RIDGE

RESEARCH INSTITUTE FOR DEVELOPMENT, GROWTH AND ECONOMICS

SINCE 2014

Complexity from Modular Legal Code: Applications in Financial Regulation

Malte Bammert,
Dietmar Leisen

RIDGE Working Paper 051

This paper was presented at the
2025 RIDGE December Forum

Complexity from Modular Legal Code: Applications in Financial Regulation*

Malte Bammert† and Dietmar Leisen‡
Johannes Gutenberg University

September 30, 2025

Abstract

This paper introduces a novel measure of regulatory complexity that extends the operator-based framework of Colliard and Georg (2025). Whereas their approach conceptualizes regulatory provisions as algorithms defined by the number of operators, our measure also incorporates the reference structure that arises from the modular drafting of legal code. Applying this methodology to Titles 12 and 17 of the Code of Federal Regulations, which govern banking and the U.S. Securities and Exchange Commission (SEC), we construct complexity scores at both the paragraph and aggregate levels. Our findings reveal that regulatory complexity has increased sharply since 1970, at a pace far exceeding the growth in word counts or operators alone. This rising complexity is mirrored in the costs borne by listed firms and banks.

JEL Classification: G18, G28, G41, K23.
Key words: complexity, regulation, references.

*We are grateful for the access to WRDS that Goethe University provided Malte Bammert in May 2024.
†Gutenberg School of Management and Economics, Johannes Gutenberg University, Mainz, Germany. Email: lubammer@uni-mainz.de; tel: +49-6131-39-22197
‡Gutenberg School of Management and Economics, Johannes Gutenberg University, Mainz, Germany. Email: leisen@uni-mainz.de; tel: +49-6131-39-25542

> "Any information included or incorporated by reference in a report filed under section 13(a) of the Act (15 U.S.C. 78m(a)) that is required to be disclosed pursuant to Item 402, 403, 404 or 407 of Regulation S-K (§ 229.402, § 229.403, § 229.404 or § 229.407 of this chapter) must be presented in a clear, concise and understandable manner." (Excerpt from Securities and Exchange Commission regulation on plain English, 17 CFR § 240.13a-20)

1 Introduction

In 1997, the Securities and Exchange Commission mandated that reports be written in plain English (see quotation above). Shortly thereafter, the Gramm–Leach–Bliley Act of 1999 extended this requirement to banking regulators[1]. In practice, agencies such as the Federal Reserve and the OCC often include a boilerplate assurance along the lines of: "The Board has sought to present the final rule in a simple, straightforward manner. The Board did not receive any comment on its use of plain language." While such statements appear to suggest that financial regulation is not unduly complex, this sits uneasily with the longstanding critique that the regulatory framework remains excessively complicated.

What, then, are we missing when we talk about complexity? The excerpt above points directly to the central issue of this paper. To begin, consider the Flesch–Kincaid Reading Ease Score (henceforth FRES), a widely used metric for evaluating plain English writing, see Kincaid et al. (1975) and Flesch (1979). The quoted passage receives a FRES of 63.5, suggesting that a reader with an eighth- or ninth-grade education should be able to follow it with ease. By this measure, the text qualifies as plain English.

Yet comprehension is not merely a function of sentence length or word choice. Full understanding is complicated by the dense web of cross-references to other regulatory provisions. This reliance on interlinked provisions reflects a modular drafting style that, while efficient for rulemakers, imposes additional burdens on readers. Our paper focuses on measuring precisely this added dimension of complexity, the complexity generated by the structure of regulatory referencing.

[1] "Each Federal banking agency shall use plain language in all proposed and final rulemakings published by the agency in the Federal Register after January 1, 2000. (...) For purposes of this section, the term 'Federal banking agency' has the meaning given that term in section 3 of the Federal Deposit Insurance Act." (Excerpt from Section 722 of the Gramm-Leach-Bliley Act).

Our measure builds on the concept of workload: the effort required to read and understand regulation. Without references, we define workload as the product of the FRES score and the number of "actions" (hereafter operators) that the regulated entity must observe, identified using the dictionary of Colliard and Georg (2025). When references are present, we augment this workload by adding that of the referenced provisions, weighted by a fixed, non-negative scaling factor. We then define regulatory complexity as the total workload, normalized by the number of operators.

We apply this approach separately to Titles 12 and 17 of the U.S. Code of Federal Regulations (CFR), yielding distinct measures of complexity for banking regulation and for securities regulation. We then situate these measures within the existing literature on regulatory complexity and show that both have increased manifold since 1970. Next, we relate our complexity measures empirically to two cost measures for non-bank firms and banks: general and administrative expenses for sales, and salaries. Our results demonstrate that the number of references and, ultimately, our composite complexity measure, possess explanatory power for both cost measures.

The literature has previously examined three dimensions of regulatory text complexity. First, the scope of regulatory requirements (Al-Ubaydli and McLaughlin (2017); Colliard and Georg (2025)). Second, the nesting resulting from references contained in the text (Li et al. (2008); Katz and Bommarito (2014)). Third, the linguistic difficulty of the text (Kincaid et al. (1975) and Flesch (1979)). This paper introduces a new measure that integrates these three aspects into a unified framework, thereby providing a more comprehensive assessment of regulatory complexity.

The remainder of the paper is structured as follows. The next section reviews the related literature. Section 3 introduces our novel reference-based measure of complexity. The subsequent section examines its empirical properties in the context of banking and securities regulation. The final section 5 concludes.

2 Literature Review

Starting with Christainsen and Haveman (1981) and Gray (1987) there is a large literature that looks at regulatory burden. Following Colliard and Georg (2025), we classify this broadly according to the analytical starting point into two lines of research.

The first line are text-based approaches that attempt to assess the complexity of the written formulation of regulatory texts, with the aim of determining the effort required to understand and comply with regulation on the basis of its textual expression. To this end, a range of measures has been proposed in the literature for capturing the complexity of regulatory texts (Li et al. (2008), Colliard and Georg (2025)). These measures focus on different aspects of textual complexity and are therefore referred to as "text-based" measures (Colliard and Georg (2025), pp. 2–3).

In contrast, another line of research takes the perspective of the regulated entities and seeks to quantify the (additional) costs they incur due to regulation. The indicators proposed in this strand of literature are thus referred to as "output-based" measures (Colliard and Georg (2025), p. 2). Several such measures have been developed in recent years (Calomiris et al. (2020); Kalmenovitz (2023); Trebbi et al. (2023)).

2.1 Text-based Approaches

The text-based approach to regulatory complexity is arguably the more intuitive one, as it seeks to identify the specific aspects of regulation that render it "complex." The main challenge of this approach, however, lies in justifying the "correct" definition of complexity, since establishing a quantitative link to the costs induced by regulation is generally difficult. This may be why the literature has introduced a wide variety of complexity measures over time. Essentially, these can be grouped into three broader categories[2], depending on the particular aspect of complexity they address.

