicitly states that rulemaking is solely within the purview of agencies. In this paper, the terms "regulation" and "rule" are used interchangeably to denote documents developed and adopted by executive or independent agencies, containing legally binding provisions and ranking below primary legislation, or laws, in the judicial hierarchy.

[6] OECD, *The OECD Report on Regulatory Reform*, (Paris: OECD, 1997), 6.
[7] Marc Allen Eisner, Jeffrey Worsham, and Evan J. Ringquist, *Contemporary regulatory policy*, (London: Lynne Rienner Publishers, 2000), 3.
[8] “What is Secondary Legislation?”, UK Parliament, accessed April 29, 2024, https://www.parliament.uk/about/how/laws/secondary-legislation/ ;
“European Union regulations”, EUR-lex, published March 16, 2022, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=LEGISSUM:l14522
[9] United States. Administrative Procedure Act. Pub. L. No. 79-404, 60 Stat. 237 (1946).

Rulemaking, as it is defined above, has became a significant phenomenon affecting citizens' lives in the 20th century. Although various executive bodies were established much earlier as states developed, they had quite different form. It was only in the last 100-150 years that more modern-looking regulatory bodies emerged. When these bodies first appeared, their operations often lacked regulation and structure, with unclear credentials. Generally, there was no standardized set of rules for their operation, they had limited powers to issue regulations, and their activities were non-transparent.[10] The latter means that they were barely accountable to citizens, as their operations were largely autonomous from them and citizens could only influence their decisions very indirectly, through legislative oversight. The situation changed in the mid-20th century, primarily after the enactment of the aforementioned APA in the USA in 1946. This act brought structure to the work of agencies, standardizing it and establishing rules on how agencies should conduct regulatory policy. One of the innovations was the introduction of "notice and comment" rulemaking, which required agencies to notify the public about new regulations, provide important information about them, and offer adequate opportunities for interested parties to comment.[11]

The introduction of this tool for engaging interested individuals and groups represents a significant milestone in rulemaking. Particularly towards the end of the 20th century, the practice of involving citizens in the development of regulations came to be recognized as a best practice as it is viewed as a means to enhance the transparency of the rulemaking process. Following its introduction in the USA, "notice-and-comment" practice was later adopted by Canada and Portugal. By the end of the 20th century, at least 19 OECD countries were using this procedure in some form.[12]

Generally, notice-and-comment and public commenting are one form of public consultations, but it can be very different and include range of communicative activities with public, including public opinion surveys, hearings, negotiations, panels, focus groups etc.[13] Public consultations can also be incorporated into Regulatory Impact Assessments, or vice versa, which highlights their interconnectedness.[14]

[10] Kathryn E. Kovacs, "Rules about Rulemaking and the Rise of the Unitary Executive", *Administrative Law Review*, Vol. 70 (2018): 519.
[11] Susan E. Dudley, "Milestones in the Evolution of the Administrative State", *Daedalus*, Vol. 150, 3 (2021): 37.
[12] OECD, *A MENA-OECD Practitioners' Guide for Engaging Stakeholders in the Rule-making Process*, (Paris: OECD, 2012).
[13] Gene Rowe, Lynn J. Frewer, "Public Participation Methods: A Framework for Evaluation", *Science, Technology, & Human Values*, Vol. 25 (2000): 8-9.
[14] OECD, *Regulatory Impact Analysis: Best Practices in OECD Countries*, (Paris: OECD, 1997), 17.

OECD emphasize the need for public engagement and consultation as it gives the opportunity to the public to express opinion on laws and rules and it is bringing the space for improvement of them as broad groups of interests can better assess what these laws and rules will mean for them in practice.[15] Moreover, public participation in the rulemaking process is critical for increasing public support for regulations and the subsequent reduction of enforcement costs as well as improved compliance. The provision also offers a learning experience to the citizens about the nuances of regulations, deepening their knowledge base and their engagement in governance. The purpose of consultations is to take into account the impact of regulations, align various conflicting interests early as well as to keep notions about the public good in constant motion by encouraging an open dialogue. The engagement fosters trust in government, todding the legal security and social cohesion as the different groups come together to solve community issues. Therefore, feedback from the public ensures the lawfulness of governance and guarantees the fulfillment of community's requirements.[16]

US Context

In this paper, as I explore the context of rulemaking in the US, I focus on the "notice-and-comment" type of public consultations. However, to analyze this effectively, it is essential to understand why and how it emerged and how it functions.

Regulatory policy in the USA dates back to the late 19th - early 20th century when the federal government began establishing regulatory agencies to address various societal and economic challenges. The earliest was the Interstate Commerce Commission, created in 1887 to regulate the railroad industry.[17] Initially, these agencies faced significant operational limitations because courts interpreted the separation of powers provisions in the Constitution as prohibiting the delegation of Congress's powers to agencies. However, this interpretation softened in 1928 when the Supreme Court allowed Congress to delegate legislative powers to agencies, provided that it established clear standards for rulemaking.[18]

Although there were concerns about the extent of regulation, support for anti-crisis measures of current isteblishment prevailed, leading to the adoption of the APA in 1946 as a

[15] OECD, *Regulatory Policy Outlook 2021: Evidence-based policy making and stakeholder engagement,* (Paris: OECD, 2021).
[16] OECD, *A MENA-OECD Practitioners' Guide for Engaging Stakeholders in the Rule-making Process*, (Paris: OECD, 2012), 9-10.
[17] Susan E. Dudley, "Milestones in the Evolution of the Administrative State", *Daedalus*, Vol. 150, 3 (2021): 34.
[18] Ibid, 35.

compromise between bureaucratic expertise and legislative accountability.[19] The APA introduced few main changes in the rulemaking: first, regulations must be grounded in statutory law, second, final rules became subject to judicial review and most importantly for the analysis, the administrative record must include a public notice-and-comment period.[20]

Nowadays the process of rule development has not changed significantly and continues to be based on the APA. However, with the advancement of technology, e-rulemaking has emerged, and procedures now incorporate the use of technological tools. Not all the rules are passing the procedures depicted in APA, as it depends on the type of the rule. Rules can be divided into two categories: legislative rules, which are "the product of an exercise of delegated legislative power to make law through rules," and non-legislative rules, which include interpretative rules and policy statements.[21] Legislative rules require publication in the Federal Register, the official daily publication for rules and other executive documents,[22] and generally must adhere to all requirements and procedures outlined in the APA. In contrast, non-legislative rules are a different category that do not undergo the full suite of procedures, most notably the notice-and-comment process so there is no public feedback on those rules. [23]

This paper focuses exclusively on legislative rules and procedures of their development. Firstly, when an agency identifies the need for new regulation based on various factors, it initiates the development process. This may involve devising a plan and gathering information about the issue through informal processes, such as collecting public opinions. Also, the agency might notify the public in advance by publishing a notice in the Federal Register. This allows anyone to contribute to the development of the proposed rule.[24] Once it is formulated, the Notice of Proposed Rulemaking or just proposed rule is published in the Federal Register. This document provides justification and explanation for the potential regulation and typically includes the regulatory text that proposes amendments to the Code of Federal Regulations (CFR). The CFR is a codification of the general and permanent rules published by federal agencies and departments.[25] The agency then announces the start of the commenting period,

[19] Ibid, 36.
[20] Ibid.
[21] Robert A. Anthony, "Interpretive Rules, Policy Statements, Guidances, Manuals, And The Like-Should Federal Agencies Use Them To Bind The Public", Duke Law Journal, Vol. 41 (1992): 1322-1323.
[22] "Federal Register", GovInfo, accessed December 5, 2024, https://www.govinfo.gov/app/collection/fr
[23] Robert A. Anthony, "Interpretive Rules, Policy Statements, Guidances, Manuals, And The Like-Should Federal Agencies Use Them To Bind The Public", Duke Law Journal, Vol. 41 (1992): 1322-1323.
[24] Office of the Federal Register, *A Guide to the Rulemaking Process*, (Washington D.C: Office of the Federal Register, n.d.), 3.
[25] "Code of Federal Regulations List of Subjects", National Archives, accessed December 13, 2023, https://www.archives.gov/federal-register/cfr/subjects.html

during which anyone can submit comments via the Regulations.gov website or through more traditional methods.[26] The agency may also hold public hearings and other forms of consultation to gather broader input. After the commenting period, the agency must consider all feedback received. If the process advances to the Final Rule, the agency uses all the arguments from the comments and other collected information to finalize the document. The Final Rule is then published in the Federal Register, including an effective date and the regulatory text which embodies amendments to the CFR. Depending on the public input, agencies may also decide to terminate the regulation process.[27]

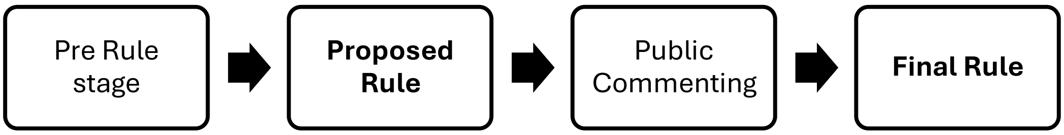


Figure 1. Stages of US Federal Rulemaking.

This outline describes the basic and usual procedural pathway for rulemaking, though variations do occur. For instance, agencies may expedite the process through fast-track regulation when the changes are not substantive or pertain to uncontroversial matters. In such cases, they might post the final rule and enact it immediately after the commenting period, provided there are no negative comments.[28] Additionally, agencies sometimes can extend the commenting period or publish supplementary rules.[29]

Do comments matter?

Given the widespread use of public commenting and notice-and-comment procedures, it is crucial that they function effectively and yield positive outcomes. Should these procedures fail to have an influence, or worse, predominantly produce negative effects, the entire practice could be called into question, challenging its necessity. Although the influence of comments

[26] "Regulations.gov", U.S. General Services Administration, accessed December 13, 2023, https://www.regulations.gov/
[27] Office of the Federal Register, *A Guide to the Rulemaking Process*, (Washington D.C: Office of the Federal Register, n.d.), 8-9.
[28] "Regulation Room", Cornell University, accessed April 15, 2024, http://archive.regulationroom.org/learn-more/stages-of-the-rulemaking-process/index.html
[29] Office of the Federal Register, *A Guide to the Rulemaking Process*, (Washington D.C: Office of the Federal Register, n.d.), 5-6.

possibly can be different, the most evident effect I refer to here is the influence of comments on regulations. This demonstrates that agencies genuinely consider feedback and make adjustments to the rules based on that input.

The absence of influence from public comments can typically be attributed to two main factors: resistance by agencies to adjust rules based on the feedback and the low quality of the comments, which do not facilitate meaningful changes in regulations. Concerning the former, if commenters perceive that their input has no visible effect due to a lack of agency responsiveness, they may become disillusioned and either participate less or disengage completely in the future. This phenomenon is one of the primary causes of "consultation fatigue," which leads to the withdrawal of individuals and groups from the participation process.[30] As for the quality of comments, it has been repeatedly observed that mass comments organized by cohesive groups often offer little value due to their lack of substantiveness.[31]

For various governments, understanding whether and how comments influence regulations is crucial. This is especially true for the USA, where the system of public commenting is "extremely open and accessible",[32] and because of that developing a regulation requires "substantial time and effort".[33]

Researchers have been trying to gauge the influence of the public commenting procedure for decades. It has been discovered that early commenters significantly influence the development of regulations, and that generally public can help by thwarting or eliminating undesirable regulations.[34] Also, the greater participation in the public comment process, especially for highly complex and low-salience rulemakings, can yield significant revisions of the regulation according to comments.[35] Agencies are more inclined to adjust rules following recommendations from larger, well-resourced lobby groups, especially for rules with low policy salience, although the influence of members of Congress and industry groups has also been noted.[36] Additionally, the volume of comments matter and agencies weigh comments

---

[30] Rex Deighton-Smith, “Regulatory transparency in OECD countries:Overview, trends and challenges”, *Australian Journal of Public Administration*, Vol. 63 (2004): 69.

[31] Steven J. Balla, Bridget C.E. Dooling, Reeve Bull, Emily Hammond, Michael Herz
Michael Livermore, Beth Simone Noveck, “Responding to Mass, Computer-Generated, and Malattributed Comments”, Administrative Law Review, Vol. 74 (2022):105-106.

[32] OECD, *Background Document on Public Consultation*, (Paris: OECD, n.d.), 4.

[33] “Pilot database on stakeholder engagement practices”, OECD, accessed March 20, 2024, https://www.oecd.org/gov/regulatory-policy/pilot-database-on-stakeholder-engagement-practices.htm

[34] Keith Naughton, Celeste Schmid, Susan Webb Yackee, Xueyong Zhan, “Understanding commenter influence during agency rule development”, *Journal of Policy Analysis and Management*, Vol. 28 (2009): 273-274

[35] Stuart Shapiro. “Does the amount of participation matter? Public comments, agency responses and the time to finalize a regulation”, *Policy Sciences*, Vol. 41 (2008): 43.

[36] Maraam A. Dwidar, “Diverse Lobbying Coalitions and Influence in Notice-and-Comment Rulemaking”, Policy Studies Journal, Vol. 50, No. 1 (2022): 201;

from expert organizations higher that from other types of actors.[37] Generally, it was found that opinions expressed during the notice-and-comment period can impact regulations.[38] Hence, the effect is present, but is highly multifaceted and nuanced.

