Measuring Regulatory Complexity

Jean-Edouard Colliard*
HEC Paris

Co-Pierre Georg†
Frankfurt School of Finance and Management

July 14, 2025

Abstract

We propose a framework to study regulatory complexity, based on concepts from computer science. We distinguish different dimensions of complexity, classify existing measures, develop new ones, compute them on three examples—Basel I, the Dodd-Frank Act, and the European Banking Authority's reporting rules—and test them using experiments and a survey on compliance costs. We highlight two measures that capture complexity beyond the length of a regulation. We propose a quantitative approach to the policy trade-off between regulatory complexity and precision.

Keywords: Regulatory Complexity, Financial Regulation, Basel Accords.
JEL classification: G18, G28, G41.

*E-Mail: colliard@hec.fr; Corresponding author.
†E-Mail: co.georg@fs.de

Acknowledgements

We are grateful to Tobias Berg, Bruno Biais, Claire Célérier, Dana Foarta, Stephen Karolyi, Igor Kozhanov, Iman van Lelyveld, Christoph Merkle, Vitaly Meursault, Sébastien Pouget, Sébastien Saurin, Shikhar Singla, Michael Troege, as well as to participants to the 2024 AFA, 2024 PSB Workshop on Banking and Finance, 2023 WFA, 2023 Chapman University's Finance Conference, 2022 SFS Cavalcade, 2022 ESEM, Paris Webinar on Financial Technology and Crypto, 2019 BCBS-CEPR Workshop on "Impact of regulation in a changing world: innovations and new risks", 2019 NY Fed Fintech Conference, 2016 "Journée de la Chaire ACPR", seminar audiences at ACPR, AMSE, AMF, Bank of England, HEC Paris, De Nederlandsche Bank, EBA, ECB, IESEG, OCC, University of Bonn, University of Exeter, and University of Luxembourg, and the Webinar "Metrics of Regulation", for helpful comments and suggestions. We are grateful to Olena Bogdan, Pietro Fadda, Ali Limon, Jane Douat, and Sabine Schaller for excellent research assistance. This work was supported by the French National Research Agency (F-STAR ANR-17-CE26-0007-01, ANR EFAR AAP Tremplin-ERC (7) 2019), the Investissements d'Avenir Labex (ANR-11-IDEX-0003/Labex Ecodec/ANR-11-LABX-0047), the Chair ACPR-HEC-TSE "Regulation and Systemic Risk", the Natixis Chair "Business Analytics for Future Banking", and the Europlace Institute of Finance. Part of this research was conducted while Colliard was visiting INSEAD, the hospitality of which is greatly appreciated.

The regulatory overhaul that followed the global financial crisis has triggered a hefty debate about the complexity of financial regulation. Haldane and Madouros (2012), for instance, argue that bank capital regulation has become so complex as to be counterproductive and likely to favor regulatory arbitrage. The Basel Committee on Banking Supervision acknowledges that there is a trade-off between the simplicity and the precision of regulation (BCBS (2013)). In the United States, similar concerns have led to the exemption of smaller banks from several provisions of the 2010 Dodd-Frank Act.[1]

Although there is a widespread concern that financial regulation has become too complex, "regulatory complexity" remains an elusive concept. Debates about the complexity of different rules and contracts have arisen in other contexts, such as structured financial products (Célérier and Vallée, 2017), securitizations (Ghent *et al.*, 2017), loan contracts (Ganglmair and Wardlaw, 2017), compensation contracts (Bennett *et al.*, 2019), or corporate taxes (Zwick, 2021). A growing number of papers propose measures and theories of the complexity of rules, but they focus on different dimensions of complexity, and a unifying framework is lacking. We propose such a framework and develop a toolkit including measures of complexity, validation experiments, and normative analyses. We use these ingredients to approach the trade-off studied by the BCBS (2013) in a quantitative manner.[2]

Our working hypothesis is to see a regulation as an algorithm: It is a sequence of instructions that is applied to an economic entity and returns a regulatory action. Previous research has used this analogy and focused on adapting some measures of algorithmic complexity to the study of law (see, e.g., Li *et al.* (2015)). We go further and use this approach to dis-

[1]See Gai *et al.* (2019) for a comprehensive discussion of the policy issues at stake and Calomiris (2018) for the case of the United States.

[2]To encourage further work within the same framework, we make the toolkit we developed available online: `https://cogeorg.github.io/MeasuringRegulatoryComplexity_Replication.zip`.

tinguish between different dimensions of complexity and derive five measures of regulatory complexity in a unified model of regulation. We compute these measures on two large-scale regulatory texts, the US 2010 Dodd-Frank Act (DFA) and the EU 2021 "Implementing Technical Standards" (ITS) for the Capital Requirements Regulation. We test their validity using both an experiment and a survey of banks conducted by the European Banking Authority (EBA). We find support in particular for two measures, *quantity* and *potential*, which we detail below. Finally, we outline how to include such measures in a normative model of the trade-off between regulatory precision and complexity.