The first group of measures primarily addresses the readability of a regulatory provision, focusing on what may be termed its lexicographical complexity. Indicators proposed in this category include, for example, the average length of words (de Lucio and Mora-Sanguinetti (2022)) or the frequency of word repetitions (Amadxarif et al. (2019)). Shorter words or a higher degree of repetition, and thus a reduced overall volume of text, are assumed to induce lower complexity. These measures ultimately build on the ideas of Kincaid et al. (1975) and Flesch (1979), whose approaches to assessing readability and textual difficulty remain widely applied to this day.

The second group of measures is concerned primarily with the scope of regulation. In

[2]Some contributions in the literature also adopt a hybrid approach, combining elements of all three groups of measures (Li et al. (2008); de Lucio and Mora-Sanguinetti (2022)).

its simplest form, this is captured by the number of words or lines contained in a regulatory text (Li et al. (2008)). Beyond this, however, the literature also discusses approaches that restrict attention to specific words implying instructions (Al-Ubaydli and McLaughlin (2017), Colliard and Georg (2025)). For instance, Al-Ubaydli and McLaughlin (2017) count the occurrences of only five terms (shall, must, may not, prohibited, required), whereas Colliard and Georg (2025) consider only words classified as logical, regulatory, or mathematical operators. The underlying intention is to conceptualize regulation as a sequence of instructions to be implemented in a defined order, akin to an algorithm. Accordingly, regulatory complexity is determined by the number of such instructions, identifiable through respective "keywords". This perspective also draws on decision analysis, particularly the EIP approach, which evaluates the effort of decision-making processes based on the number of elementary information processes involved (Bettman et al. (1990)).

The third group of measures examines the reference structure of regulation, often by analyzing the number of cross-references per paragraph (de Lucio and Mora-Sanguinetti (2022)). Here, regulatory texts are frequently modeled as graphs in which edges represent references within the document. Complexity is then captured by graph-theoretic metrics such as density (Li et al. (2008)). This approach also connects to psychological perspectives, as dense reference structures hinder the decomposition of regulation into smaller, meaningful units. From the standpoint of information theory, this may lead to information overload (Eppler and Mengis (2004)), which in turn can reduce decision quality (Chewning and Harrell (1990); Hwang and Lin (1999)) and, in this context, compliance with regulation.

2.2 Output-oriented Approaches

The counterpart to text-based approaches is provided by output-oriented approaches. These explicitly address the difficulty faced by text-based approaches in linking the complexity of a given regulation to the actual costs borne by the regulated. Instead, output-oriented approaches place emphasis on quantifying the costs incurred by regulated entities, without necessarily disentangling which specific aspects of regulation generate these costs.

A common strategy in this literature is to treat the introduction of new regulation as an exogenous shock. For instance, Zhang (2007) analyzed the costs induced by the Sarbanes-Oxley Act (SOX) by examining stock market reactions to SOX-related news, whereas Ahmed

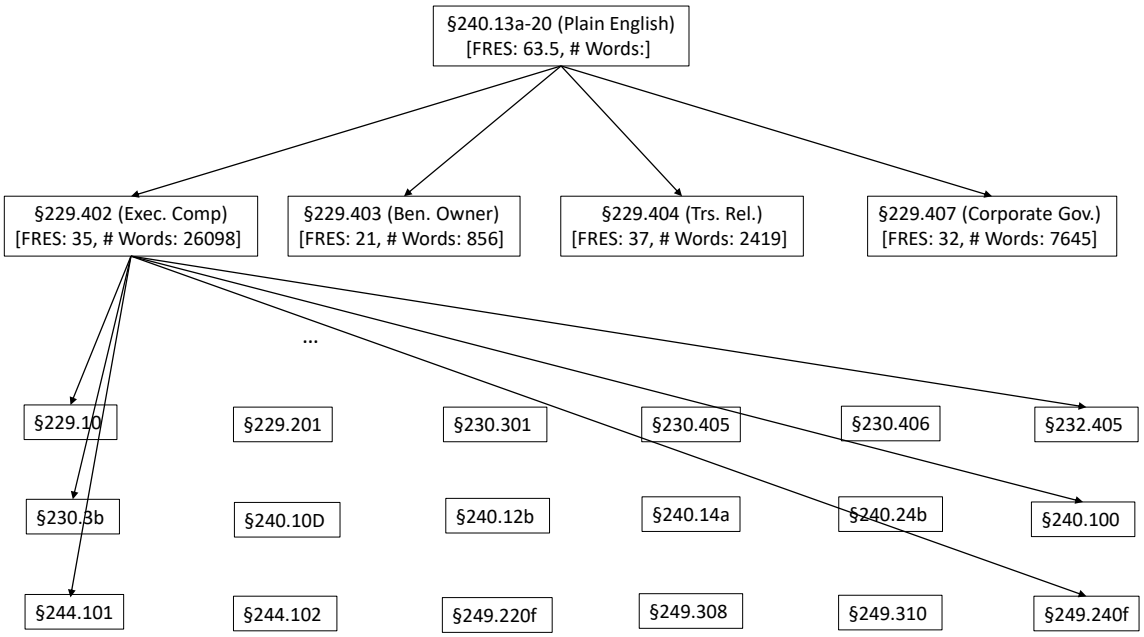
et al. (2010) studied the associated costs through the profitability channel (cash flows). Both studies provide evidence of substantial costs linked to the enactment of SOX, which tightened accounting and disclosure requirements for listed firms in the wake of the early 2000s financial scandals. Similarly, Hogan and Burns (2019) provide evidence of the costs associated with the Dodd-Frank Act of 2010, enacted to strengthen banking regulation after the 2008 financial crisis, by examining banks' non-interest expenditures.

Beyond the examination of regulatory introductions, the literature has also proposed measures to estimate the costs of existing regulation (Calomiris et al. (2020); Kalmenovitz (2023); Kalmenovitz et al. (2025); Trebbi et al. (2023)). For example, Kalmenovitz (2023) relies on compliance cost estimates reported by federal agencies on each mandatory corporate filing. Using a machine-learning approach, he aggregates these estimates across all forms required of a given firm to derive total compliance costs, finding that these estimated totals are associated with higher overall costs for firms. Trebbi et al. (2023), by contrast, employ U.S. Bureau of Labor Statistics data to calculate labor costs associated with regulatory activities, showing that regulatory labor expenses account for between 1.33% and 3.33% of total wage costs, depending on the definition used. Calomiris et al. (2020), in turn, construct a measure based on the share of regulatory content in earnings warnings and demonstrate that greater regulatory salience is associated with slower revenue and earnings growth and lower profitability.