There are two considerations. First, most of these studies employ relatively narrow data samples. This limitation is manifested either in the size of the datasets, which typically do not exceed dozens of regulations and hundreds of comments, or in the sampling—studies often focus on rules from specific agencies rather than a common pool of regulations. This approach allows researchers to employ qualitative methods to delve deeper into the content of rules and comments but limits the external validity of the findings. Given that US agencies create thousands of rules every few months, these small samples provide limited insights for generalizable results.

Secondly, existing research often focuses on the attributes of the commenters, but less attention has been given to the attributes of the comments themselves. Some studies have investigated the emotionality of comments posted on regulations. For instance, research in Germany used dictionary coding of comments to analyze the distribution of different emotions. It was found that ordinary citizens are more likely to write emotional comments compared to other types of actors, and comments on specific topics tend to elicit more emotional responses.[39] Although it is not directly related to public comments, sentiments of tweets were analyzed during the public consultation period for a controversial rule in the USA. The study found that the types of emotions and levels of positivity in Twitter discussions on the topic can vary before the start of public consultations, during, and after them.[40] However, influence of emotional aspect of comments on regulations was not studied.

Considering all of this, I aim to continue research efforts in this direction. In this paper, I address the question of *does content of public comments is associated with revision of regulatory texts, and which attributes of these comments are the mostly significant in this relationship.* The main novelty of this study that I utilize a much bigger and more extensive

---

Mia Costa, Bruce A. Desmarais, and John A. Hird, "Public Comments' Influence on Science Use in U.S. Rulemaking: The Case of EPA's National Emission Standards", *American Review of Public Administration,* Vol. 49 (2019): 46

[37] Alex Ingrams, "Do public comments make a difference in open rulemaking? Insights from information management using machine learning and QCA analysis", Government Information Quarterly, Vol. 40 (2023): 8.

[38] Susan Webb Yackee, "Sweet-Talking the Fourth Branch: The Influence of Interest Group Comments on Federal Agency Rulemaking", *Journal of Public Administration Research and Theory,* Vol. 16 (2006): 118.

[39] Simon Fink, Eva Ruffing, Tobias Burst, Sara Katharina Chinnow, "Emotional citizens, detached interest groups? The use of emotional language in public policy consultations", *Policy Sciences,* Vol. 56 (2023): 489.

[40] Kayla Schwoerer, "An exploratory study of social media's role in facilitating public participation in e-rulemaking using computational text analysis tools", *Policy & Internet*, Vol. 15 (2023): 187-188.

dataset, taking regulations from a common pool rather than focusing on specific domains or agencies. It enhances the external validity of the study and allows the use of quantitative and computational methods.

I primarily investigate whether we can demonstrate the general presence of the of association between comments on regulations and regulations themselves. Previous studies confirm the presence of relationship but highlight its complexity. I hypothesize that comments and their content is associated with revisions of regulatory texts after the commenting period (Hypothesis 1). Additionally, existing studies show that the number of comments is important both in generall, but also in special circumstances. Hence, I hypothesize that more comments are associated with greater revisions of regulatory texts (Hypothesis 2). It was also found that the sentiments of comments are important and change depending on the situation. Although there is a lack of studies on the effect of comments' sentiments in the rulemaking sphere, marketing studies use customer sentiments to detect attitudes toward products.[41] I assume that higher positivity (positive sentiments) are associated with fewer revisions of regulatory texts (Hypothesis 3) as it probably corresponds to a better attitude towards them. Additionally, I expect that higher emotionality in comments correlates with negative expressions, leading to more revisions of texts (Hypothesis 4).

[41] Ana Catarina Forte, Pavel B. Brazdil, "Determining the Level of Clients' Dissatisfaction from their Commentaries", *PROPOR* (2016): 75.

Data

Regulations

The data of US federal regulations were extracted using the official Application Programming Interface (API) provided by the Open General Services Administration (OpenGSA).[42] This service requires the creation of access tokens, allowing for up to one thousand requests per hour.

For each regulation, I extracted the proposed rule and the final rule.[43] Proposed rules are the initial versions of a regulation. Therefore, if multiple versions exist, the earliest one is selected. The final rule should be the definitive version, and if there are multiple versions, the latest one is chosen. Occasionally, several proposed and final Rules can be published under one regulation, covering different aspects, and having distinctly different titles, necessitating the correct pairing of proposed and final rules. Additionally, there are sometimes final rules that contain only corrections to earlier versions. To accurately pair different versions of rules without manually searching each case, I linked them by the smallest string distance between their titles as the titles of proposed and final rules in a pair are usually very similar, and by their publication dates. For instance, when selecting a final rule for a proposed rule and faced with multiple options, I choose the one whose title most closely resembles that of the proposed rule, and if there are several, I select the most recently published.

I use in the research not the entire rules but only their specific parts. Both proposed and final rules typically contain an explanation of the regulation, its justification, and other components such as impact assessments. The most crucial section is usually found at the end of the rule: the list of subjects to the Code of Federal Regulations. As it was said in previous section, each US federal regulation amends this document by adding or changing provisions. Generally, most of published rules include a section with these provisions as there is requirement to include them into final rule but in proposed rules it is also often presented.[44] This part constitutes the actual changes made by the rule—these provisions regulate the activities of societal actors and are incorporated into the CFR, while other sections of the rule do not contain legal provisions and aim to justify and explain the implications of those

[42] "Regulations.gov API", U.S General Services Administration, accessed September 16, 2023, https://open.gsa.gov/api/regulationsgov/

[43] For clarity, in this text, the term "regulation" refers to an individual regulatory project. The term "rule" pertains to the various versions of that regulation, including proposed and final versions. Essentially, "rule" refers to the documents within the regulation.

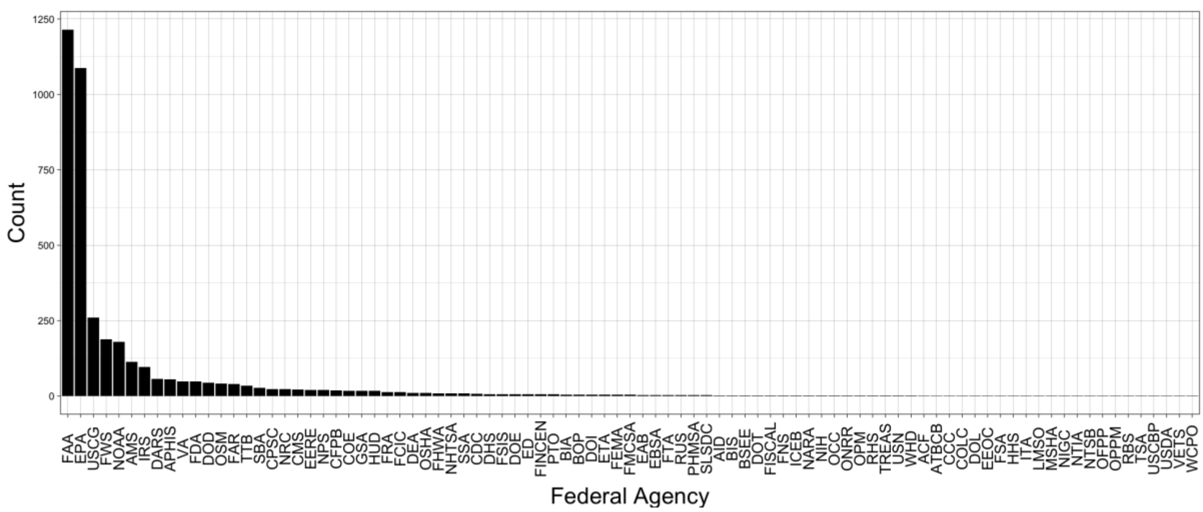
[44] Office of the Federal Register, *A Guide to the Rulemaking Process*, (Washington D.C: Office of the Federal Register, n.d.), 4.

provisions included in the CFR. This is why I focus my analysis solely on this section—other parts vary significantly between the proposed and final rules as documents are serving different purposes. However, the section containing the CFR provisions remains consistent, and any differences there between proposed and final rule are solely attributable to revisions and corrections.

I extract regulations for the period from 2010 to 2016. This interval was selected to capture the earlier stage of e-rulemaking development. There are 14895 federal regulations published for that period, but I extracted only 9124 regulations, or 18248 rules (twice as many, since there are pairs of rules for each regulation) as only regulations with clearly connected proposed and final rules were extracted. From the text of each rule, I retained only the section listing subjects related to the CFR, using regular expressions to isolate this part. Then, I removed all regulations where either rules lacked a list of subjects' section or where this section was structured differently from usual practice. This led to the dataset being halved, leaving 4442 regulations. Recognizing that such significant filtering based on a specific condition could introduce bias, I balanced the dataset to make the proportions between agencies more reflective of the initially extracted data, which also resulted in the removal of an additional 500 regulations.

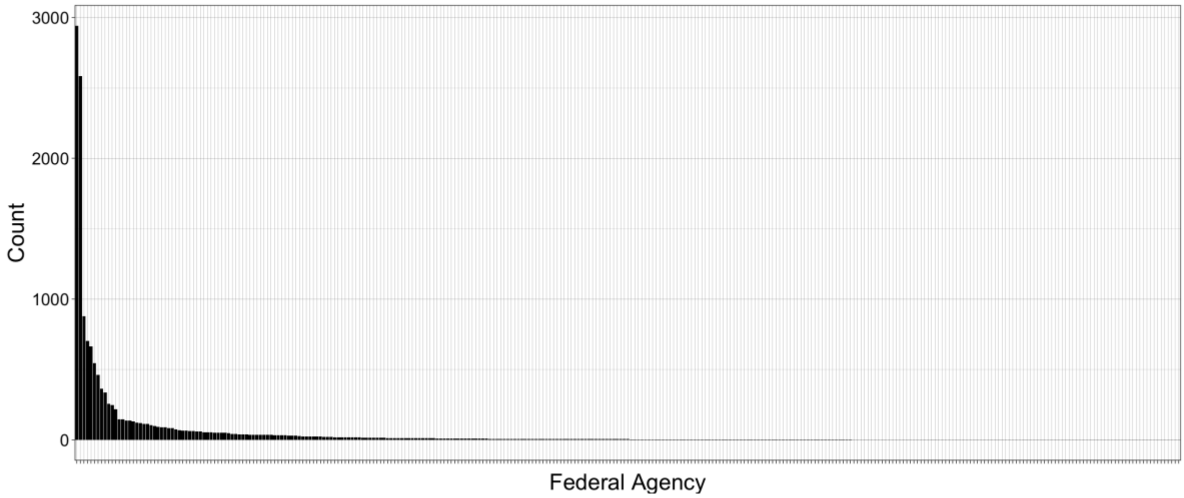
However, even after balancing the data, we observe significant skewness, with slightly more than half of the regulations in the dataset being developed by two agencies: the Federal Aviation Administration and the Environmental Protection Agency (Figure 2).

Figure 2. Distribution of rules by agency in the final dataset.



This is explained by two factors. Firstly, the distribution closely mirrors that of the original dataset, which comprises all the published in that period 14895 regulations (Figure 3). Here, 37% of the rules are developed by the same two agencies, indicating that the data inherently reflects the imbalanced regulatory activity of US federal agencies. Secondly, the filtering process slightly altered the distribution. As shown in the two figures, many agencies that produced relatively few rules were omitted from the final dataset. This pattern may reflect that agencies less active in regulatory policy tend not to write rules in a conventional manner or that the API fails to capture those rules for certain reasons.

Figure 3. Distribution of rules by agency for all regulations published from October 14, 2010 to April 5, 2016.

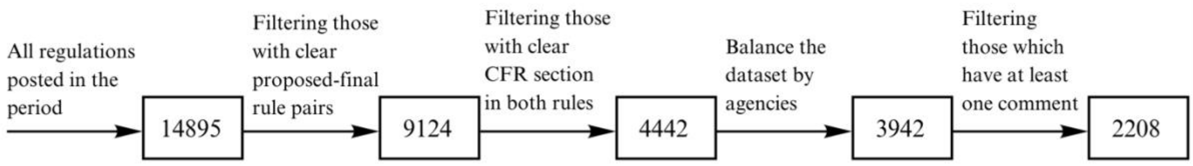


Comments

Extraction of comments was done for all 3942 regulations. These comments were posted in response to each proposed rule in the dataset. The range of comments varies significantly: 2208 regulations received at least one comment on their proposed rules, while 1417 regulations had five or fewer comments. The comments were also extracted using an API. There are two issues. Firstly, since this research focuses on the content of comments, there must be at least one comment published on the proposed rule. This requirement reduces the dataset to include only those regulations that have received at least one comment. Secondly, there are 75 regulations received more than 200 comments, with some garnering over 10000. Extracting all these comments would require considerable time and I decided to sample the comments: if more than 200 comments were posted on a regulation, I took a random sample of

200 comments from the whole population of comments so the maximum number of comments per regulation is capped at 200 in the dataset.

Figure 4. Filtering of the Dataset with numbers of Regulations.



Overall, I extracted 43757 comments on federal US regulations. These were accessed either in simple text form if the comments were submitted through the website's input form or as documents in various formats. For documents containing non-copyable text, I used Optical Character Recognition to read these texts. The comments were pre-processed by converting them to lowercase, removing non-alphabetical symbols, and performing both stopword removal and lemmatization.