We first use our framework to formally define measures of regulatory complexity and distinguish between different dimensions. In particular, we make a distinction between: (i) "intrinsic complexity"—a regulation is complex because it aims at imposing many different rules on the regulated, independently of the language used; (ii) "psychological complexity"—a regulation is complex because it is difficult for a reader to understand; and (iii) "computational complexity"—a regulation is complex because it is resource intensive to implement. We relate these dimensions to empirically observable quantities, namely the occurrence of mistakes in the regulatory process and the effort necessary to understand and apply the regulation. We call these quantities "outcome-based" measures of regulatory complexity.

The literature has studied several outcome-based measures of complexity, but, by definition, they measure the consequences of regulatory complexity, not its causes. To identify what is causing complexity and provide guidance to regulatory authorities, we also need "text-based" measures based on regulatory documents. Most existing measures are based on linguistics and only cover psychological complexity. To also derive measures of intrinsic complexity, we use the approach developed by Halstead (1977) to measure algorithmic complexity. As we detail in Section 1, this approach represents an algorithm as a sequence of "operators" (e.g., $+$, $-$, logical connectors) and "operands" (variables, parameters), and the measures of complexity aim at capturing the number of operations, inputs, and outputs. In the context of regulation, these measures can capture the number of different rules in a regulatory text, whether these rules are repetitive or different, whether they apply to different economic entities or to the same ones, etc. We show that within this model we can encompass two already existing measures of regulatory complexity and go on to introduce three new ones.

As a proof of concept, we first consider the design of risk weights in the Basel I Accords, which is close to following an actual algorithm. We write a computer code corresponding to the instructions of Basel I and measure the algorithmic complexity of this code directly. We then analyze the text of the regulation, classify words according to whether they correspond to operands or operators in the algorithm, and compute the same measures. This exercise shows that the concepts of "operands" and "operators" apply to a regulatory text. Moreover, we find that the measures supposed to capture intrinsic complexity, which should in principle not depend on the language used, are indeed close in both versions.

The next step is to compute our measures at scale. We first apply our textual approach to the DFA, which has been the focus of much of the debate on regulatory complexity and earlier literature (e.g., Li *et al.* (2015)). We compute the different measures for each title of the DFA and document new facts about the distribution of complexity measures across titles. We then perform a similar analysis with the ITS.

Text-based measures of regulatory complexity are only valid and useful if they are related to outcomes. To test our measures, we follow two complementary approaches. The first one is again inspired by the field of computer science, which uses experiments to test whether measures of algorithmic complexity predict the mistakes programmers make or the time they need to code a program.[3] We give participants to an experiment different regulatory instructions consisting of (randomly generated) Basel-I type rules and the balance sheet of a hypothetical bank. They have to compute the bank's risk-weighted assets. We analyze how different measures of complexity explain whether a participant returns a wrong value and the time taken to give a correct answer. We find that only two of the five measures in our framework have explanatory power beyond the length of the regulation, suggesting that our experimental design is a powerful touchstone to test the validity of new measures. The first measure is *quantity*, which captures the number of regulatory operators (words that indicate a new rule or constraint) in the text and is related to the existing RegData measure (Al-Ubaydli and McLaughlin, 2017). The second measure is *potential*, which captures the number of unique operands and hence the diversity of economic concepts. While *quantity* explains both the mistakes and the time taken to answer, *potential* explains only the latter. These two measures are those meant to capture intrinsic complexity, which validates the idea that this is indeed a dimension not captured by length alone.

Our second test is based on a survey conducted by the EBA, which asked EU banks to indicate which ITS templates are a source of high compliance costs, a major concern and outcome variable in the literature on regulatory complexity.[4] We compute our complexity measures for each template and test which measures are significantly correlated with high costs. Remarkably, the results are qualitatively similar to those obtained for mistakes in the

[3]See, e.g., Canfora *et al.* (2005) or Zhang and Baddoo (2007).

[4]As we detail in Section 3.2, the EBA estimates the compliance costs with these rules for banks at 19.6 bn EUR. For comparison, Hogan and Burns (2019) estimate the compliance costs related to the Dodd-Frank Act at 60 bn USD. Trebbi *et al.* (2023) estimate total compliance costs in the US at 239 bn USD for 2014.

experimental setting, despite the two settings being very different. In particular, we find that *quantity* is the only measure that can explain compliance costs beyond length, which is in line with the experimental results, as the respondents to the survey list the difficulty of understanding the rules as the main driver of compliance costs. Moreover, the effect of *quantity* is mainly driven by large and medium banks, whereas small banks react instead to *potential*. This is new evidence that firms may face different types of complexity costs depending on their size.

We complement these results with the analysis of a related survey by the EBA, which asked EU regulators which ITS templates are more important to them. We find that *potential* is strongly positively related to "importance", controlling for *length*. This supports the hypothesis that regulators draft intrinsically more complex regulations on topics they deem more important and that *potential* can capture this effect. On the contrary, *quantity* is negatively correlated with importance. This suggests that revising the ITS templates by reducing *quantity* could decrease costs for banks without negatively affecting the regulators.