Several studies also document a marked increase in the costs of compliance over time. Kalmenovitz (2023), using federal agency estimates, finds that compliance costs for firms have increased elevenfold since 1980 and doubled since 1996. Trebbi et al. (2023) similarly report rising labor costs attributable to a rising trend in regulatory activities during the period under examination, 2002 to 2014.

Beyond this evidence of escalating compliance costs, the literature also points to broader economic consequences of expanding regulation. For example, Singla (2023) provides evidence that increasing regulatory intensity has contributed to greater market concentration. This finding is supported by Trebbi et al. (2023), who show that the share of regulatory labor costs in total labor costs follows an inverted U-shape with firm size. Thus, compliance costs may act as a barrier to firm growth and foster market concentration. Another example of negative effects of excessive regulation is provided by Bailey and Thomas (2017), who find that industries facing higher regulatory burdens experience fewer firm entries and lower job

Figure 1: Illustrating the Reference Structure of the SEC's Plain English Regulation. All references are within title 17 of the United States Code of Federal Regulations.



creation. Moreover, Bailey and Thomas (2017) show that small firms are disproportionately affected, as they are more likely to exit heavily regulated industries. In a similar vein, Engel et al. (2007) conclude that increasing regulatory requirements have contributed to the delisting trend in the United States. The perception of rising economy-wide costs of regulation is further reinforced by Coffey et al. (2020), who argue that the expansion of regulation since 1980 has reduced the annual U.S. growth rate by 0.8 percentage points.

In addition to direct economic effects, the literature also highlights socially adverse consequences of excessive regulation. Chambers and O'Reilly (2022) find that U.S. states with a higher density of regulatory provisions exhibit greater income inequality. Likewise, Chambers et al. (2019) present evidence that higher levels of regulation at the state level are associated with increased poverty rates.

3 Measuring Complexity

This paper argues that cross-references within regulations are a key driver of regulatory complexity. Our analysis takes the paragraph as the basic unit of study, examining both references within a paragraph and those between paragraphs. This choice is motivated by practical considerations: regulations themselves typically use the paragraph as the smallest explicit unit of reference, making it a natural starting point. Conceptually, however, our approach is flexible and could be adapted to coarser or more granular units.[3]

Figure 1 provides an illustration of "references" and its structure. Therein, our starting point is the Securities and Exchange Commission's regulation on plain English, title 17 § 240.13a-20 in the U.S. Code of Federal Regulations (CFR). This paragraph references four other paragraphs within the same title 17 that detail specific issues: §229.402 (entitled "Executive Compensation"), §229.403 (entitled "Security ownership of certain beneficial owners and management"), §229.404 (entitled "Transactions with related persons, promoters and certain control persons") and §229.407 (entitled "Corporate Governance"). Each of these four paragraphs refers to other paragraphs within title 17; for simplicity Figure 1 exhibits *only* the 18 paragraphs that §229.402 refers to: §229.10, §229.201, §230.301, §230.405, §230.406, §232.405, §230.3b, §240.10D, §240.12b, §240.14a, §240.24b, §240.100, §240.101, §240.102, §249.220f, §249.308, §249.310 and §249.240f. We do not depict this here, but each of these 18 paragraphs may reference other paragraphs of title 17.

Figure 1 focuses on a single paragraph ("plain English"). Yet even understanding this paragraph in isolation quickly becomes a tedious task, as it requires consulting at least four other referenced paragraphs, which in turn point to many more. (Here we depict only 18 referenced by one of them.) The process demands working through long stretches of text. Although each is, in principle, written in "plain English," it is difficult to hold all of them in mind simultaneously. As the notion of "mental overload" suggests, each additional reference compounds the difficulty. This observation underlies our approach of treating references explicitly as a key component of regulatory complexity.

[3]Using coarser units would shift some references from occurring between units to occurring within them, whereas finer units would have the opposite effect. The former adjustment is straightforward, while the latter would require machine reading techniques, which fall beyond the scope of this paper.

3.1 Workload of a Paragraph without References

The Flesch-Kincaid Reading Ease Score (FRES) of a text (or of any paragraph of it), denoted L, was introduced by Kincaid et al. (1975) and Flesch (1979). It is a widely used measure of the readability of texts. Higher FRES values correspond to easier texts, with scores above the conventional cutoff of 60 typically classified as "plain English."

While the FRES does not account for references, we adopt it as the foundation of our approach. For our purposes, we require a measure F of reading complexity, such that higher values indicate greater complexity. We therefore define F by a simple linear transformation of the FRES L:

$$F = 2 - \frac{L}{60}.$$

Under this definition, higher F values correspond to more complex texts, and the "plain English" cutoff aligns with $F = 1$.

By construction, the FRES is unbounded, and values outside the interval $[0, 1]$ are possible. In fact, the literature documents examples with FRES values in the range of -500, which would translate into complexity levels around $F \approx 6$ under our transformation.

The FRES, and therefore our complexity measure, rests on two components: the average number of words per sentence and the average number of syllables per word. Put differently, text complexity is determined by the (average) length of sentences and the (average) length of words.

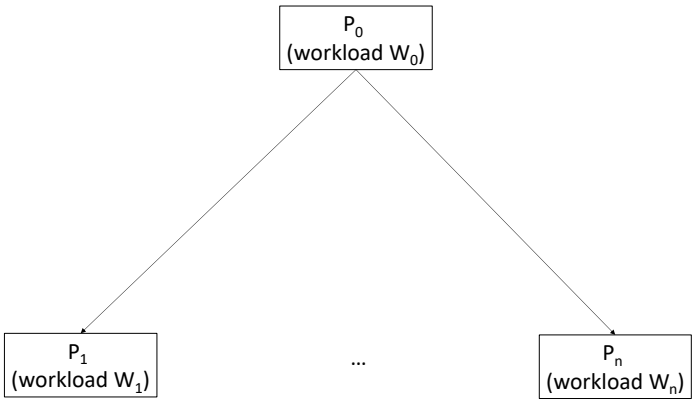
However, legal texts such as laws, regulations, and orders introduce an additional source of complexity. They prescribe actions using specialized vocabulary: terms such as "(dis-)allow", "deem", "exempt", "mitigate" and related nouns as well as adverbs. Such words, though rare in everyday prose, play a central role in regulatory language and make interpretation more demanding. In short, they contribute to complexity and should be reflected in a suitable measure.

We build on Colliard and Georg (2025), who cataloged these specialized terms, referred to as operators. Absent references, a straightforward approach is to count the number of operators N_A of a text "A" and define its "workload" W_A as the product of this count with the complexity score:

$$W_A^N = F_A \cdot N_A \text{ and } C_A = W_A / N_A, \tag{1}$$

where C_A represents the relative (or proportional) workload, i.e., the average complexity per

Figure 2: Illustrating the Reference Structure of a Paragraph.



regulatory operator.[4]

In this way, $C_A = F_A$ corresponds to the baseline case: the average complexity per operator in a text without references.