Methodology

I utilize a similar approach to the method used of computational comparison of proposed and final rules for a given regulation.[45] Given that there are two different versions of the regulation and comments are posted on the proposed rule and the final rule is released only after the commenting period, it is feasible to explore which attributes and textual elements of comments correlate with greater or lesser differences between the first and final versions of regulations and is there generally such a relationship. The difference between the two versions of the texts serves as the variable to be explained, with the proposed rule acting as the initial version and the final rule as the final version. I compare these two and quantify their differences, designating this difference as the explanatory variable. Additionally, I have several variables that could explain this difference, making them response variables. Given the presence of one explanatory variable and several response ones, I employ a Multiple Regression model to measure the impact of various variables on the dependent one.

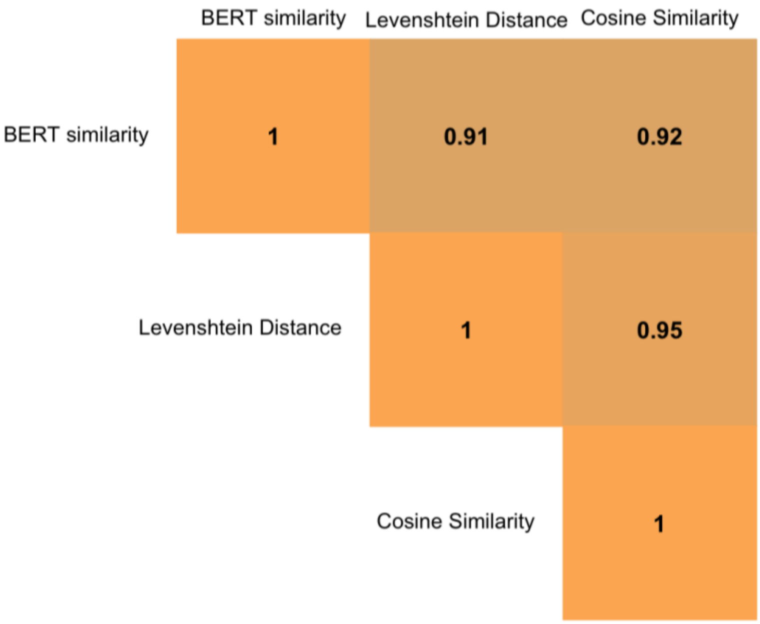
[45] Alex Ingrams, "Do public comments make a difference in open rulemaking? Insights from information management using machine learning and QCA analysis", Government Information Quarterly, Vol. 40 (2023): 5.

For computing text differences, I employ three different metrics. The first is the Levenshtein distance, which quantifies the dissimilarity between two texts by counting the minimum number of single-character edits required to transform one text into the other. The second metric is Cosine similarity, which calculates the cosine of the angle between two vectors projected in a multi-dimensional space. Texts need to be vectorized for this, and typically, the Term Frequency-Inverse Document Frequency method is used, which assigns scores to terms based not only on their occurrence in a particular document but also considering their frequency across all documents in a corpus, thus assigning lower scores to common terms, and focusing more on rare terms. However, to observe the actual changes between rules in each regulation separately, I use a basic term frequency calculation where words in documents are scored based on their occurrences. The third metric is BERT-based semantic similarity, which utilizes a pre-trained deep learning language model developed by Google researchers, with word embeddings as a fundamental part of its architecture.

All three metrics are normalized, ranging from 0 (indicating total dissimilarity) to 1 (indicating identical documents). For the Levenshtein distance, I do not preprocess the texts at all, as this metric is used to measure *superficial difference,* and it captures any literal difference between texts, including case sensitivity. For Cosine similarity, I undertake comprehensive preprocessing, removing all non-alphabetic symbols, including numbers and punctuation marks, and performing lowercasing, lemmatization, and stopword removal. This metric focuses on *textual difference*, detecting word changes and checks how the content of the documents differs. For BERT-based semantic similarity, I only perform basic preprocessing such as lowercasing and punctuation removal, as this model requires contextual information which can be lost after stopword removal and lemmatization. This metric is unique here in capturing not only the literal changes in the text but also the *semantic dissimilarities* between documents. The use of all these metrics is justified by the reason that they measure differently similar things and then can overlap. The presence of significant results across multiple metrics enhances the confidence in the findings. Additionally, since these metrics assess different dimensions of textual differences, they can provide deeper multifaceted insights. For instance, differences between texts can be superficial rather than textual, and vice versa.

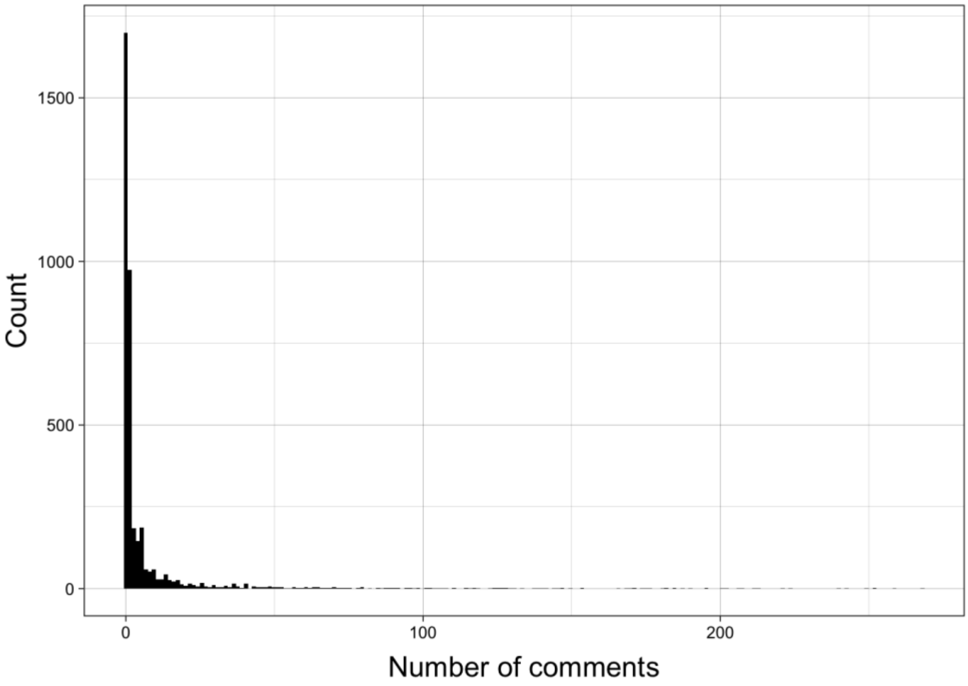
Surprisingly, all the metrics are highly correlated, even though each quantifies the differences between texts in different ways. It is likely explained by the overlap in what these metrics measure: if a text exhibits any superficial differences, probably it includes the change of content and then the meaning also varies, leading to a correlation between these differences. Additionally, it can be explained by the formal style of regulations: as they are written in a standardized style with structured formatting and lack stylistic diversity, the more literal differences between texts can be highly correlated with semantic differences.

Figure 5. Correlation matrix of Text Similarity metrics.



For the attributes of comments, I consider several different variables. The first one is the number of comments. As mentioned previously, various stakeholders can post comments on rules. I count only those comments that were posted on the proposed rule within the regulation, and only regulations with at least one comment are taken to analysis. The distribution of comments is highly skewed; most rules receive just a few comments, but there are regulations with thousands of comments, with a maximum of 20812 comments for one proposed rule. On the Figure 6, the highly skewed distribution is apparent, where extreme values were excluded from the figure to make it more readable.

Figure 6. Distribution of the number of comments posted on Proposed Rules (max. 300).



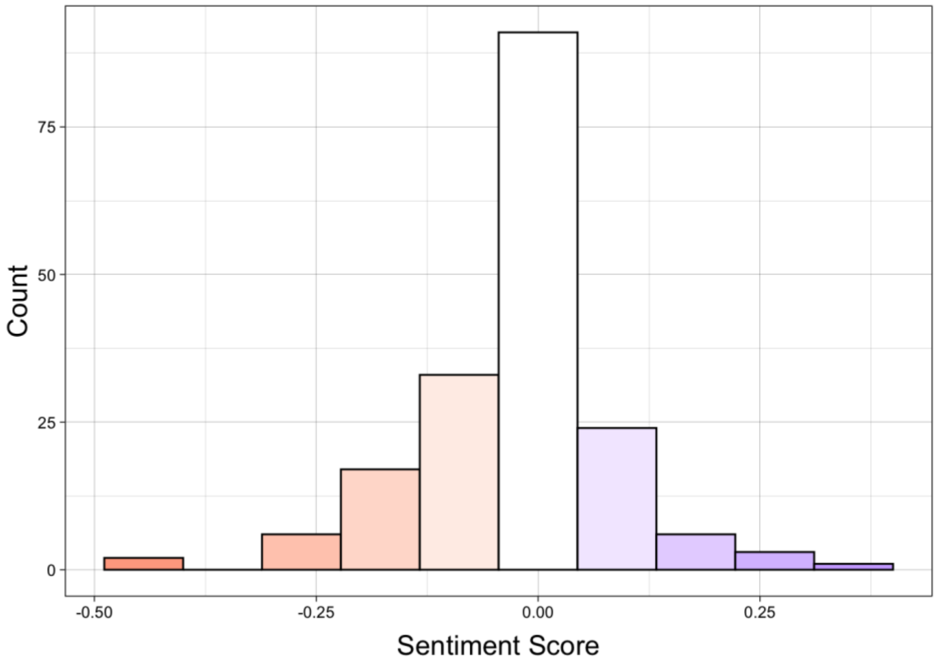
As mentioned earlier, I extracted the texts of comments with a cap of 200 comments for each regulation. This cap does not affect the variable representing the number of comments as I used the actual numbers of comments that were posted on the proposed rule. Given the presence of high outliers and significant skewness in the distribution, I have applied a logarithmic transformation to the variable to capture the relationship more effectively.

The subsequent variables are sentiment score and sentiment magnitude. Considering the limited efficacy of dictionary-based methods for sentiment analysis, I employed the Google Cloud Natural Language API. This service offers a pre-trained machine learning model that analyses text, assigning sentiment scores that take into account contextual information and the sequence of words, rather than evaluating each word separately.[46] Sentiment score ranges from -1 (entirely negative) to 1 (entirely positive), with 0 denoting a neutral sentiment. Since the analysis is conducted at the level of regulations rather than individual comments, it is necessary to average the scores from comments for each regulation. The distribution of sentiment scores for each regulation is likely Gaussian, as indicated in Figure 7 for one of regulations. This suggests that calculating the mean score of comments' sentiment scores per regulation is an

[46] "Analyzing Sentiment", Google Cloud, accessed December 10, 2023, https://cloud.google.com/natural-language/docs/analyzing-sentiment

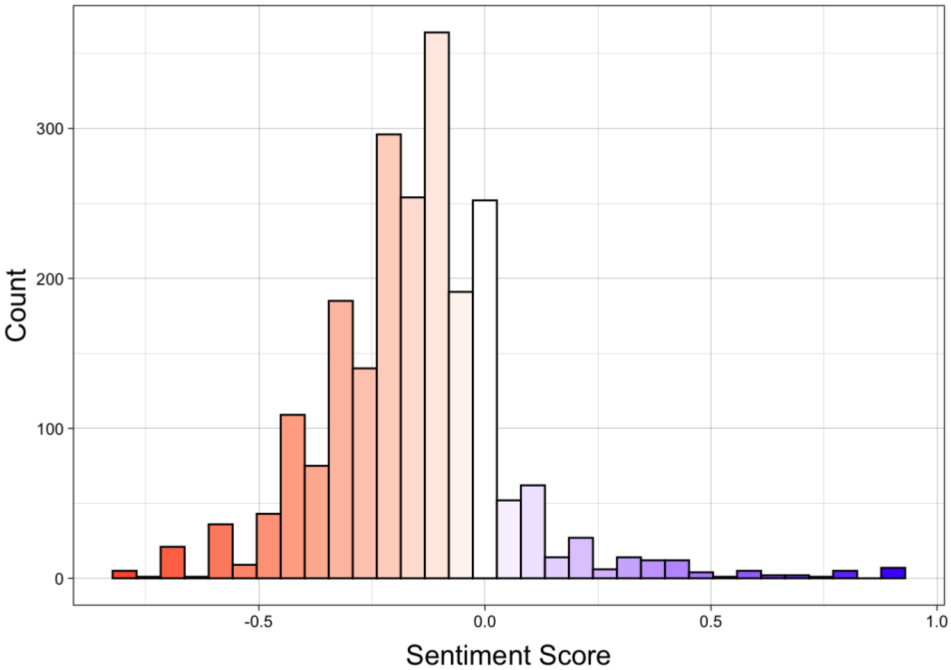
appropriate measure of central tendency, as it closely approximates both the median and the mode.

Figure 7. Distribution of Sentiment scores for comments posted on Proposed Rule BSEE-2015-0002-0001.