We systematically compare the performance of *potential* and *quantity* with the performance of *length*. We find that for each of the 4 dependent variables considered in the paper, the specification associated with the highest R^2 includes either *length* and *potential* or *length* and *quantity*. In this favorite specification (but not in others), *length* is not significant at the 10% level and contributes much less to the R^2 than the alternative measure (from none at all to at most 5 times less).

These results mean that a policymaker aiming at simplifying a regulatory text should consider measures of intrinsic complexity, not only the length of the text. To further demonstrate the usefulness of our approach, we propose a research program that would eventually lead to a quantitative model of the trade-off between the precision and complexity of regulation mentioned in BCBS (2013). As a proof of concept, we build a simple model of capital regulation relying on risk buckets, as in Basel I. We use our measures and experimental estimates to compute the complexity cost of regulation and the optimal number of risk buckets.

Making further progress on this research program will require better connecting different text-based measures to outcomes at the firm level. In line with our belief that open science can accelerate this progress, we make all the material used in this paper available online. This includes the DFA dictionary with 5627 operands and 608 operators related to financial regulation, plus the EBA dictionary with an additional 982 operands and 234 operators. Researchers can use both to study other texts or as a training sample to identify other operands and operators using machine learning. We also share our data and code so that other researchers can test alternative text-based measures under the exact same conditions.

We review the literature on measures of regulatory complexity in Section 1.3, where we show how the different measures fit into our framework.[5] Our contribution to that literature is not only to propose new measures of intrinsic complexity (of which there are few), but also to use our framework to organize this literature and highlight the complementarities between different measures that capture different dimensions of complexity. Similarly, we review the theoretical literature on the causes and consequences of regulatory complexity in Section 1.2 and clarify the dimensions of regulatory complexity that each article is studying. Our measures of regulatory complexity make it possible to test these theories.[6]

[5]We do not include measures of algorithmic complexity more generally, and refer the interested reader to Zuse (1990), and Yu and Zhou (2010) for a more recent survey.

[6]Some empirical papers study the increase in the stringency or quantity of regulations. For example, Kalmenovitz (2023) shows that an increased regulatory intensity leads to a significant reduction in firm-level

As mentioned above, a number of papers have studied the complexity of financial products and contracts more generally. Our framework also applies to these applications, to the extent that they consider rules describing how to perform a certain operation.[7] There is more broadly a theoretical literature on complexity in product markets (see, in particular, Gabaix and Laibson (2006), Carlin (2009), and Ellison (2016) for a survey). The economic mechanisms studied in this literature are not easy to transpose to the complexity of regulation. In addition, Arora *et al.* (2009) argue that computational complexity creates a new form of asymmetric information when one agent is able to solve a computational problem and the other is not. Carlin *et al.* (2013) find support for this idea in a trading experiment, with adverse selection being larger for more complex assets.

In addition to finance applications, the experimental approach we use in Section 3.1 is related to the literature that tries to measure the complexity of solving mathematical problems for humans. In particular, Murawski and Bossaerts (2016) and Franco *et al.* (2021) ask experimental participants to solve different versions of the knapsack problem. Our experimental approach is conceptually similar, but the Halstead model we use is a more flexible representation of an algorithm, allowing us to apply our approach to entire regulatory texts and not only to well-identified mathematical problems and algorithms.

Finally, a literature on behavioral economics dating back to Rubinstein (1986) models the strategies of economic agents as automata and measures the cognitive costs of a strategy using the complexity of an automaton. Recent experimental work by Oprea (2020) and

investment and hiring. Gutiérrez and Philippon (2019) argue that the increase in regulation can account for the decline in the elasticity of entry with respect to Tobin's Q since the late 1990s.

[7]For example, we have applied our framework to study the complexity of the OECD's blueprints on the tax challenges arising from digitalization (Colliard *et al.*, 2021). In contrast, our approach does not in principle apply to the complexity of objects that are not rules, for instance firm disclosures, where complexity is probably better captured by stylistic or linguistic measures (e.g., Loughran and McDonald (2014)).

Kendall and Oprea (2024) validates this approach. We do not represent regulation as an automaton, as this would be costly to do on a large-scale text, and believe that the Halstead representation is an easier alternative for empirical studies.

1 Defining Regulatory Complexity

We first propose a model of the regulatory process that formally defines different dimensions of complexity, connects them to existing research, and makes the analogy between regulations and algorithms explicit. Then, we apply this comparison to introduce complexity measures from computer science and review how existing measures fit within our framework.

1.1 Framework

We start by defining the economic environment. The economy is in an observable state $\tilde{x} \in X$, where for convenience X is a finite set.[8] We denote $\nu(x) = \Pr(\tilde{x} = x)$. In every state $x \in X$, a regulator can choose a regulatory instrument y in a finite set Y. An economic outcome $\tilde{z} \in Z$ is obtained, where the probability distribution of $\tilde{z}$ depends on the state and the regulatory instrument, and is denoted $\mu(z, x, y) = \Pr(\tilde{z} = z|x, y)$. The regulator has preferences over economic outcomes, represented by a utility function U.