3.2 Workload of a Paragraph with References

We now extend our complexity measure to account for the case of a paragraph that references other paragraphs.

Without loss of generality, let paragraph P_0 reference n other paragraphs, denoted P_i, $i = 1, \ldots, n$. Let F_0 and N_0 denote the reading complexity and the number of operators of P_0, respectively. For each referenced paragraph P_i, we denote its workload by W_i. Figure 2 provides an illustration.

Reading paragraph P_0 requires not only processing its own content but also consulting all referenced paragraphs P_i ($i = 1, \ldots, n$), each with workload W_i. We therefore define the total workload of P_0, incorporating its references, as

$$W_0^R = W_{P_0 \text{ with references}} = F_0 N_0 + c \sum_{i=1}^{n} W_i, \tag{2}$$

[4]Alternative specifications are possible. For instance, one could assign fractional weight to operators, e.g. $W_A = F_A \cdot (1 + N_A/10)$.

where $c > 0$ is a fixed parameter that controls how much a referenced paragraph contributes to the workload of the paragraph that "origin".

A "naïve" way to assess the complexity in the situation depicted in Figure 2 is to consider the workload of paragraph P_0 in isolation, disregarding its reference structure.[5] Based on the definition in equation (1), this approach would yield

$$W_0^N = W_{P_0 \text{ without references}} = F_0 N_0. \tag{3}$$

If the referenced paragraphs P_i, $i = 1, \ldots, n$, do not themselves reference further paragraphs, then their workloads are given directly by

$$W_i = F_i \cdot N_i,$$

where F_i and N_i denote the reading complexity and the number of operators of paragraph $i = 1, \ldots, n$, respectively.

If, however, the referenced paragraphs contain further references, the scheme in equation (2) can be applied, provided that the workload W_i of each paragraph is well-defined. In this way, equation (2) offers a recursive procedure to determine total workload, as long as every chain of references eventually terminates in paragraphs for which workload can be computed as in subsection 3.1.

This assumption, however, does not always hold. A classic difficulty, known as the dictionary paradox, arises when definitions (or references) form a closed loop: a word in a dictionary refers to another, which refers to another, until eventually circling back to the original word. Analogously, regulatory references may create loops among paragraphs, preventing the workload of any paragraph in the cycle from being defined under the recursive scheme.

This assumption, however, does not always hold. A classic difficulty, known as the dictionary paradox, arises when definitions (or references) form a closed loop: a word in a dictionary refers to another, which refers to another, until eventually circling back to the original word. Analogously, regulatory references may create loops among paragraphs, preventing the workload of any paragraph in the cycle from being defined under the recursive

[5] This should not be misunderstood as ignoring the referenced paragraphs entirely. They still contribute when determining the workload of the text as a whole; see the discussion in the next subsection.

scheme.

3.3 Workload and Complexity of the Entire Text

The first step is to identify in a regulatory text the paragraphs $P_i, i = 1, \ldots, m$ and all the references. We discussed on Figure 1 the task of identifying references on the (part of) paragraph §240.13a describing "plain English" and stress that this has to be done for all paragraphs in the regulatory text analogously.

The next step is then to note that the previous two subsections defined the workload W_i^R for every paragraph i therein. We then define the workload $W_{1,\ldots,m}^R$ of the entire text as the sum of all these workloads.

Finally, using the number of operators N_i in each paragraph, we define the complexity of the entire text as

$$C_{1,\ldots,m}^R = \frac{W_{1,\ldots,m}^R}{\sum_{i=1}^m N_i},$$

One may view this as an average Flesch-Kincaid complexity of paragraphs, weighted by the referencing in the text.

3.4 Redoing the Workload in Assessing Regulatory Paragraphs

Disregarding the complications introduced by loops, the recursive structure in equation (2) implies that the "naïve workload" of a paragraph may be counted multiple times. To illustrate, consider the case shown in Figure 2, assuming that the referenced paragraphs do not themselves contain further references. In this case, the workload of a referenced paragraph P_i is simply $F_i N_i$.

Under the naïve approach (ignoring references), the workload of paragraph P_0 is

$$W_0^N = F_0 N_0.$$

By contrast, under the recursive approach that accounts for references, it becomes

$$W_0^N = F_0 N_0 + c \sum_{i=1}^{n} F_i N_i.$$

Consequently, the workload of the entire text is given by

$$W^N_{\text{text}} = F_0 N_0 + \sum_{i=1}^{n} F_i N_i$$

under the naïve approach, and by

$$W^R_{\text{text}} = W^R_0 + \sum_{i=1}^{n} F_i N_i = F_0 N_0 + (c+1) \sum_{i=1}^{n} F_i N_i$$

with our recursive approach, the one that takes into account references.

The key difference lies in the multiplicative factor: the naïve approach counts the contribution of the referenced paragraphs only once, whereas our recursive approach counts it $c + 1 > 1$ times. This reflects the core insight of our framework: referencing generates additional workload because each referenced paragraph must be reassessed in full whenever it is encountered.[6]

4 Complexity of Regulation

The previous section introduced our novel complexity measure. We now apply it to study the evolution of complexity in two major areas of U.S. financial regulation. First, we examine the provisions of the U.S. Securities and Exchange Commission (SEC) from 1970 to 2024, codified in Title 17 of the Code of Federal Regulations (CFR). Second, we analyze capital rules in banking regulation, codified in Title 12 of the CFR.