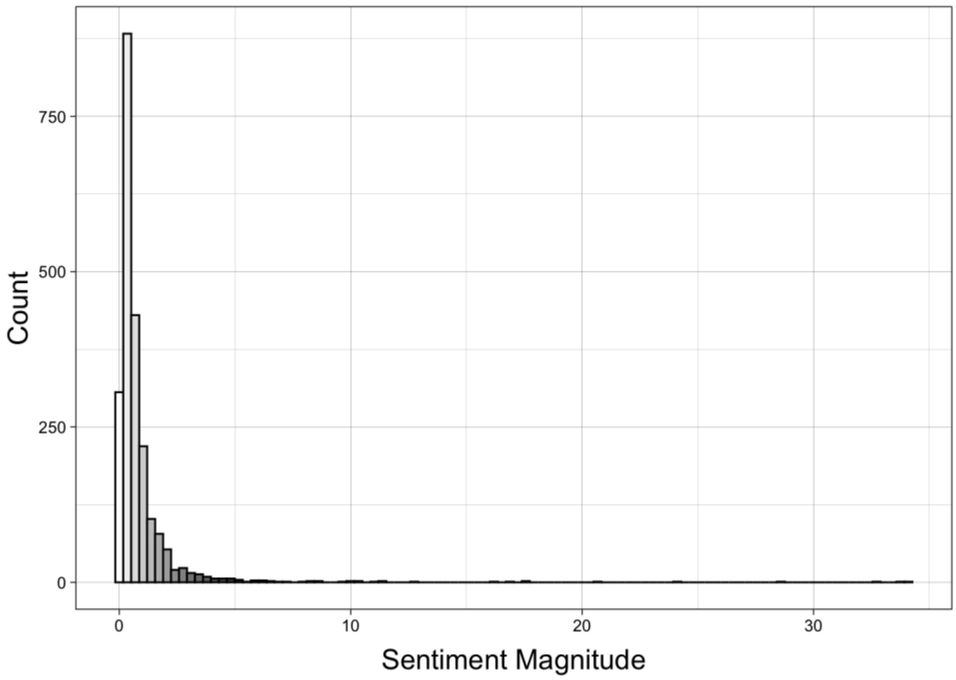
The distribution of the average sentiment scores across all regulations also approximates a normal Gaussian distribution. However, it exhibits a leftward skew, indicative of a strong tendency towards negative sentiments with the peak occurring slightly on the left (Figure 8). It reflects that stakeholders tend to use more negative language in their comments on most regulations.

Figure 8. Distribution of Sentiment scores averaged per regulation.



The sentiment magnitude is meant to capture the extent or intensity of emotional content within a text. While it is normalized on a per-sentence basis, it is not adjusted for entire texts, which can lead to potentially high values when texts exceed a single sentence. The distribution of these values is highly right skewed, with a majority clustering near zero, suggesting that most comments are devoid of strong emotion. However, a substantial proportion of the comments fall between the values of 2 and 4, indicating that many regulations collect relatively emotional comments.

Figure 9. Distribution of Sentiment magnitudes averaged per regulation.



Considering that longer comments are likely to have a higher score simply due to their length, I included comment length in the form of word count as a control variable in the regression analysis. This accounts for comment length to handle the non-normalized sentiment magnitude metric and may reveal whether comment length influences the difference between documents. Some studies have used comment length as a variable, showing that different actors tend to write comments of varying lengths on average, with industries providing the lengthiest comments and citizens the shortest. [47] However, the effect of comment length on regulations has not been studied. As each comment has its own length, I calculate word count for each of comments, and then average by regulation. Due to the high skewness of the distribution and the presence of extreme outliers, I have applied a logarithmic transformation to the variable before including it in the model.

The next variable is of primary interest for that research and its creation required several steps. Initially, I concatenate all the comments by regulation. Consequently, for each regulation, I have long texts consisting of all comments posted on that regulation. Then, I calculate term frequencies of each word in those comments per regulation and normalize them

[47] James Andrew Smith, Roxanna Abhari, Zain Hussain, Carl Heneghan, Gary Collins, Andrew Carr, “Industry ties and evidence in public comments on the FDA framework for modifications to artificial intelligence/machine learning-based medical devices: a cross sectional study”, *BMJ Open* (2020): 4; Simon Fink, Eva Ruffing, Tobias Burst, Sara Katharina Chinnow, “Emotional citizens, detached interest groups? The use of emotional language in public policy consultations”, *Policy Sciences*, Vol. 56 (2023): 489.

by the length of the text, resulting in scores for each word appearing in comments for a particular regulation. Importantly, the calculation of normalized term frequencies is done in isolation for each regulation. Therefore, the scores of a word in comments for one regulation do not affect the score for the same word in comments for another regulation.

Next, I take proposed-final rules pairs from each regulation and derive the words that appear only in the proposed rule but not in the final rule, and vice versa—the words that appear only in the final rule but do not appear in the proposed rule. Then, I concatenate these two lists of words, resulting in a single list consisting of words by which the two versions of regulations differ, and which appear or disappear in the final rule compared to the proposed rule for each regulation. Further, I use term *unique words* for referencing those words for convenience.

Subsequently, for each regulation, I have a set of normalized term frequencies for all words appearing in comments for that regulation and a set of words that appear only in one version of the regulation and by which the rules differ. Then, I sum normalized term frequencies only for those unique words. In mathematical notation, it would appear as follows:

$$\mathrm{S(r)} = \sum_{\mathrm{w} \in \mathrm{D_r}} \frac{\mathrm{T_w}}{\mathrm{L}}$$

Where T_w is the term frequency of word w in the comments, and L is the total number of words in comments. Hence, $\frac{\mathrm{T_w}}{\mathrm{L}}$ represents the normalized term frequency. D_r stands for the set of unique words which appear only in one version of regulation r. Then, the formula sums the normalized term frequencies for only words which appearing only in one version of regulation.

Essentially, I calculate the proportion of unique words in comments to gauge how frequently commenters mention those words that distinguish between two versions of regulations. A high proportion of these words in comments suggests that they are commonly used by commenters. Frequent mention of these unique words can indicate that commenters are concerned about topics related to those words, and that their feedback is significantly associated with the revisions in regulations. For instance, they might express dissatisfaction with the absence of specific words in the proposed rule, or conversely, request removal of some words.

In cases where a high relative frequency of unique words in public comments corresponds to significant differences between proposed and final rules, it suggests that public feedback is significantly associated with regulatory revisions. Specifically, this indicates that agencies are likely to make substantial revisions in response to issues that attract considerable attention from stakeholders, leading to significant alterations in the parts of the regulation that are most discussed. This shows that agencies prioritize modifications based on the intensity of public engagement on specific topics. Conversely, when a smaller share of unique words in comments corresponds to smaller differences between the texts, it may indicate that these unique words are not highlighted by the public and thus are probably not core themes of the regulation, leading only to minor corrections. Additionally, this scenario could imply that the public is focused on other topics, which were not accounted for by the agencies, but they make only minor adjustments to aspects that were not emphasized by the public. Overall, this can suggest that agencies tend to make larger changes in response to public feedback.

If relationship between share of unique words in comments and text similarities is reversed, and unique words are seldom used by commenters, but there is a significant difference in the texts, it might indicate that commenters focused on topics that were not modified in the final rule from the proposed rule, but factors other than public feedback is prompting agencies to significantly alter the documents. Moreover, when share of unique words is high but differences between texts are small, it means that agencies are taking public feedback into account to make only small corrections. Overall, this relationship indicates that larger revisions are made because of factors other than public feedback.

However, significant concerns arise. Firstly, if the list of unique words is quite long, then having more words there probably corresponds to a higher probability of those words appearing in comments, especially if those words are commonly used in the language. Moreover, if there are no differences between documents, meaning there are no unique words, then their share in comments will also be zero corresponding to no difference between rules. A high proportion of such regulations will inflate the magnitude of the relationship. Secondly, if these words literally constitute the difference between versions of regulations, a longer list likely corresponds to a greater difference between texts. I address these concerns with two adjustments. Firstly, it’s important to note that only 5 out of 2028 regulations have empty lists of unique words, while only 408 regulations have 10 or fewer unique words. Secondly, I create a custom list of stopwords for unique words, meaning that if texts differ by a word from this list, it will not be added to the list of unique words. I excluded all words that are merely artifacts of preprocessing and can represent acronyms or bureaucratic abbreviations of terms. Additionally, I removed words that do not relate to specific themes, and their alteration in the rules is probably not connected to substantial changes, such as "actually" or "regulation." The selection of these words is made manually by reviewing the list of words that most frequently appear in the lists of unique words for regulations, and the full list can be seen in Appendix 1. Thirdly, I include the variable of the length of the unique words list to control for the size of the list. Generally, the length of unique words lists can vary significantly for very close values of textual difference (as shown in Figure 10) and for identical sizes of unique words lists, differences between texts can vary greatly (as shown in Figure 11). This indicates that the length of the list of unique words is not the sole predictor of textual differences, and that there are other variables which explain the variation in differences, possibly including the share of unique words in comments. Nonetheless, accounting for this variable helps to differentiate the effects of these two possibly correlated variables.

Figure 10. Distribution of lengths of unique words lists for regulations where Cosine Similarity between Proposed and Final Rule is between 0.9 and 0.91.

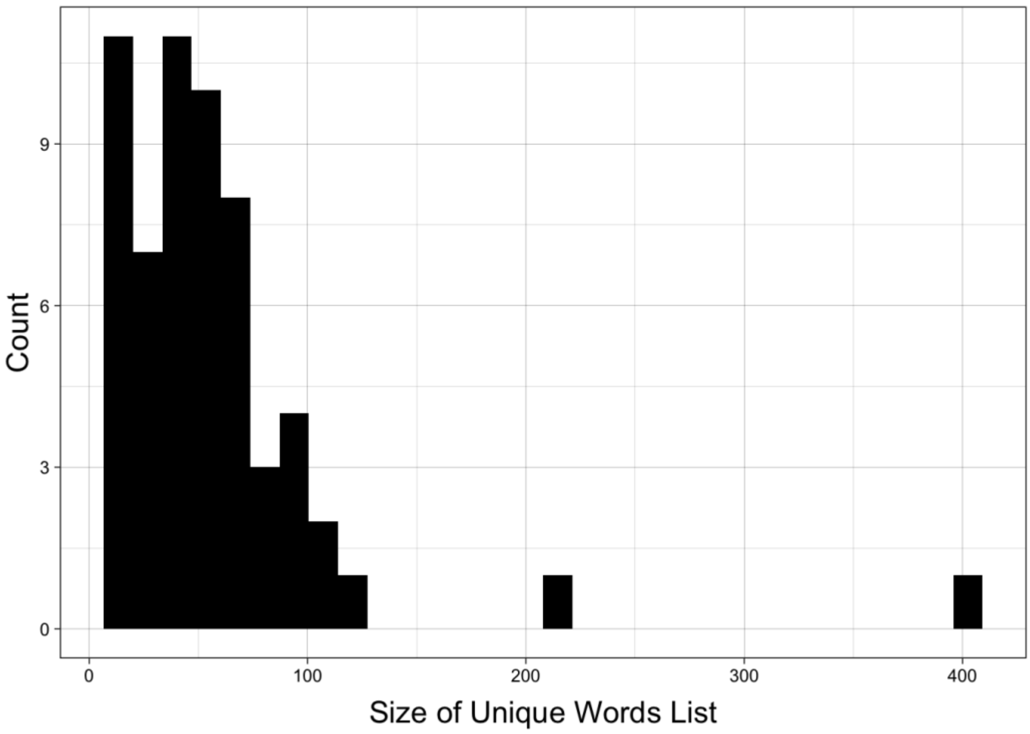
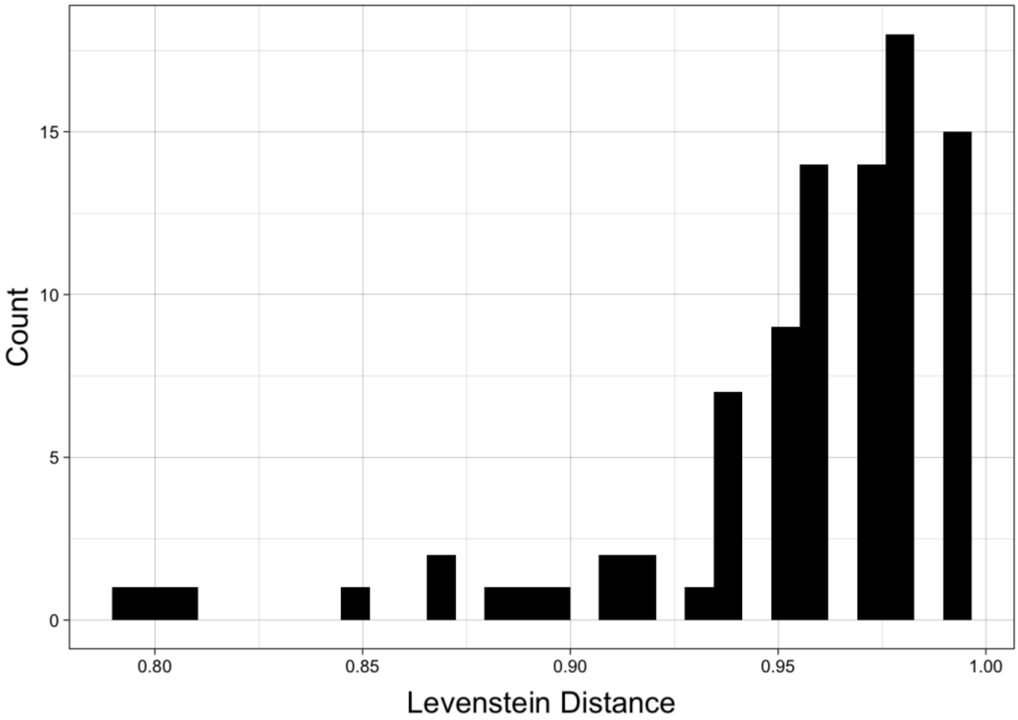


Figure 11. Distribution of Levenshtein Distance scores for unique words list with size 5.