We define *regulation* as a function φ from the set of economic states X to the set of regulatory instruments Y. We denote by $\Phi = Y^X$ the set of such functions. In the example of bank capital regulation, x is the state of the balance sheet of a bank, and y is the level of capital requirements. A "capital regulation" is then a set of rules that determines the level

[8]This assumption is made to sidestep issues of measurability and plays no relevant economic role.

of capital requirements that applies for any possible balance sheet.

In a standard economic model, the mapping φ would be sufficient to describe the regulatory process, and φ derived as the optimal solution to the following maximization problem:

$$\max_{\varphi(.)} \sum_{x \in X} \nu(x) \left[\sum_{\tilde{z} \in Z} \mu(\tilde{z}, x, \varphi(x)) U(\tilde{z}) \right]. \tag{1}$$

We enrich this standard approach by considering that the process of writing down the rules and then implementing them is costly and error-prone. We distinguish the regulation φ, a function, from the *regulatory text* F that describes it. Formally, F is a list of elements (words) in a finite *vocabulary* $\mathcal{V}$: $F \in \bar{\mathcal{V}}$, with $\bar{\mathcal{V}} = \bigcup_{n=1}^{+\infty} \mathcal{V}^n$.

The first step in the regulatory process is *drafting*: the regulator drafts a regulatory text F to describe the regulation φ. To account for mistakes in the draft, we assume that the text $\tilde{F}$ is random, with μ_D representing the probability distribution over possible drafts. This distribution depends on the regulation φ, the regulator's type θ (e.g., skill), and the drafting effort e (e.g., hours of work). Formally:

$$\forall F \in \bar{\mathcal{V}}, \Pr(\tilde{F} = F) = \mu_D(F, \varphi, \theta, e). \tag{2}$$

The second step of the regulatory process is *interpretation*: an economic agent must interpret the regulatory text F to comply with the regulation. The interpretation $\tilde{f} \in \Phi$, maps states of the economy to regulatory tools. The *correct interpretation*, denoted $I(F) \in \Phi$, is the function actually defined by the text.[9] However, interpretations can be incorrect. As in the

[9]Depending on the distribution μ_D, it could be that F does not describe a proper function, so that $I(F) \notin \Phi$. We abstract from this possibility: as our definitions of complexity in Section 1.2 rely on correctly describing φ, it does not matter whether F describes a proper function.

drafting step, we assume that $\tilde{f}$ is random, with μ_I representing the probability distribution over Φ. This distribution depends on the regulatory text F, the interpreter's type θ, and effort e:

$$\forall f \in \Phi, \Pr(\tilde{f} = f) = \mu_I(f, F, \theta, e). \tag{3}$$

The final step in the regulatory process is *compliance*: a regulated agent implements the regulation f in the current state (e.g., the characteristics of the firm) to compute $f(x)$. Errors are again possible: the regulatory instrument $\tilde{y}$ follows a probability distribution μ_S over Y, which depends on the interpretation f and the supervisor's type θ and effort e:

$$\forall (x, y) \in X \times Y, \Pr(\tilde{y} = y) = \mu_D(y, x, f, \theta, e). \tag{4}$$

Returning to the example of bank capital regulation, we begin with a regulation φ, such as the Basel I capital regulation. This regulation is detailed in a regulatory text F (e.g., BCBS (1988)). If the text F is drafted accurately, then its correct interpretation $I(F)$ should align with the intended regulation φ, so that $I(F) = \varphi$. A compliance officer must read and interpret this text. If her interpretation f is correct, then $f = I(F) = \varphi$. Then, she applies the regulation to the characteristics x of the bank and computes the regulatory instrument y. If the computation is correct, then $y = f(x) = I(F)(x) = \varphi(x)$, so that the actual regulatory instrument is exactly as the regulator intended.

To underline the parallel between regulations and algorithms, consider the problem of sorting a vector. The input is a real vector x and the output z includes the sorted vector and other metrics such as computation time. The programmer has preferences over these metrics. The programmer may also give different weights $\nu(x)$ to various inputs, such as "worst-case", "average", and "best-case" scenarios. The first step in solving this problem is to choose an algorithm φ, such as "InsertionSort" or "QuickSort", each offering a distinct method. The second step is to code an actual program F that implements this algorithm. For a given input x, a computer then interprets the code into $I(F)$, executes the computations, and outputs the sorted vector $\varphi(x)$, along with other performance metrics in z.

In summary, we conceptualize regulation as an "algorithm" that uses economic variables as inputs, applies rules to these inputs, and produces a regulatory action as output. The regulatory text functions like a program describing this algorithm, and applying the regulation to a regulated entity is akin to executing the program.

1.2 Dimensions of regulatory complexity

We define three dimensions of complexity based on the costs and errors made at each stage in the regulatory process.

Drafting: We call *intrinsic complexity* properties of the regulation φ that make the regulator less likely to draft a text with an accurate interpretation of φ.