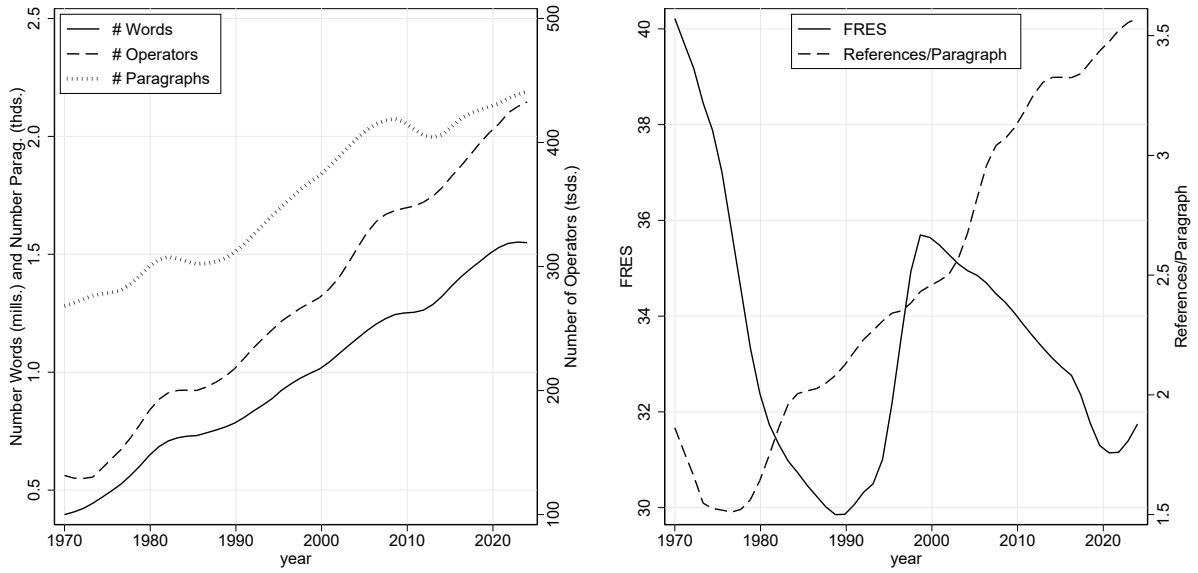
The SEC's regulations primarily concern listed companies.[7] This is of particular interest because rising regulatory burdens have been linked in the literature to both the decline in the number of listed companies (Engel et al. (2007)) and increased market concentration (Singla (2023)) in the United States.

Capital rules were dispersed across Title 12 until 1996, but since 1997 have been consolidated into a dedicated subpart. Extracting the relevant provisions prior to 1997 is tedious,

[6] Of course, deep familiarity with a regulation may allow specialists to navigate such references without re-reading them from scratch. Nevertheless, our framework captures the workload faced by a non-specialist reader.

[7] They also address other aspects of creditor protection, which we omit for simplicity.

Figure 3: Complexity of Securities Regulation according to basic measures of it.



error-prone, and would distract from the focus of this paper. For this reason, our analysis of banking capital regulation begins in 1997.

This section is structured as follows. The first subsection presents descriptive statistics on the complexity of securities and banking regulation. The subsequent subsections then examine the costs of regulation separately for securities and for banking.[8]

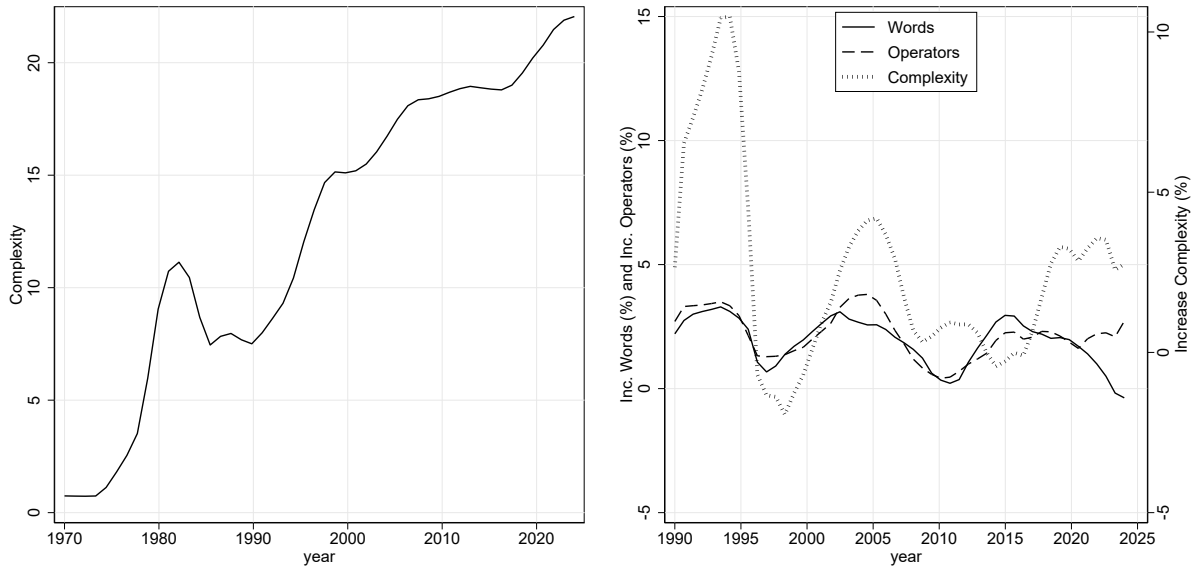
4.1 Descriptive Statistics

4.1.1 Securities Regulation

Figure 3 illustrates how the complexity in *securities* regulation evolved from 1970 to 2024 according to various basic measures from the literature. Left-hand we see that over that time period the number of words roughly tripled, that the number of paragraphs increased by almost one-half, that the number of operators increased almost four-fold and that all these

[8]Regulatory costs can be divided into two components. First, one-time costs arise from understanding new rules or amendments and establishing compliance processes. Second, ongoing costs result from maintaining compliance, such as staffing compliance functions, collecting data, and filing paperwork. Our analysis focuses on ongoing costs. While one-time costs could in principle be studied for banks given available data, we abstract from this to simplify our discussion.

Figure 4: Illustrating our complexity measure of securities regulation over the years since 1970 relative to that base year (left-hand) and YoY percentage changes in words, operators and complexity since 1990.



increased almost linearly over time. Note that the latter one (number of operators) is the measure put forward by Colliard and Georg (2025) and confirms that complexity increased considerably over time, more than a simple counting of words would suggest.

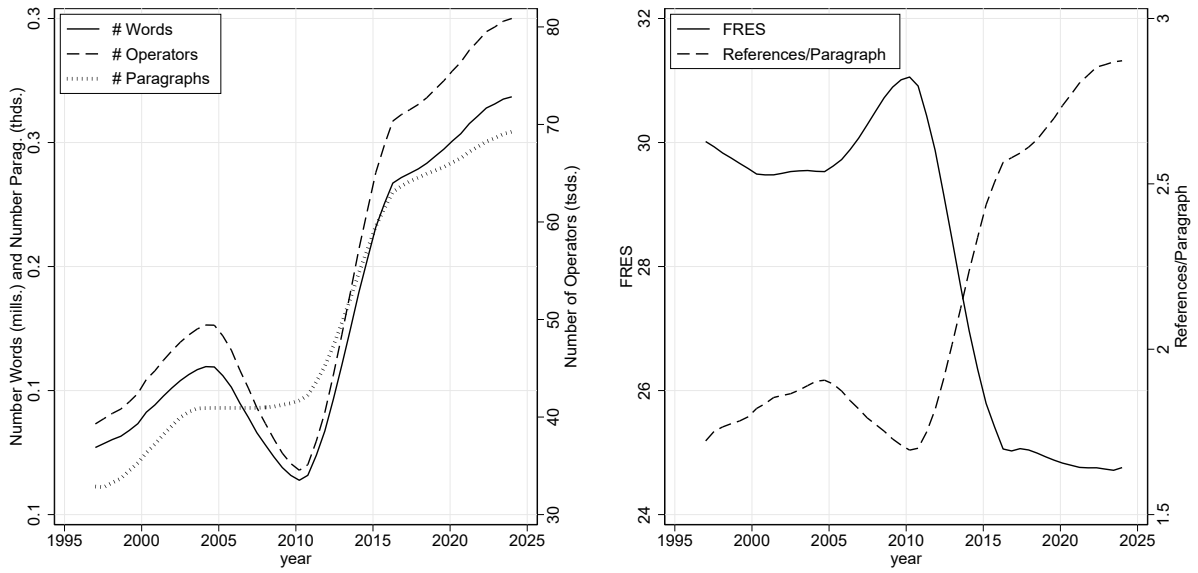
Getting closer to our complexity measure, the right-hand plot illustrates the two main parameters that drive it: the Flesch-Kincaid reading easy score (FRES) and the average references in paragraphs, defined as ratio of the number of references to the number of paragraphs.