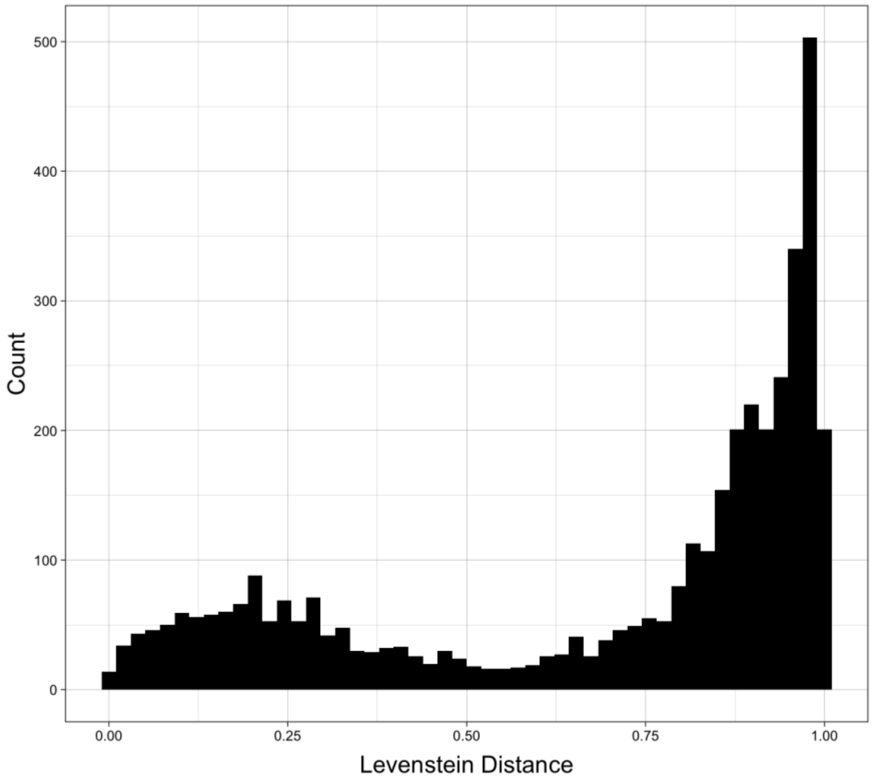


In the regression I use agency fixed effects. This means that a variable is created for each agency where all observations are set to zero, except for those pertaining to a specific agency. These variables, in the form of one-hot encoding, allow control for agency-specific effects that could be confounders affecting both independent and dependent variables. By controlling for these agency tendencies and characteristics, it is possible to achieve more robust and reliable results, thereby reducing the bias in the model.

However, there are other few issues with the regression model which should be discussed or addressed. The first issue is related to the high correlation between two variables: the share of unique words in comments and the length of the unique words list, which have a Pearson correlation coefficient of 0.57. As it was said, this correlation is quite expected because having more words in the list likely corresponds to a higher probability of those words appearing in comments, but including both variables is essential. Therefore, I will perform a Variance Inflation Factor (VIF) test to check for the presence of multicollinearity in the regression model. If significant multicollinearity is found between these variables, I consider residualizing those variables. This is done because if the variables are correlated and it inflates the coefficients of those variables, then predicting one variable by another and deriving residuals from both models allows us to include not the variables in their entirety, but their specific parts that do not depend on each other—essentially, the variance of one variable that is not explained by the other[48]. This approach slightly changes the interpretation of the model's results, and we obtain not the effect of the whole variable, but only its part.

The second issue pertains to the distribution of the explanatory variable. It is evident that all three metrics of the independent variable exhibit a bimodal distribution, with two peaks, where the right peak is significantly higher and populated than the left one (as shown in Figure 12). This observation suggests three things. Firstly, most regulations either have quite minor differences between their versions or quite substantial ones, with moderate differences being the least common. Secondly, among these two peaks, there is a greater number of regulations that are very similar than those with significant differences. Thirdly, when examining the distribution of all three metrics, the most pronounced bimodal trend and the greatest distance between the two peaks are observed in the Levenshtein distance.

Figure 12. The distribution of Levenshtein Distance between Proposed and Final Rules.

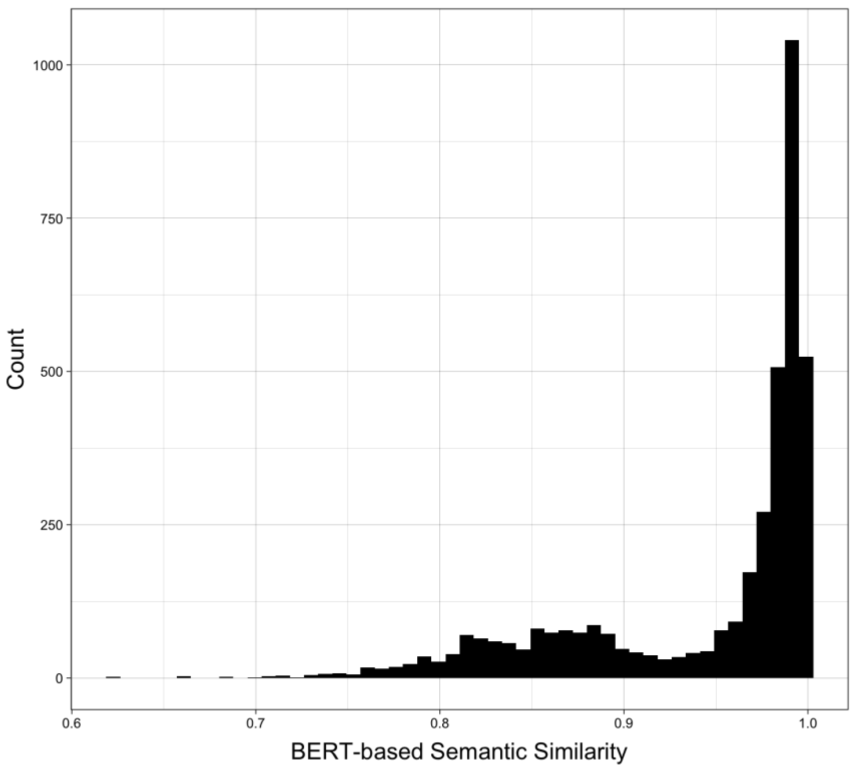


Upon examining the other metrics, the more I progress towards metrics that analyze the text more substantially, the narrower the distribution becomes, and the less pronounced is the

[48] Canh Phuc, Nguyen., Christophe Schinckus, and Thanh Dinh Su, "Determinants of Economic Complexity: A Global Evidence of Economic Integration, Institutions, and Internet Usage", *Journal of Knowledge Economy* 14 (2023): 4201.

left peak representing regulations with significantly different versions. This likely indicates that although the superficial differences between documents, such as the order of words, numbers, or punctuation (as indicated by the Levenshtein distance), may increase, the more substantial differences, such as those of semantics, do not shift as greatly, which is reflected in the BERT-based semantic similarity (as seen in Figure 13). This similarity in distributions can be explained by the overlap in what these metrics measure and factor of regulatory texts formality. If there is a superficial differences in the texts, they are likely to correspond to semantic changes. The Levenshtein distance, which captures any variation in symbols, tends to yield a higher score because it detects any differences, even change of case. However, in terms of semantic similarity, these differences are not significant unless the document has undergone a complete overhaul as the semantic does not differ much because of lexical diversity lack. The Cosine Similarity metric, which does not capture all literal differences yet is not a measure of semantic similarity, reflects dynamics that lie between these two extremes.

Figure 13. The distribution of BERT-based Semantic Similarity between Proposed and Final Rules.



This bimodal type of distribution suggests that it encompasses two underlying distributions, each with its own central tendency, and each subgroup can reflect different behaviors influencing the variable. Regressing on this entire variable can be problematic due to potential heteroscedasticity and the omission of distinct subgroup behaviors. For this reason, I split my metrics into two parts, each corresponding to one of the distributions. The cutoff for separation is set at the minimum bridge between the two peaks, where values are lowest, and this cutoff varies for each metric. Consequently, I will conduct regression analyses on the full dependent variable as well as two separate regressions for the divided parts.

Hence, the basic regression model here will have a following form:

$$\mathrm{TextDifference_i} = \beta_0 + \beta_1 \mathrm{SentimentScore_i} + \beta_2 \mathrm{SentimentMagnitude_i} + \beta_3\, log\left(\mathrm{UniqueWordsNumber_i}\right) + \beta_4 U\mathrm{niqueWordsShar}e_i + \beta_5 \log(\mathrm{CommentsNumber_i}) + \beta_6 \log\left(\mathrm{CommentsLength_i}\right) + \gamma_{\mathrm{Agency_i}} + \epsilon_\mathrm{i}$$

For regulation i, the variables include the average sentiment score and magnitude, the proportion of unique words in comments, and a control for the length of the unique word list. Additionally, the model incorporates the logarithms of the number of comments and the average word count representing comment length. There is also an agency fixed effect and an error term. The dependent variable, which is a text difference metric, changes for each model. Hence, when dividing the dependent variable into two parts for regression, the right side of the equation remains constant, but the data is differing for the subgroups of the dependent variable.

The current regression model primarily captures linear relationships. While the logarithmic transformation of two variables may better show non-linearities in their relationship with text differences, the remaining variables are only indicative of linear associations. Nevertheless, non-linear interactions might exist, and more advanced methods may be required to detect them. To this end, I employ a machine learning approach, specifically a gradient boosting model, which is adept at uncovering non-linear relationships between the target variable and features. Gradient boosting is an ensemble method that builds decision trees in sequence, with each tree learning from the errors of its predecessor, thereby potentially enhancing results.

To compare the R^2 scores of linear and non-linear models fairly, I train both models on the same train-test splits and evaluate their R^2 on the test data. This ensures their comparability as R^2 reflects the models' performance on unseen data, rather than using the original full data for the linear model and the test data for the non-linear model, which would be unfair. Fixed effects are removed because the linear regression performs poorly with many hot-encoded variables, while the gradient boosting model benefits from them. If the non-linear model significantly outperforms the linear model, it indicates the presence of non-linear patterns that the linear model cannot capture effectively.[49] If this is the case, I can further investigate the true form of the relationship between text differences and each variable using Partial Dependence Plots. These plots illustrate the relationship between a single feature and the target variable, averaging out the effects of all other features.

An additional benefit of the gradient boosting model is its feature importance tool, which indicates which variables or features the model most relied upon for prediction and hence, which contributed most to the resulting R^2. Therefore, beyond statistical significance, which only indicates the presence of a relationship, and coefficients, which describe the nature and strength of the relationship, this metric is particularly useful in models where multiple variables are considered, as it helps to identify which factors are most influential in predicting the outcome.

Results

Initially I run three models with full data not dividing it on two sets. VIF test for those models show that there is no variable in any model with score higher than 2.5 what shows that there is no multicollinearity and no need to residualize variables. Models with all three metrics show quite optimistic results, as all variables are significant in almost all models. We see that for all models, the directions of relationships are consistent among different metrics, and the effect of the share of unique words in comments is the largest across all models. The control variable, number of unique words, has a significant relationship with text differences, but its coefficients are the smallest among all the variables. In the third model, it is not significant, and in the second model, it is significant only at the alpha = 0.01 level, while all others are significant at the alpha = 0.001 level. We see that all coefficients for the third model are smaller, which does not suggest that the impact is lower on semantic similarity, but rather that the distribution is much narrower, leading to a smaller variation. The number of comments, word count in comments, and level of comment positivity show small but significant positive effects, whereas the presence of emotional text shows negative effects. Those are already intriguing

[49] Dehua Liang, David A. Frederick, Elia E. Lledo, Natalia Rosenfield, Vincent Berardi, Erik Linstead, Uri Maoz, "Examining the utility of nonlinear machine learning approaches versus linear regression for predicting body image outcomes: The U.S. Body Project I", Body Image, Vol. 41 (2022): 34.

and important results, but I want to move to other models where datasets are divided based on two peaks of dependent variables, as these three basic models can miss important tendencies within the two distributions.

Table 1. Output of Regression Models with Full Dataset.

	Dependent variable:		
	Levenshtein distance	Cosine similarity	BERT-based semantic similarity
	(1)	(2)	(3)
Share of unique words in comments	-2.077***	-1.931***	-0.391***
	(0.074)	(0.065)	(0.017)
Number of unique words	-0.0003***	-0.0001**	0.00001
	(0.00004)	(0.00003)	(0.00001)
Number of comments (log)	0.010***	0.013***	0.002***
	(0.002)	(0.002)	(0.0004)
Number of words in comments (log)	0.013***	0.017***	0.003***
	(0.002)	(0.002)	(0.001)
Sentiment score	0.045***	0.055***	0.014***
	(0.017)	(0.015)	(0.004)
Sentiment magnitude	-0.011***	-0.006***	-0.002***
	(0.002)	(0.002)	(0.0004)
Constant	0.614***	0.925***	0.943***
	(0.151)	(0.132)	(0.035)
Agency fixed effects	Yes	Yes	Yes
Observations	2,208	2,208	2,208
R^2	0.747	0.728	0.677
Adjusted R^2	0.736	0.717	0.664
Residual Std. Error (df = 2119)	0.150	0.131	0.035
F Statistic (df = 88; 2119)	70.915***	64.390***	50.562***
Note:			*p<0.1; **p<0.05; ***p<0.01

On the Table 2 there are the models where I divided the dependent variables into two distributions, and to divide them I found the lower middle point between two distributions for each metric: 0.56 for Levenshtein distance, 0.71 for Cosine similarity, and 0.91 for BERT-based semantic similarity. Hence, each model is run on data group of more similar (the first column for each metric) and on group of less similar (the second column for each metric) documents. Further in the text I will call those models as *higher similarity group* and *lower similarity group* for convenience. The number of observations in each model show those two distributions in each of metrics are probably highly overlapped as numbers of observations

Table 2. Output of Regression Models with Divided Dataset.