Definition 1. *For two regulations* $(\varphi, \varphi') \in \Phi^2$*, associated regulatory texts* $\tilde{F}$ *and* $\tilde{F}'$*, and a regulator with given* (θ, e)*, we say that* φ' *has a higher intrinsic complexity than* φ *if* $\Pr(I(\tilde{F}') = \varphi') < \Pr(I(\tilde{F}) = \varphi)$.

In our approach, intrinsic complexity is relative to the individual or institution drafting the regulation. We see this as a desirable property. For example, consider two systemic risk regulations: (i) a systemic risk tax on banks, proportional to the negative externality they impose, as measured, e.g., by SRISK (Brownlees and Engle (2016)); (ii) the FSB's approach, which aggregates multiple criteria into a synthetic index, ranks banks, and imposes a capital add-on based on a bank's rank. Financial economists may find regulation (i) easier to draft, while regulators with a background in accounting or law might find regulation (ii) simpler.

Interpretation: We call *psychological complexity* properties of the regulatory text F that make interpretation difficult, in the sense that $\tilde{f}$ is more likely to be incorrect.

Definition 2. *For two regulatory texts* $(F, F') \in \bar{\mathcal{V}}^2$*, the associated interpretations* $\tilde{f}$ *and* $\tilde{f}'$*, and an interpreter with given* (θ, e)*, we say that* F' *has a higher psychological complexity than* F *if* $\Pr(\tilde{f}' = I(\tilde{F}')) < \Pr(\tilde{f} = I(\tilde{F}))$.

Compliance: We call *computational complexity* properties of the regulatory text F that make arriving at the correct result more difficult.

Definition 3. *For two regulatory texts* $(F, F') \in \bar{\mathcal{V}}^2$*, a given state* $x \in X$*, the associated regulatory instruments* $\tilde{y}$ *and* $\tilde{y}'$*, and a supervisor with given* (θ, e)*, we say that* F' *has a higher computational complexity than* F *if* $\Pr(\tilde{y}' = I(\tilde{F}')(x)) < \Pr(\tilde{y} = I(\tilde{F})(x))$.

Note that computational complexity depends on the state of the world x (e.g., bank characteristics). Imagine for instance a regulation that exempts small banks from most rules. It could be the case that the regulatory text is complex, that applying it to large banks is costly, but that applying it to small banks is simple.

We can again draw a comparison with algorithmic complexity.[10] To solve a given problem, some algorithms are intrinsically simpler to program without errors. For example, "InsertionSort" likely has lower intrinsic complexity than "QuickSort". There are also many

[10]We take the terms "intrinsic complexity", "psychological complexity" (e.g., Zuse (1990)), and "computational complexity" from that literature and give them a similar meaning. To our knowledge, a formal framework giving a definition of these three terms does not exist in that literature.

ways to code a given algorithm (e.g., using different languages, loops, or modularity), and some choices make the program easier to understand, reducing its psychological complexity. Finally, algorithms differ in computational complexity, often measured by the time required to execute the program.

[Insert Fig. 1 here.]

Figure 1 summarizes our model of the regulatory process and which mistakes can occur. Observe that mistakes are cumulative. A complex problem may lead to an incorrect draft, which in turn may be wrongly interpreted, leading to mistakes at the compliance stage. It is important to develop empirical measures that can separate these different dimensions, as they have different theoretical properties and implications.

Indeed, a growing number of recent theory papers study the causes and the consequences of regulatory complexity. Hakenes and Schnabel (2012) develop a model of "capture by sophistication" in which some regulators cannot understand complex arguments and "rubber-stamp" claims from the industry so as not to reveal their lack of sophistication. In Asriyan *et al.* (2023), a policymaker proposes a regulation to, e.g., a parliament. Making the regulation more complex makes the regulation more complicated to study, so that members of parliament will rely more on their prior regarding the regulator's competence and less on their own understanding. In both papers, the focus is on psychological complexity.

Oehmke and Zawadowski (2019) develop a model in which regulatory complexity is in itself desirable (e.g., it allows for more risk-sensitivity), but regulators neglect that a more complex regulation consumes the limited attention of agents, and crowds out other activities. In our framework, this would be a trade-off between intrinsic complexity and psychological complexity. Foarta and Morelli (2022) model the dynamics of regulatory complexity. In their model, more complex regulations imply both higher implementation costs (computational complexity) and more noisy information (psychological complexity).