We see that the linguistic complexity of SEC regulation is always at college level. We recall that lower FRES readings correspond to more complicated texts. Thus, initially complexity increased mostly up to the early 1990's, decreased heavily in the second half of that decade and has been slowly increasing again, since then.

The other main driver of our measure of complexity are references. The right-hand plot illustrates that the average number of references in paragraphs increased almost linearly and has almost doubled since 1970.

Figure 4 illustrates left-hand the relative increase in our complexity measure since 1970.

Figure 5: Complexity of Banking Regulation according to basic measures of it.



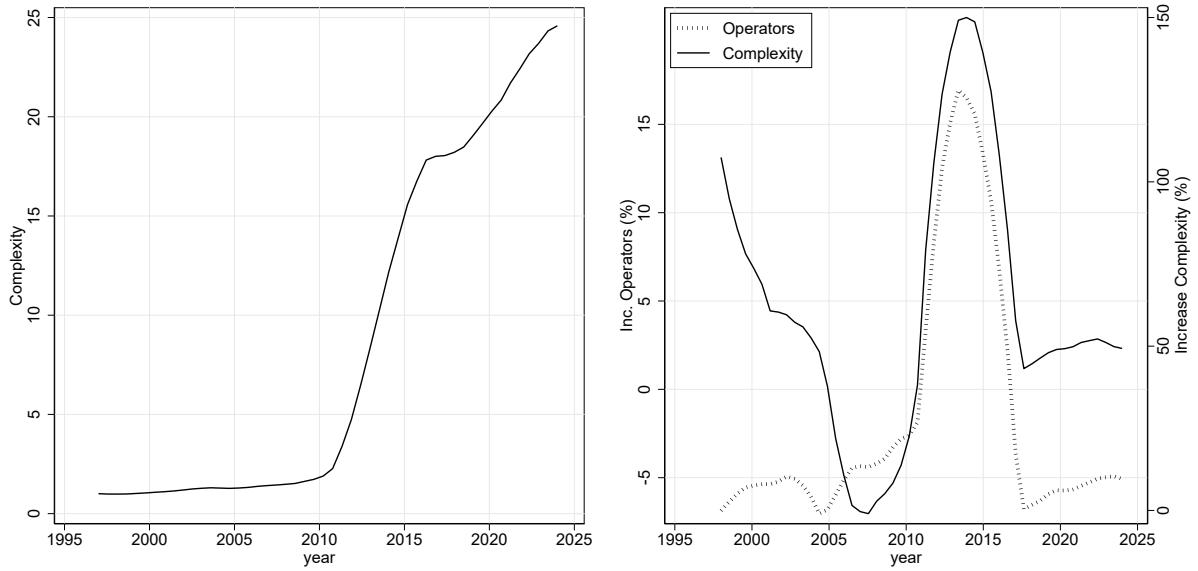
In the late 1970's, it exhibits a large increase in complexity that was partially reversed in the early 1980's and then staid somewhat flat in the later parts of that decade. Up to then complexity increased roughly 10-fold. Since the 1990's then, there is a continuous increase in complexity, roughly doubling it. The right-hand plot illustrates YoY change in complexity since[9] 1990, which hovers on average around 5.2%. For comparison, we plot also the YoY percentage changes in the number of words and operators, which hover around 2.0% and 2.3%. Note that the average rate of YoY changes in our measure is almost double that of words and operators, thus measuring much stronger increases YoY and, ultimately, a much higher increase over the entire time period under consideration.

4.1.2 Bank Regulations

Figure 5, analogous Figure 3, shows how the complexity of *banking* regulation evolved from 1997 to 2024, based on several standard measures from the literature. Recall that Figure 3 studied a longer period 1970 to 2024, nearly twice as many years.

[9]As noted in the left-hand plot of Figure 4, complexity decreased strongly in the early 1980's, leading to large YoY percentage changes that deteriorate the scale in the right-hand plot. For this, we start plotting with year 1990.

Figure 6: Illustrating our complexity measure of banking regulation over the years since 1970 relative to that base year (left-hand) and YoY percentage changes in words, operators and complexity since 1990.



We note that securities regulation seems far more complex than banking regulation based on two basic measures of it: taking account of the smaller time period in Figure 3, the number of paragraphs (operators) of securities regulation is on a scale of 1.0 to 1.5 (200 to 400), whereas in the corresponding Figure 5 the scale is on a scale of 0.10 to 0.25 (40 to 80). However, getting closer to our measure of complexity, both financial regulations are roughly comparable of the in terms of FRES banking regulation and the FRES is much lower (hence more complex) for the banking regulation when compared to securities regulation.

Figure 6 is analogous Figure 4, again looking only at years after 1996. Some notable differences are that banking regulatory complexity according to our measure was at a much lower level up to 2010 than securities regulation. (In fact, our measure suggests that complexity decreased between 2005 and 2010.) However, since 2010 the complexity of banking regulation increased considerably, ending in 2024 at a level comparable to securities regulation. (Note the large YoY increase in complexity of banking regulation, which runs around 50% to 150%.)

4.2 The Cost of Regulation for Listed Non-Banks

Costs of regulation are here the costs induced by compliance with the existing regulatory requirements. This includes wages for employees whose activities are part of the internal compliance processes. Moreover, we include periodic costs arising from regulatory compliance, such as fees for auditors.

As the data basis for the ongoing regulatory costs we use a CRSP/Compustat dataset covering the years 1986 to 2023, accessed through WRDS. We restrict ourselves to companies that are not banks, are headquartered in the United States and listed on one of the major exchanges (NYSE, American SE, NASDAQ).