	Dependent variable:					
	Levenshtein distance		Cosine similarity		BERT-based semantic similarity	
	(1)	(2)	(3)	(4)	(5)	(6)
Share of unique words in comments	-1.600*** (0.086)	-0.603*** (0.083)	-0.962*** (0.053)	-0.561*** (0.100)	-0.206*** (0.014)	-0.055* (0.033)
Number of unique words	-0.001*** (0.00005)	-0.0001*** (0.00004)	-0.0001*** (0.00002)	-0.0002*** (0.00005)	-0.00002*** (0.00001)	-0.00001 (0.00002)
Number of comments (log)	-0.004*** (0.001)	0.004** (0.002)	0.0001 (0.001)	0.003 (0.002)	-0.001*** (0.0003)	-0.001 (0.001)
Number of words in comments (log)	-0.004** (0.001)	0.007** (0.003)	0.001 (0.001)	0.006 (0.004)	0.001** (0.0003)	-0.001 (0.001)
Sentiment score	-0.010 (0.010)	-0.006 (0.025)	-0.0001 (0.006)	0.036 (0.028)	0.003 (0.002)	0.009 (0.009)
Sentiment magnitude	0.00001 (0.002)	-0.006*** (0.002)	-0.001 (0.001)	-0.003 (0.002)	-0.0002 (0.0003)	-0.0002 (0.001)
Constant	0.946*** (0.072)	0.544*** (0.124)	1.023*** (0.046)	0.708*** (0.084)	0.969*** (0.015)	0.898*** (0.034)
Agency fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,663	545	1,698	510	1,750	458
R^2	0.537	0.393	0.354	0.291	0.363	0.151
Adjusted R^2	0.513	0.339	0.320	0.240	0.330	0.083
Residual Std. Error	0.072 (df = 1580)	0.121 (df = 500)	0.045 (df = 1612)	0.137 (df = 475)	0.015 (df = 1664)	0.045 (df = 423)
F Statistic	22.393*** (df = 82; 1580)	7.347*** (df = 44; 500)	10.384*** (df = 85; 1612)	5.740*** (df = 34; 475)	11.133*** (df = 85; 1664)	2.209*** (df = 34; 423)

Note: *p<0.1; **p<0.05; ***p<0.01

are very similar for different metrics for same groups, what is also explained by very high correlations between those metrics. VIF test shows almost the same results as in the first three models, so there is no problem of multicollinearity.

In the new models with divided data, it is visible that most variables become much less significant. For example, only two variables are significant for both groups in the Cosine similarity metric. Coefficients and their significance can vary depending on the text difference metric what suggest that those metrics with their similarities do not overlap completely. We see that sign of coefficients vary for different distributions within the same metrics, which confirms concerns about overlooked trends inherent to distributions within metrics. Moreover, the effect of a variable can be significant for one group but insignificant for another within the same metric. For example, a higher number of comments significantly leads to less semantic similarity in the higher similarity group, but it is not significant for the lower similarity group. This indicates that more comments under the proposed rule cause semantic changes in the final rule, but only for regulations where the versions are quite similar. Interestingly, for superficial changes in regulations, a higher number of comments has an oppositely different effect depending on the group of regulations: higher similarity or lower. If a regulation's versions are similar, a higher number of comments corresponds to more superficial changes in the regulation, but if the versions are relatively different, then more comments lead to less revision of the final rule.

The variable of comment length also shows a complex effect: it has significant coefficients in the same models as the number of comments. For the Levenshtein distance metric, it has the same directions of effect as the number of comments for the same groups. For the higher similarity group, it shows that lengthier comments lead to more semantic similarity between versions of the regulation.

The effect of these two variables is hard to interpret as it is likely multifaceted. They have the strongest influence on superficial differences between versions of regulations. Therefore, more comments or their length probably affect changes in punctuation, numbers, and the order of words. However, with more substantial kinds of changes, both variables only affect semantic similarity in the higher similarity group. Moreover, it is important to note that for both these variables, even the significant effects are very small. As these are taken as logarithms in the regression, for example, the largest coefficient for length of comments (0.007 for the lower similarity group with Levenshtein distance) means that increasing the length of comments by 1% leads to a 0.00007 increase in similarity, which is extremely small. The same applies to the number of comments: for the same metric and group, an increase in the number of comments by 1% corresponds to an increase in similarity by 0.00004. Hence, even though the effect of these variables is significant and important to note, it is minuscule and produces little impact. Then, *I confirm the Hypothesis 2*, but I emphasize that the effect is very complex and present only in specific circumstances.

Regarding the emotional aspect of the text, the effect is almost completely absent. Sentiment scores show no effect in any model, in contrast to the initial full models. This suggests that in models with divided data, the positivity of comments does not influence the decisions of agencies to revise rules. Regarding sentiment magnitude, it has a significant effect in only one model, specifically for the lower similarity group, suggesting that more emotional comments lead to more superficial revisions in texts, which do not correspond to any more substantial kinds of changes.

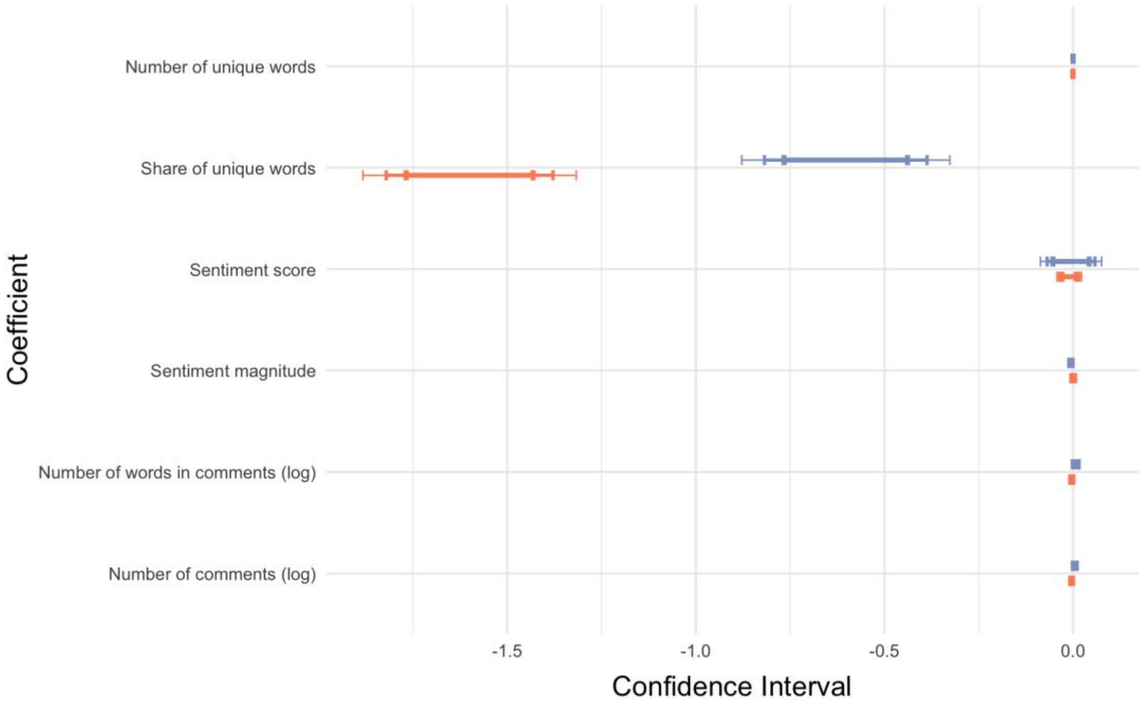
Considering that all those variables have significant effects in the initial three models with full data, the almost complete absence of effects of those variables in models with divided data can be explained by the fact that the divided models have less data. This less variation in the data can result in reduced coefficients, and fewer observations in higher standard errors. Both factors contribute to smaller coefficients and their reduced significance, what says that even with low statistical significance in models with divided data those variables can have an effect. However, we see that they already have tiny coefficients in the models with full data, and if the reduction of data results in the disappearance of significance for coefficients, it likely means that even though these relationships exist, they are not strong and do not have a great impact. Considering this, I state that having those results *it is not possible to conclusively confirm or reject Hypotheses 3 and 4*: results in the models with full data contradict the results of models with divided data, and it is difficult to derive conclusions with such inconsistent results.

The biggest effect in models with divided data is the same as in the models with full data, so of the variable of the share of unique words in comments. It has the always significant and the largest coefficient in all models.

Generally, considering that this variable can range from 0 to 1 (as it is the share of specific words in comments) and the dependent variables are also measured on a scale from 0 to 1 (indicating completely similar or dissimilar texts), coefficients close to or greater than 1 are not practical: an increase in the share by 1 in the higher similarity group in Levenshtein distance leads to a -1.6 reduction in similarity, which is impossible in a real scenario. Dividing it by 100 and presenting it as percentages shows that an increase of the share of unique words in comments by 0.01 leads to a decrease in similarity between versions of the regulation by 0.016

what shows the large impact. The difference in effects is visible in Figure 14, where the effect of the share of unique words is much larger than for other variables.

Figure 14. Confidence Intervals of Coefficients at 95%, 99%, and 99.9% Levels for Higher Similarity (Blue) and Lower Similarity (Red) Groups with Levenshtein Distance metric.



As mentioned, this relationship is at high risk of being explained by the sheer number of unique words, which can lead to higher differences in regulatory texts as those words embody those differences. The models show that the number of unique words has the tiny coefficient in almost all the models, which is also visible in the figure. However, this does not imply a small impact. The small coefficient is due to the different scales of variable measurement. For the number of unique words, the coefficient indicates the increase in text difference for each additional unique word. Given that these lists can be quite lengthy, consisting of dozens of words, the coefficient becomes not so small and even substantial. This suggests that both the number of unique words and their proportion in the comments impact the text difference.

With the aim to even more exclude the possibility of a misinterpretation perceiving the share of unique words more influential variable than it is, I conducted a robustness check: I ran the three regression models with full data, as the share of unique words in comments consistently had significant coefficients among the six models with divided data, but instead of using the share of unique words in comments, I use averages of their normalized term frequencies. This means that I did not sum the term frequencies of the words by which the versions of regulations differ but took their mean instead. This variable is harder to interpret, but generally, it also shows how often those unique words appear in comments. The main point is that this variable is not related to the number of unique words, as averages of term frequencies of those words cannot be related to their number, which is reflected in the 0.009 correlation coefficient between them. If this variable has significant and negative coefficients, similar to the variable of the share of unique words, then we can with greater confidence conclude that the interpretation is correct.

Models show the same results as the original ones, so the three models show significant negative relationship. Coefficients of this variable are much bigger than for share of unique words what is explained by the fact that means of term frequencies are much smaller than sums and increase of variable by one expectedly lead to much higher change in dependent variable. Thus, *with high confidence I confirm the Hypothesis 1.*

R^2 are quite high for the first three full models. Considering that original data has bimodal distribution in all three metrics and that coefficients differs greatly between data from two distributions, it is expected that linear regression captures the effect not fully. However, even in six models with divided data I expect the non-linear relationship of dependent variable with independent ones. I used gradient boosting models (tuned separately for each model) using the same set of variables to check whether there are significant non-linearities in effect of variables. I detect it using R^2 scores: much higher R^2 of a gradient boosting model than of linear model will evidence for strong non-linear relationships. In comparison from linear regression side, I use normal R^2 score instead of adjusted R^2 as in the non-linear model R^2 doesn't proceed any transformations and adjustments so comparison will be fairer.

On the Table 3 it is visible that non-linear model shows greater fit, showing at least the same, but almost always higher R^2 scores. It means that some variables have non-linear relationships with text differences metrics.

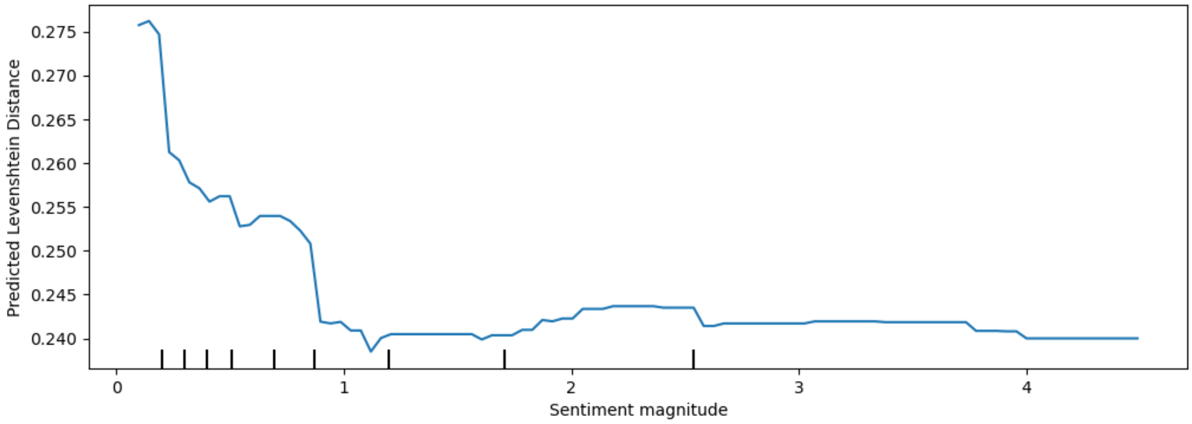
Table 3. R^2 scores for Multiple Linear Regression and for Gradient Boosting model.