Note that we focus on measuring the complexity of existing regulations, not on understanding why a complex regulation was chosen in the first place. In reality, complexity is the result of a choice by the regulator, who may for instance make regulation more detailed and complex for dimensions that are more important. Similarly, we treat all mistakes by regulated entities equally but in reality some may be more consequential than others, and regulated agents will allocate more attention to those. Gabaix and Graeber (2024) propose a general model of economic decision-making with complexity to formalize this type of trade-off. We go in this direction in Section 4.1 in the context of a model of banking regulation, where we can model the environment faced by the regulator when solving (1). In our empirical analyses we do not observe the underlying problem the regulator was trying to solve, and hence we just take the existing regulatory text F and the underlying regulation φ as given. Section 3.2 gives some evidence on how measures of complexity correlate with the importance of a rule for the regulator.[11]

1.3 Measures of regulatory complexity

There are two ways to measure regulatory complexity: the *text-based approach* and the *outcome-based approach*. These methods are complementary: the text-based approach measures properties of regulatory texts that cause complexity, whereas the outcome-based ap-

[11] Some papers discuss the complexity of the problem (1) faced by the regulator in the context of banking regulation. For instance, Gai *et al.* (2011) and Haldane and Madouros (2012) discuss the difficulties of regulation in complex economic systems, which relates to the probability distribution μ. The complexity of the banking industry (X in our model) is also studied for instance by Cetorelli *et al.* (2014).

proach measures costs that are consequences of this complexity.

A comprehensive research program on the measurement of regulatory complexity must include text-based measures, but these should correlate with the consequences measured by the outcome-based approach. In what follows, we define several text-based measures based on modeling regulation as an algorithm. We then test whether they correlate with outcome-based measures in Section 3.

Measures derived in the Halstead framework. We derive several text-based measures of regulatory complexity by modeling a regulation like an algorithm, adapting Halstead (1977). We consider a regulatory text F as a sequence of words $F = \{w_1, w_2...w_N\}$, which we can classify into "operands", "operators", and "others". Using Halstead's definition, operands in a program are "variables or constants" and operators are "symbols or combinations of symbols that affect the value or ordering of an operand". Consider, for instance, the following "pseudo-code" to compute the vector norm of an n-dimensional vector $x = (x_1, x_2...x_n)$:

```
y = sqrt(x_1^2+x_2^2...+x_n^2)
```
(5)

Here, the operators are $=, sqrt, +, \hat{}$, and the operands $y, x_i, 2$. We extract from F a sequence of N_{OR} operators and a sequence of N_{OD} operands, the remaining words having no algorithmic value (e.g., function words). The sets $\{o_1, o_2...o_{\eta_{OR}}\}$ and $\{\omega_1, \omega_2...\omega_{\eta_{OD}}\}$ are the sets of all operators and operands that appear in F, where η_{OR} is the total number of unique operators, and η_{OD} the total number of unique operands. In the example above, we have $\eta_{OR} = 4$, $N_{OR} = 2n+1$, $\eta_{OD} = n+2$, $N_{OD} = 2n+2$.

To better take into account some differences between regulatory texts and computer programs, we propose a slightly finer partition than Halstead's. Already in Halstead's work, the assignment operator (the = sign in (5)) plays a different role from other operators. Similarly, a regulation will necessarily contain words that indicate a rule, an obligation, a permission, etc. We call such words "regulatory operators". We also define "logical operators" that represent logical operations such as "if", "when", etc., and "mathematical operators" that represent operations like addition, product, subtraction, and so on. We denote $N_R, \eta_R, N_L, \eta_L, N_M, \eta_M$ the number of total/unique regulatory operators, total/unique logical operators, total/unique mathematical operators, respectively. We have $N_R + N_L + N_M = N_{OR}$ and $\eta_R + \eta_L + \eta_M = \eta_{OR}$.

We can encompass three already existing measures of complexity in this framework:

- Length: The total number of words $length = N$ in a regulation, as used for instance in Haldane and Madouros (2012).

- Cyclomatic complexity: The total number of conditional statements in a code (McCabe, 1976), which can be computed as $cyclomatic = N_L$, the total number of logical operators (as in, e.g., Li *et al.* (2015)).

- Quantity of regulations: The total number of regulatory operators $quantity = N_R$. This is similar to the RegData measure of Al-Ubaydli and McLaughlin (2017), who count the number of words indicating a binding constraint in the US Code of Federal Regulations.[12] We propose to interpret *quantity* as a measure of intrinsic complexity. In principle, one could replace two rules by a single rule and the logical operator "and", so that *quantity* depends on the way a text is written. However, we think it is difficult in practice to significantly

[12]See also McLaughlin *et al.* (2021) for a recent study using this measure. A related example is Herring (2018), who measures complexity through the number of different capital ratios Global Systemically Important Banks need to comply with.

reduce or increase the number of rules in a text while keeping the intrinsic content the same.

Next, we derive three measures inspired by Halstead (1977) and which are new to the literature on regulatory complexity:

- Potential volume: Halstead (1977) proposes to consider the shortest possible program that can solve a given problem, in the best possible programming language. Defining this program is easy. In the example of computing the vector norm, it is:

```
y = vecnorm(x_1,x_2...x_n),
```
(6)

where `vecnorm` is a function returning the vector norm, which has to exist in the "best programming language" to solve this problem. This is the shortest possible program because any program to compute the norm of a vector would need to specify the input, the output, an assignment rule, and an operation.