To measure the ongoing costs of regulation, we use the general and administrative expenses of sales (SGA). Within this cost category, the ongoing compliance costs associated with SEC regulation are (most likely) only a (small) share. Nevertheless, the compliance effort should be reflected here. The aim of the analysis is therefore to examine the effect of regulatory complexity, measured with our metric introduced here and alternative measures, on this cost component of firms, while controlling for firm-specific characteristics. The model for analyzing the general costs is as follows:

$$\text{SGA}_{it} = \alpha_0 + \alpha_i + \beta_1 \text{Complexity}_t + \beta_2 \log(\text{Sales}_{it}) + \beta_3 \text{GDP-growth}_{it} + \varepsilon_{it}.$$

Table 1 presents four regression in the four columns. They differ in how we measure "Complexity". It is is either our measure of it, the number of operators, the number of references, or the FRES index. For better comparability, SGA has been adjusted for inflation since 1998. Moreover, as it is highly skewed, we study its natural logarithm, throughout. The same applies to Sales.

Table 1 shows that regulatory complexity increases SGA expenses and this is evident through all four measures. First of all, lower FRES score (corresponding to more "difficult" texts) increase costs. Moreover, increasing the number of operators, a measure put forward by Colliard and Georg (2025), increases costs.

Throughout this paper we argue that references matter for complexity. Hence, most important to us, increasing the number of references does increase costs. Finally, increases in our complexity correspond to higher costs for firms. These results suggests that the complexity of regulation impacts listed companies and that reference as well as our complexity

Table 1: Studying how the (logarithm of the) SGA component of non-bank firms is affected by various complexity measures of SEC regulation. Coefficient statistics are calculated using robust methods; t-statistics are reported in parentheses; significance levels refer to * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

	(1) Log SGA	(2) Log SGA	(3) Log SGA	(4) Log SGA
Complexity	0.000428***			
	(52.47)			
Number Operators		0.0000347***		
		(70.66)		
Average Reference			0.448***	
			(71.12)	
FRES Index				-0.00672***
				(-12.13)
Log Sales	0.727***	0.658***	0.667***	0.774***
	(195.05)	(158.76)	(163.95)	(231.94)
Number Employees	0.00129***	0.00129***	0.00128***	0.00128***
	(14.52)	(14.16)	(14.20)	(14.60)
GDP Growth	-1.189***	-1.060***	-0.962***	-2.469***
	(-27.63)	(-26.78)	(-24.19)	(-51.09)
Constant	-0.264***	-0.691***	-0.621***	0.0944***
	(-16.30)	(-56.18)	(-48.80)	(3.83)
Observations	142529	142529	142529	142529
Adjusted R^2	0.972	0.974	0.974	0.970
F	42886.2	58404.6	58011.0	26561.1

measure capture some aspects of it.

4.3 The Cost of Regulation for Banks

Whereas the previous subsection studied the complexity of securities regulation, we now take a look at the complexity of banking regulation. Therein, we also take a look at listed banks, which are affected by the complexities of both regulations.

In the United States, Bank Holding Companies (BHCs) have been required to report their salary expenses within the FR Y-9C reports, which they must submit quarterly to

Table 2: Studying how (the logarithm of) total salaries in banks are affected by various complexity measures of Bank regulations. Coefficient statistics are calculated using robust methods; t-statistics are reported in parentheses; significance levels refer to * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

	(1) Log Salaries	(2) Log Salaries	(3) Log Salaries	(4) Log Salaries
FED: Complexity	0.000333**			
	(2.04)			
FED: Number Operators		0.00000624**		
		(2.05)		
FED: Average Reference			0.0507**	
			(2.19)	
FED: FRES Index				-0.00531**
				(-2.19)
Log Assets	0.872***	0.878***	0.877***	0.885***
	(36.25)	(41.39)	(41.33)	(49.97)
Number Employees	0.00000131***	0.00000129***	0.00000130***	0.00000126***
	(2.80)	(2.81)	(2.82)	(2.82)
GDP Growth	-0.0704	-0.125	-0.0453	-0.0620
	(-1.05)	(-1.58)	(-0.70)	(-0.93)
Constant	-2.401***	-2.503***	-2.499***	-2.410***
	(-7.27)	(-8.89)	(-9.02)	(-7.63)
Observations	23353	23353	23353	23353
Adjusted R^2	0.979	0.979	0.979	0.979
F	8760.1	7925.1	7775.9	6405.1

the Fed, making these data publicly available. Throughout, we control for the size of the BHCs through the (logarithm of) total assets, for the number of employees and the current economic environment in the form of the most recent GDP growth. Hence, we run the following regression:

$$\begin{aligned} \text{Salaries}_{it} &= \alpha_0 + \alpha_i + \beta_1 \text{Complexity}_t \\ &\quad + \beta_2 \log(\text{Assets}_{it}) + \beta_3 \text{Number Employees}_{it} + \beta_4 \text{GDP-growth}_{it} + \varepsilon_{it}. \end{aligned}$$

Table 2 presents the regression results. Again, they differ in how we measure banking "complexity". It is is either our measure of it, the number of operators, the number of references, or the FRES index.

The results are similar to those in the previous subsection: lower FRES, i.e. more complex texts in terms of wording increases salaries. Increasing the number of operators, the measure of complexity put forward by Colliard and Georg (2025) increases salaries. Moreover, the number of reference and our measure of complexity built on these impact salaries. These results confirm that studying (the number of) references in regulatory is important and that our now complexity measure captures important aspects of it.

For further analysis, we revisit securities regulation. The SEC section of Title 17 CFR imposes regulatory requirements both on all companies whose shares are publicly traded (on exchanges), including financial institutions, especially banks. Hence, listed banks should be impacted by both financial regulations and we now study these jointly. For this, we extend the above regression to include both a measure of securities complexity and one of banking complexity

Table 3 the results of such regressions, analogous to our prior two tables. The results match our all previous ones. Overall, our main insights so far are strengthened: the number of operators, the measure of complexity put forward by Colliard and Georg (2025) is important. Yet, it misses the number of reference which turns out to be important. Ultimately, our measure of complexity built on these captures important aspects of the complexity of regulation, both of securities and banking regulation.

5 Conclusion

This paper introduces a new measure of the complexity of regulatory texts, building on the approach proposed by Colliard and Georg (2025). In their framework, a regulatory provision is conceived as an algorithm, with its complexity determined by the number of regulatory instructions (operators) it contains. The measure presented here extends this framework by incorporating the reference structure embedded in regulation. The underlying idea is that the complexity of a single provision depends not only on the operators within it, but also on the aggregated complexity of all cited provisions, weighted by an appropriate factor.