	R^2 of Linear model	R^2 of Gradient boosting model
Levenshtein distance: Higher similarity group	0.467	0.54
Levenshtein distance: Lower similarity group	0.15	0.276

Cosine similarity: Higher similarity group	0.242	0.256
Cosine similarity: Lower similarity group	0.098	0.24
BERT-based semantic similarity: Higher similarity group	0.268	0.321
BERT-based semantic similarity: Lower similarity group	0.009	0.203

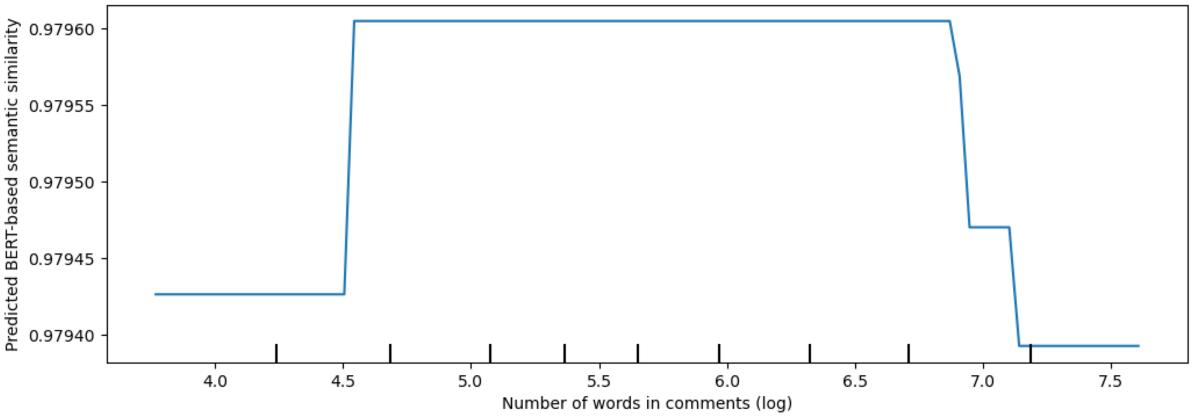
I use Partial Dependency Plots to detect which variables have a non-linear relationship with text difference metrics. Checking variables such as positivity or sentiment magnitude indeed shows non-linearities in relationships for all similarity metrics and similarity groups. This is important considering that I use divided data, so it is not explained by the presence of different peaks in the distribution of the dependent variable, but rather inherent in the relationships.

Figure 15. PDP of Sentiment magnitude feature and its prediction of Levenshtein distance in Lower Similarity group.



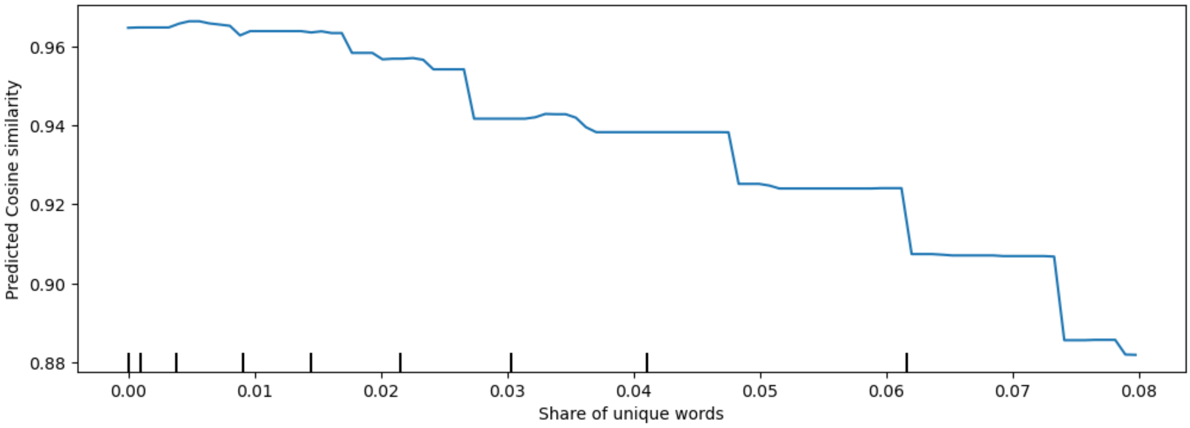
However, as shown on the Y-axis of the plots, only a tiny portion of text difference metrics is explained using those variables, even for Levenshtein distance, which has considerable variation. This suggests that these variables have very limited predictive power. The same situation applies to the number of comments and the length of comments—there are non-linearities in the relationships, but these variables predict an extremely narrow portion of the dependent variable.

Figure 16. PDP of Comments length feature and its prediction of BERT-based semantic similarity in Higher Similarity group.



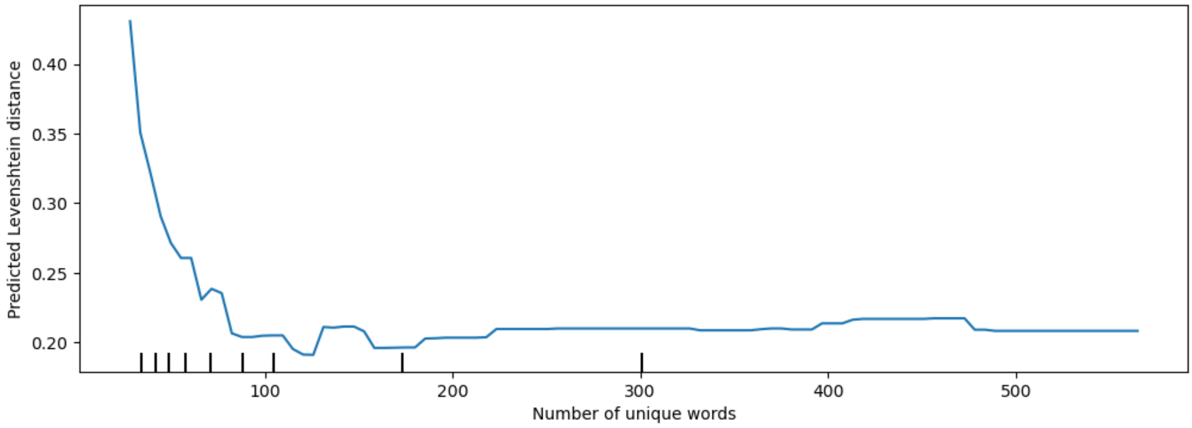
The share of unique words, on the other hand, shows a relatively linear relationship across all metrics. It exhibits a stepwise pattern, but the "steps" are small, and the direction of the curve remains consistent throughout the graph with minor deviations. Additionally, this variable predicts a significant portion of the dependent variable, as shown in the figure, considering that this group consists only of values higher than 0.71.

Figure 17. PDP of Share of unique words feature and its prediction of Cosine similarity in Higher Similarity group.



However, the variable representing the sole number of unique words has strongly non-linear relationships with almost all metrics and in nearly all groups. As shown in Figure 18 the curve drops rapidly and then levels off with almost no slope. Despite this, it also predicts significant portions of the dependent variables. This variable likely contributes the most to the better fit of non-linear models.

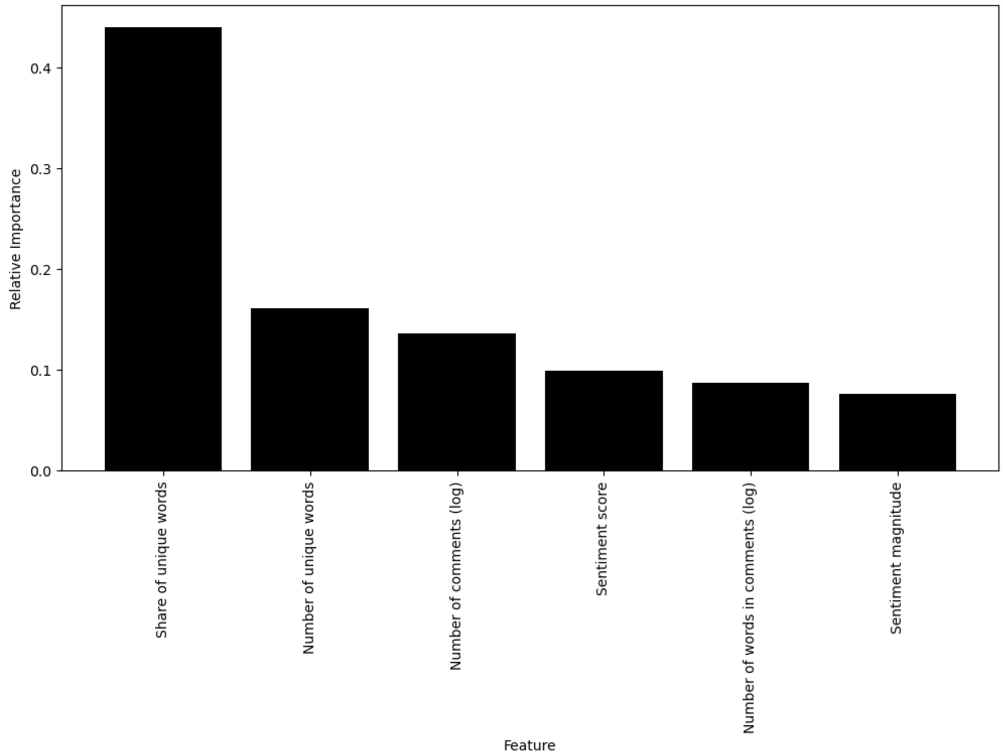
Figure 18. PDP of Number of unique words feature and its prediction of Levenshtein distance in Higher Similarity group.



Inspection of feature importance shows the expected results. In models where the number of unique words exhibits a very visible non-linear pattern of relationship with text difference metrics, the importance of this feature is the highest, significantly surpassing other features. However, in models where the relationship is more linear, this feature becomes less important, and the share of unique words takes the first place, as shown in Figure 19. In this case, the relationship between the number of unique words and text differences has a more linear pattern, making this feature less important in the model's predictions.

When I add and exclude variables in Multiple Linear Regression models, the addition of the number of unique words variable expectedly does not increase the R^2 more than the share of unique words variable. Other variables, apart from these two, show very limited predictive power in each model, which aligns with their small coefficients and infrequent statistical significance.

Figure 19. Feature importance graph with Cosine similarity metric for Higher Similarity group.



All of this suggests that the number of unique words has a complex and non-linear impact on the differences between two versions of regulations. In cases where the relationship is highly non-linear, the model relies more on this feature, resulting in higher importance. When the relationship is more linear, the variable's importance decreases because prediction can be more easily approximated by other features, primarily the share of unique words. The smaller explanatory power of the number of unique words compared to the share of unique words in all linear models also suggests that linear models capture linear patterns better, contrasting with the more complex patterns of the number of unique words.

Importantly, the share of unique words in comments shows a high explanatory power for text difference metrics in linear regressions and is consistently the first or second most important feature in non-linear models. It ranks second only when the number of unique words has a non-linear relationship with the dependent variable. This indicates a strong linear relationship between the share of unique words in comments and the difference between the first and final versions of the regulations.

Discussion

By employing the metric of the share of unique words—representing the proportion of words by which versions of regulation differ within the comments—I aimed to estimate if commenters use of words that subsequently appear or disappear in the final version of the regulation is correlated with text differences. The regression results show a strong negative relationship. This effect is significant, has a large coefficient, and contributes substantially to explaining variations in text similarity within the model. The effect is mostly linear and suggests that if commenters use words that then appear or disappear in the final version of the regulation compared to the proposed rule, the difference between versions increases. This can mean that words belonging to topics of public concern are then changed in the regulation, resulting in many changes on superficial, textual, and semantic levels. A decently high R^2 value, along with the high importance of this feature, indicates that it explains the variance in text similarities well. Several measures used in the research to separate the effect from the mere number of unique words increase confidence in the result. The high linearity of the effect suggests that the strong impact is consistent across different parts of the text similarities distribution, and not limited to specific sections.

The results indicate that when commenters frequently use words by which versions of regulations differ, the versions of the rule become less similar. When the public less frequently uses words that differ between versions of the regulations, the rules tend to become more similar. This could indicate that agencies either disregard topics not emphasized by the public, making fewer revisions to them, or that commenters highlight other parts of the rule for revision, but agencies do not take this feedback into account and not revise them. In any case, larger revisions are not associated with these situations. Therefore, the results allow for the possibility that agencies may make decisions disregarding public input in some situations. However, *the key finding is that larger revisions of regulatory texts are strongly associated with greater stakeholders' attention to the subsequently revised parts*. I establish a correlation, not a causal link, between the variables. Given that this method is being applied for the first time, additional research using various approaches is necessary to confidently determine the causal impact of comments on the development of regulations using big data.