More generally, for any problem, the shortest program would still contain a minimum number of operands η^*_{OD} that represent the number of inputs and outputs of the program. All the operations transforming the inputs into outputs would already be part of the language as a single built-in function. The number of operators is then $\eta^*_{OR} = 2$. If one assumes that the list of inputs and outputs never includes some unnecessary ones, then we also have $\eta^*_{OD} = \eta_{OD}$. Halstead (1977) calls the length of this minimal program "potential volume". We denote it $potential = 2 + \eta_{OD}$. It does not depend on the way the program is written and can thus capture intrinsic complexity.[13]

[13]In Halstead (1977) the notion of "volume" is more involved than mere length, due to an attempt at having a formula that would be valid in any computer language. As this seems artificial in the context of regulation, we simplify this aspect of Halstead's analysis.

- Level: It is natural to normalize *potential* by the length, a measure Halstead (1977) calls "level", which gives $level = potential/length$. The level measures whether a program is close to the shortest one possible. This measure has an intuitive interpretation in the context of regulatory complexity. If *level* is high (close to 1) this means that the regulation has a very specific vocabulary, probably opaque to outsiders. Conversely, a low value means that the regulation starts from elementary concepts and operations. The impact of level is ambiguous: a high-level regulation may be easier to process for experts, but more difficult for others.

- Operator diversity: By symmetry with *potential* we propose to also consider the number of unique operators $diversity = \eta_{OR}$, as a measure of psychological complexity. Intuitively, there might be increasing returns to scale in always processing the same operations, whereas a regulation that describes many distinct operations could be more difficult to understand.

In the next sections, we will often compare regulatory texts of very different sizes. To better apprehend the heterogeneity across these texts, we will sometimes report our measures scaled by length: $\overline{cyclomatic} = cyclomatic/length$, $\overline{quantity} = quantity/length$, and $\overline{diversity} = diversity/length$. We keep the notation *level* for $potential/length$.

Review of other measures. For completeness, we briefly review other measures that have been used in the literature but do not directly fit within our framework.

A number of other text-based measures have been used that do not rely on the algorithmic value of words in a text. Kalmenovitz *et al.* (2024) propose a measure of regulatory fragmentation, *RegFragmentation*, which relies the number of different regulatory agencies mentioned in the Federal Register on a given topic. This measure is best interpreted as a measure of computational complexity, the idea being that the overlap between different authorities makes compliance more costly.

Amadxarif *et al.* (2019) use measures from the linguistics literature, in particular *average word length*, the Maas' index of *lexical diversity* (Maas, 1972), and the Flesch-Kincaid *readability metric* (Kincaid *et al.*, 1975). Katz and Bommarito (2014) and Li *et al.* (2015) also use *Shannon's entropy*. All these measures do not rely on a partition of words between operands and operators. They aim at capturing the complexity of the style used by an author, which is part of psychological complexity.

Boulet *et al.* (2011), Katz and Bommarito (2014), Li *et al.* (2015), and Amadxarif *et al.* (2019) propose to analyze the network formed by different legal texts or regulations that reference each other. Network measures such as the *in-degree*, *out-degree*, or different *network centralities* can then be interpreted as measures of psychological complexity. These network-based measures of complexity are quite different from our approach because they are based on references between different legal texts in a corpus.

Finally, there is a growing literature that proposes outcome-based measures of regulatory complexity. For instance, Trebbi *et al.* (2023) build a *Regulation Index* based on the proportion of regulation-related employees in different sectors. Kalmenovitz (2023) proposes four *RegIn* indices of regulatory intensity, based on the number of forms required by Federal regulatory agencies in the US, the number of completed forms they receive, and the associated time costs and dollar costs. Calomiris *et al.* (2020) propose to measure the cost of regulation to US firms by *NetReg*, a measure based on the mention of regulatory topics in transcripts of earnings calls. Singla (2023) uses estimates of regulatory costs provided by US regulatory agencies themselves at the industry level.

Table 1 summarizes the different measures surveyed in this section. The table also serves to illustrate how different measures can be classified according to the dimension of complexity they capture, following Section 1.2.

[Insert Table 1 here.]

2 Building the Measures

The Halstead measures we propose to use were initially designed for algorithms, in which the classification of elements into operands and operators is unambiguous. In this section, we show that it is possible to meaningfully adapt these measures to regulatory texts. We start with the example of the computation of capital requirements under the 1988 Basel Accords. This is a natural starting point as this regulation is very close to describing an algorithm. We then extend our approach to two important and complementary texts, namely the US Dodd-Frank Act and the 2021 version of the European Banking Authority's "Implementing Technical Standards on Supervisory Reporting".

2.1 Basel I as an algorithm

The Basel I Accords are a 30-page long text specifying how to compute a bank's capital ratio. We focus here on Annex 2, "Risk weights by category of on-balance-sheet asset". This Annex maps different asset classes to risk buckets, and different capital instruments to weights. The regulation then compares the risk-weighted sum of assets to the weighted sum of capital, and the ratio has to be higher than 8%.