To illustrate the application of our measure, we analyzed Titles 12 and 17 of the United

Table 3: Studying how (the logarithm of) total salaries in banks are affected by various complexity measures of SEC and Bank regulations. Coefficient statistics are calculated using robust methods; t-statistics are reported in parentheses; significance levels refer to * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

	(1) Log Salaries	(2) Log Salaries	(3) Log Salaries	(4) Log Salaries
FED: Complexity	0.000243***			
	(4.46)			
SEC: Complexity	0.000173***			
	(3.95)			
FED: Number Operators		0.00000180		
		(1.54)		
SEC: Number Operators		0.00000729***		
		(4.41)		
FED: Average Reference			0.00974	
			(1.15)	
SEC: Average Reference			0.139***	
			(9.01)	
FED: FRES Index				-0.00342***
				(-2.87)
SEC: FRES Index				-0.0115***
				(-3.56)
Log Assets	0.830***	0.829***	0.812***	0.844***
	(79.79)	(76.54)	(86.12)	(94.29)
Number Employees	0.00000101***	0.000000986***	0.000000997***	0.000000985***
	(2.86)	(2.80)	(2.80)	(2.81)
GDP Growth	-0.412***	-0.329***	-0.0744	-0.393***
	(-3.71)	(-2.92)	(-0.65)	(-3.52)
Constant	-1.780***	-1.880***	-1.728***	-1.327***
	(-13.15)	(-14.99)	(-14.47)	(-5.93)
Observations	4915	4915	4915	4915
Adjusted R^2	0.992	0.992	0.992	0.992
F	4464.3	4404.9	4380.0	4423.9

States Code of Federal Regulations (CFR), which contain the banking regulations and the regulations of the U.S. Securities and Exchange Commission (SEC), respectively. Using the dictionary developed by Colliard and Georg (2025), we identified all regulatory operators and cross-references in these titles. Based on this information, we calculated the complexity of each paragraph and then derived the overall complexity of banking and securities regulation as the sum across paragraphs.

Our analysis of securities regulation since 1970 shows a clear upward trend in complexity. Although the pace of growth has moderated somewhat since the mid-1990s, it continues to far outstrip the growth in word counts or operator counts. This rising burden is reflected in the costs borne by non-bank firms, which are most strongly affected by our measure of securities regulation complexity. Banking regulation has also become significantly more complex over time. We further show that regulatory costs in banks are directly influenced by the complexity of banking regulation, while listed banks face the additional burden of securities regulation complexity.

References

Ahmed, A.S., McAnally, M.L., Rasmussen, S., Weaver, C.D., 2010. How costly is the Sarbanes Oxley Act? evidence on the effects of the act on corporate profitability. Journal of Corporate Finance 16, 352–369.

Al-Ubaydli, O., McLaughlin, P.A., 2017. Regdata: A numerical database on industry-specific regulations for all United States industries and federal regulations, 1997–2012. Regulation & Governance 11, 109–123.

Amadxarif, Z., Brookes, J., Garbarino, N., Patel, R., Walczak, E., 2019. The language of rules: textual complexity in banking reforms. Bank of England working papers 834. Bank of England.

Bailey, J.B., Thomas, D.W., 2017. Regulating away competition: the effect of regulation on entrepreneurship and employment. Journal of Regulatory Economics 52, 237–254.

Bettman, J.R., Johnson, E.J., Payne, J.W., 1990. A componential analysis of cognitive effort in choice. Organizational Behavior and Human Decision Processes 45, 111–139.

Calomiris, C.W., Mamaysky, H., Yang, R., 2020. Measuring the Cost of Regulation: A Text-Based Approach. Working Paper 26856. National Bureau of Economic Research.

Chambers, D., McLaughlin, P.A., Stanley, L., 2019. Regulation and poverty: an empirical examination of the relationship between the incidence of federal regulation and the occurrence of poverty across the US states. Public Choice 180, pp. 131–144.

Chambers, D., O'Reilly, C., 2022. Regulation and income inequality in the united states. European Journal of Political Economy 72, 102101.

Chewning, E.G., Harrell, A.M., 1990. The effect of information load on decision makers' cue utilization levels and decision quality in a financial distress decision task. Accounting, Organizations and Society 15, 527–542.

Christainsen, G.B., Haveman, R.H., 1981. Public regulations and the slowdown in productivity growth. The American Economic Review 71, 320–325.

Coffey, B., McLaughlin, P., Peretto, P., 2020. The cumulative cost of regulations. Review of Economic Dynamics 38, 1–21.

Colliard, J.E., Georg, C.P., 2025. Measuring regulatory complexity. Forthcoming: Journal of Financial Economics .

de Lucio, J., Mora-Sanguinetti, J.S., 2022. Drafting "better regulation": The economic cost of regulatory complexity. Journal of Policy Modeling 44, 163–183.

Engel, E., Hayes, R.M., Wang, X., 2007. The Sarbanes-Oxley Act and firms' going-private decisions. Journal of Accounting and Economics 44, 116–145.

Eppler, M.J., Mengis, J., 2004. The concept of information overload: A review of literature from organization science, accounting, marketing, MIS, and related disciplines. The Information Society 20, 325 – 344.

Flesch, R., 1979. How to write Plain English: A book for Lawyers and Consumers. Harper and Row.

Gray, W.B., 1987. The cost of regulation: OSHA, EPA and the productivity slowdown. American Economic Review 77, 998–1006.

Hogan, T.L., Burns, S., 2019. Has Dodd-Frank affected bank expenses? Journal of Regulatory Economics 55, 214–236.

Hwang, M.I., Lin, J.W., 1999. Information dimension, information overload and decision quality. Journal of Information Science 25, 213–218.

Kalmenovitz, J., 2023. Regulatory intensity and firm-specific exposure. The Review of Financial Studies 36, 3311–3347.

Kalmenovitz, J., Lowry, M., Volkova, E., 2025. Regulatory fragmentation. Journal of Finance 80, 1081–1126.

Katz, D.M., Bommarito, M., 2014. Measuring the complexity of the law: the United States code. Artificial Intelligence and Law 2, 337–374.

Kincaid, P., Fishburne, R.P., Rogers, R.L., Chissom, B.S., 1975. Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel. Technical Report.

Li, H., Pincus, M., Rego, S.O., 2008. Market reaction to events surrounding the Sarbanes-Oxley Act of 2002 and earnings management. Journal of Law and Economics 51, 111–134.

Singla, S., 2023. Regulatory costs and market power. Working Paper.

Trebbi, F., Zhang, M.B., Simkovic, M., 2023. The Cost of Regulatory Compliance in the United States. Working Paper.

Zhang, I.X., 2007. Economic consequences of the Sarbanes-Oxley Act of 2002. Journal of Accounting and Economics 44, 74–115.