The number of comments has a complex relationship with differences between proposed and final rule. It varies depending on the circumstances but remains consistently minor overall. This likely supports previous findings in the field regarding the multifaceted nature of the number of comments, indicating that this association is significant in certain circumstances but not apparent in others. Similarly, the length of comments yields inconsistent results, and even when significant, the effect size is extremely small.

The emotional aspect of comments is questionable . Positivity has a significant relationship when the full data used, but this effect disappears and becomes insignificant when data is divided into subsamples. The sentiment magnitude shows consistent negative effects across models, and in one case, it is significant. However, even when significant, the effect size remains very small. Considering that, it is hard to state that relationship is present.

Conclusion

This research aimed to find strong and generalizable evidence of the association between the regulatory texts and public comments, posted on them, in the United States. The results provide evidence that significant revisions of regulatory texts are correlated with the public's attention to those words being revised, as expressed through public commenting. However, they also indicate that specific aspects, such as the emotionality or number of comments, likely have an undefined or complex relationship with revisions of regulations, that is present in some circumstances but not in others, with the nature of the effect varying accordingly. This suggests the need for using different, narrower samples of data to identify these circumstances. Consequently, the existing trend in the field should continue to focus on specific domains and circumstances.

This research and its results pave the way for further investigation into the interconnection between rulemaking and public commenting. As the association between comments and regulatory texts has been established, it is not sufficient to assert a causal relationship. Results from different approaches or using other analytical tools should be obtained to confidently claim causality. Also, it is important to determine whether comments with a high use of words that differentiate various versions of the regulation are authored by specific groups. It is quite possible that those important comments are written by specific types of actors, probably organizations. Calculation of the share of these key words in each individual comment, allows us to identify which comments have a higher share and who the authors are.

Moreover, existing limitations of the research can be addressed in the future to make results more robust. Regarding the metric of semantic similarity of texts, the "Bag of words" approach was used which disregard the order of words. It significantly reduces the information about the language used and decrease accuracy. In the future research other approach can be used, which estimate the semantic similarities of sentences with their order instead of just set of words. Also, this research covers scenarios where commenters use specific words in their comments, and these words either appear or disappear in the final version of the regulation compared to the first version. However, it is possible that commenters used synonyms to express their attention to words, not writing exactly them. Approaches that account for this synonym usage can be used in a future research.

Bibliography

Anthony, Robert A. “Interpretive Rules, Policy Statements, Guidances, Manuals, And The Like-Should Federal Agencies Use Them To Bind The Public”. *Duke Law Journal*, Vol. 41 (1992): 1311-1384.

Anttiroiko, Ari-Veikko., Lau, Gloria T., Law, Kincho H. “A Prototype Study on Electronic Rulemaking”. (2008): 1-17.

Balla, Steven J., Dooling, Bridget C.E., Bull, Reeve., Hammond, Emily., Herz, Michael., Livermore, Michael., Noveck, Beth Simone. “Responding to Mass, Computer-Generated, and Malattributed Comments”. *Administrative Law Review*, Vol. 74 (2022): 95-160.

Canh Phuc, Nguyen., Schinckus, Christophe., and Su, Thanh Dinh. "Determinants of Economic Complexity: A Global Evidence of Economic Integration, Institutions, and Internet Usage". *Journal of Knowledge Economy*, Vol. 14 (2023): 4195-4215.

Cornell University. “Regulation Room”. Accessed April 15, 2024. http://archive.regulationroom.org/learn-more/stages-of-the-rulemaking-process/index.html

Costa, Mia., Desmarais, Bruce A., and Hird, John A. “Public Comments’ Influence on Science Use in U.S. Rulemaking: The Case of EPA’s National Emission Standards”. *American Review of Public Administration*, Vol. 49 (2019): 36-50.

Deighton-Smith, Rex. “Regulatory transparency in OECD countries: Overview, trends and challenges”. *Australian Journal of Public Administration*, Vol. 63 (2004): 66-73.

Dudley, Susan E. “Milestones in the Evolution of the Administrative State”. *Daedalus*, Vol. 150, 3 (2021): 33-48.

Dwidar, Maraam A. “Diverse Lobbying Coalitions and Influence in Notice-and-Comment Rulemaking”. *Policy Studies Journal*, Vol. 50, No. 1 (2022): 199-240.

Eisner, Marc Allen., Worsham, Jeffrey., and Ringquist, Evan J. *Contemporary Regulatory Policy*. London: Lynne Rienner Publishers, 2000.

EUR-lex. “European Union regulations”. Published March 16, 2022. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=LEGISSUM:l14522

Fink, Simon., Ruffing, Eva., Burst, Tobias., Chinnow, Sara Katharina. “Emotional citizens, detached interest groups? The use of emotional language in public policy consultations”. *Policy Sciences*, Vol. 56 (2023): 469-497.

Forte, Ana Catarina., Brazdil, Pavel B. “Determining the Level of Clients’ Dissatisfaction from their Commentaries”, *PROPOR* (2016): 74-85.

Google Cloud. “Analyzing Sentiment”. Accessed December 10, 2023. https://cloud.google.com/natural-language/docs/analyzing-sentiment

GOV.UK. “Consultations - what’s new and why they are so important”. Published January 15, 2016. https://civilservice.blog.gov.uk/2016/01/15/consultations-whats-new-and-why-they-are-so-important/

GovInfo. “Federal Register”. Accessed December 5, 2024. https://www.govinfo.gov/app/collection/fr

Ingrams, Alex. “Do public comments make a difference in open rulemaking? Insights from information management using machine learning and QCA analysis”. *Government Information Quarterly*, Vol. 40 (2023): 1-10.

Kovacs, Kathryn E. “Rules about Rulemaking and the Rise of the Unitary Executive”, *Administrative Law Review*, Vol. 70 (2018): 515-566.

Liang, Dehua., Frederick, David A., Lledo, Elia E., Rosenfield, Natalia., Berardi, Vincent., Linstead, Erik., Maoz, Uri. “Examining the utility of nonlinear machine learning approaches versus linear regression for predicting body image outcomes: The U.S. Body Project I”. *Body Image,* Vol. 41 (2022): 32-45.

National Archives. “Code of Federal Regulations List of Subjects”. Accessed December 13, 2023. https://www.archives.gov/federal-register/cfr/subjects.html

Naughton, Keith., Schmid, Celeste., Yackee, Susan Webb., Zhan, Xueyong. “Understanding commenter influence during agency rule development”. *Journal of Policy Analysis and Management,* Vol. 28 (2009): 258-277.

OECD. “Pilot database on stakeholder engagement practices”. Accessed March 20, 2024. https://www.oecd.org/gov/regulatory-policy/pilot-database-on-stakeholder-engagement-practices.htm

OECD. *A MENA-OECD Practitioners’ Guide for Engaging Stakeholders in the Rule-making Process*. Paris: OECD, 2012.

OECD. *Background Document on Public Consultation*. Paris: OECD, n.d.

OECD. *Regulatory Impact Analysis: Best Practices in OECD Countries.* Paris: OECD, 1997.

OECD. *Regulatory Policy Outlook 2021: Evidence-based policy making and stakeholder engagement*. Paris: OECD, 2021.

OECD. *The OECD Report on Regulatory Reform.* Paris: OECD, 1997.

Office of the Federal Register. *A Guide to the Rulemaking Process*. Washington D.C: Office of the Federal Register, n.d.

Rowe, Gene., Frewer, Lynn J. ”Public Participation Methods: A Framework for Evaluation“. *Science, Technology, & Human Values*, Vol. 25, (2000): 3-29.

Schwoerer, Kayla. “An exploratory study of social media's role in facilitating public participation in e-rulemaking using computational text analysis tools”. *Policy & Internet*, Vol. 15 (2023): 178-203.

Shapiro, Stuart. “Does the amount of participation matter? Public comments, agency responses and the time to finalize a regulation”. *Policy Sciences,* Vol. 41 (2008): 33-49.

Smith, James Andrew., Abhari, Roxanna., Hussain, Zain., Heneghan, Carl., Collins, Gary., Carr, Andrew. “Industry ties and evidence in public comments on the FDA framework for modifications to artificial intelligence/machine learning-based medical devices: a cross-sectional study”. *BMJ Open* (2020): 1-6.

U.S General Services Administration. “Regulations.gov API”. Accessed September 16, 2023. https://open.gsa.gov/api/regulationsgov/

U.S. General Services Administration. “Regulations.gov”. Accessed December 13, 2023. https://www.regulations.gov/

UK Parliament. "What is Secondary Legislation?". Accessed April 29, 2024. https://www.parliament.uk/about/how/laws/secondary-legislation/

United States. Administrative Procedure Act. Pub. L. No. 79-404, 60 Stat. 237 (1946).

World Bank. "Global Indicators of Regulatory Governance". Accessed May 25, 2024. https://rulemaking.worldbank.org/en/key-findings

Yackee, Susan Webb. "Sweet-Talking the Fourth Branch: The Influence of Interest Group Comments on Federal Agency Rulemaking". *Journal of Public Administration Research and Theory*, Vol. 16 (2006): 103-124.

Appendix

Appendix 1

Table 4. Words and other artefacts of preprocessing, which were excluded from lists of unique words.

regulation, provide, propose, approve, add, support, rule, s, a, c, would, make, l, de, re, f, also, u, r, d, n, st, t, b, in, j, be, o, and, g, i, ka, w, v, tn, va, h, q, or, an, na, amend, fr, noaa, epa, well, use, include, know, take, may, regulatory, sip, without, receive, agency, non, meet, provision, need, see, within, must, could, the, ensure, now, document, pm, wa, et, per, action, whether, k, eg, ie, te, apply, add, hb, due, ii, ep, tr, fe, le, proposes, project, x, br, bia, able, policy, y, implement, al, to, even, of, pre, th, co, shall, either, ee, eris, acre, so, en, example, iii, www, than, on, net, note, usc, reg, notice, no, ap, another, rd, ppm, el, pa, his, caa, therefore, se, nw, hr, he, gov, upon, by, ac, iv, mw, pe, ne, ph, generally, http, id, la, since, otherwise, rather, do, hg, po, cc, mg, ira, ag, cfc, at, com, thus, mm, ky, cw, among, already, vi, ic, rf, var, er, appendix, authority, ha, mr, tal, lb, tp, sca, ab, un, there, ra, ly, via, di, mi, me, cr, dun, hh, go, tar, us, ft, ct, much, held, kkk, z, cem, org, pi, tt, around, although, if, for, pfa, ed, btu, ce, ch, doe, ny, mt, pp, pcc, ga, af, ll, bt, fl, inc, astm, xx, yet, all, cd, om, moc, ba, bav, dp, actually, lee, sha, her, ol, rt, np, nj, rta, like, out, ppa, osha, esa, iso, ci, but, tc, il, ad, ncr, bi, over, tri, hi, nc, we, ng, dot, pc, ge, oar, gg, wsr, il, fg, dol, pr, iga, gcc, sr, oc, eq, arb, sc, acm, rip, hq, bb, sec, hap, vii, mc, ation, ive, put, bo, km, md, db, get, though, nd, notify, not, xxx, keep, mwh, etc, ml, gy, oh, eae, tac, hpv, xviii, cf, fic, lv, doi, hc, vlc, lot, whose, mf, iu, hv, say, especially, ec, ov, fda, fed, gor, rts, voc, cp, hhs, soon, fd, neither, py, faa, adj, usa, please, td, dd, si, tir, ar, fi, elk, pfd, sd, can, op, ner, cm, mo, gf, pt, eh, off, ti, cbs, pb, ta, pd, hua, ler, bc, vol, npr, rc, lrt, nm, gi, sw, con, ix, mp, pass, fo, attach, bii, mon, az, hip, qm, moreover, ert, have, viii, da, cb, xy, ric, nist, lek, ke, cv, want, iz, ef, each, up, lo, rm, mcr, chu, hemi, usda

Appendix 2

Table 5. Output of Regression with Averaged Normalized Term Frequency of Unique Words variable.

	Dependent variable:		
	Levenshtein distance	Cosine similarity	BERT-based semantic similarity
	(1)	(2)	(3)
Mean frequency of unique words	-46.872***	-44.129***	-9.100***
	(4.270)	(3.791)	(0.960)
Number of unique words	-0.001***	-0.001***	-0.0001***
	(0.00003)	(0.00003)	(0.00001)
Number of comments (log)	0.010***	0.013***	0.002***
	(0.002)	(0.002)	(0.0004)
Number of words in comments (log)	0.005*	0.010***	0.002***
	(0.003)	(0.002)	(0.001)
Sentiment score	0.019	0.030*	0.010**
	(0.019)	(0.017)	(0.004)
Sentiment magnitude	-0.011***	-0.006***	-0.002***
	(0.002)	(0.002)	(0.0005)
Constant	0.759***	1.060***	0.971***
	(0.172)	(0.153)	(0.039)
Agency fixed effects	Yes	Yes	Yes
Observations	2,208	2,208	2,208
R^2	0.671	0.637	0.617
Adjusted R^2	0.657	0.622	0.601
Residual Std. Error (df = 2119)	0.171	0.151	0.038
F Statistic (df = 88; 2119)	49.052***	42.212***	38.751***
Note:			*p<0.1; **p<0.05; ***p<0.01