Annex 2 is easily described as an algorithm. Our idea is to first apply the Halstead measures described above literally, by writing a "pseudo-code" that implements the Basel I computation of risk-weighted assets. We then compute the Halstead measures on the original text instead, by classifying the words into operands and operators, and verify that we can classify a large fraction of words. Finally, we compare the two approaches and verify that measures of intrinsic complexity (which should not depend on the language used, code or plain English) are indeed similar under the two approaches.

We give the full pseudo-code of the Basel I risk weights in Online Appendix OA.1. To give an example, Basel I specifies a risk weight of 20% for "Claims on banks incorporated in the OECD". In our code this is translated into:

```
IF (ASSET_CLASS == "claims" AND ISSUER == "bank" AND ISSUER_COUNTRY == "oecd") THEN:
    risk_weight = 0.2;
```

The operands are the different asset characteristics (e.g., `ASSET_CLASS`, `ISSUER_COUNTRY`), their values (e.g., `oecd`), and risk-weights (e.g., `risk_weight`, `0.2`). The logical operators are `IF`, `AND`, `THEN`, and we distinguish between the mathematical operator `==` and the regulatory operator `=`. We thus obtain $\eta_{OD} = N_{OD} = 8$, $\eta_R = N_R = 1$, $\eta_L = 3$, $N_L = 4$, $\eta_M = 1$, $N_M = 3$. We can then apply the formulas derived in Section 1.3 and compute the complexity measures. We conduct the same exercise for the 19 items covered by Basel I, and report all the measures in Online Appendix OA.2.

We then repeat the analysis, relying this time on the actual text of Basel I's Appendix 2. A drawback of this text is that some words are left implicit. In particular, the mapping between different asset classes and their respective risk weights is only indicated by the layout of the page. To circumvent this issue, we wrote a more explicit text in which each item ends with "shall have an x% risk weight". This is the only modification we made.[14]

[14] We report this modified text in Online Appendix OA.1.

We then classify as "operands" the words or word combinations that have the same function as operands in the program. These are economic entities (e.g., "bank" or "OECD"), characteristics (e.g., "maturity" or "counterparty"), and values (e.g., "one year"). We classify as regulatory operators words that indicate an obligation or regulatory requirement, which here is "shall have". Logical operators are words that correspond to logical operations, such as "and" or "excluding". Mathematical operators are for instance "up to" and "above".

Using this approach, we classify 80 unique words or expressions out of the 95 used in the text. The remaining words are used for grammatical reasons and do not really correspond to operands or operators (e.g., "by", "on", "the"), hence we do not take them into account. In the Online Appendix OA.2 we report the most frequent words in each category, as well as the measures we compute for each item of Basel I.

We can now compare the measures derived under the two versions of the regulation, the algorithm and the actual text. Going back to the example above, its text version is:

Claims on banks incorporated in the OECD shall have a 20% risk weight.

There are $\eta_{OD} = 6$ unique operands ("claims", "banks", "incorporated", "OECD", "20%", and "risk weight") and $\eta_R = 1$ unique regulatory operator ("shall have"). The algorithmic version and the text version have the same *quantity*: one rule in each case. Their measures of *potential* are close but not exactly the same. The text version has no equivalent for the operands "ASSET_CLASS" and "ISSUER": these operands have to be explicit in the algorithm, but are implicit in the text. Similarly, the measures *cyclomatic* and *diversity* are different because the algorithmic version uses logical operators that are implicit in the text.

These observations generalize to the other 18 items of Basel I. We compute the correlations between the text-based measures and their algorithm-based counterparts and report the results in Table 2.[15] The correlation is perfect for *quantity*, high for *potential* (0.82), and medium (between 0.40 and 0.68) for *length*, *cyclomatic*, *diversity*, and *level*.

These results correspond well to the theoretical properties of the different measures. On the one hand, *length*, *cyclomatic*, *diversity*, and *level* are measures of psychological complexity and should depend on the way a regulation is written. The algorithm is more explicit than the text and thus has a different psychological complexity, which is what our measures capture. In particular, *level* is indeed lower in the algorithmic version. On the other hand, *quantity* and *potential* are measures of intrinsic complexity and should in theory not depend on the way the regulation is written. This is correct exactly for *quantity*, and correct up to a few words left implicit in the text for *potential*.

We conclude from this exercise that most words in a regulatory text can be classified into operands and operators, so that it is possible to compute the Halstead measures on a text. Moreover, *quantity* and *potential*, which are meant to capture intrinsic complexity, seem to indeed be relatively insensitive to how the regulation is written. Both observations support using the text-based approach for a full-scale regulatory text in the next sections.

[Insert Table 2 here.]

[15] Since *quantity* is constant in both the algorithm version and the text version we adopt the convention that the correlation coefficients are equal to 1.

2.2 Building a dictionary: The Dodd-Frank Act

We now compute our complexity measures for the 2010 Dodd-Frank Act (DFA). There are two reasons for this choice. First, the DFA is a long text (848 pages) that touches upon a wide range of issues in finance. A dictionary of operands and operators based on the DFA thus covers many financial regulation topics. Second, as the DFA is one of the key regulations introduced after the financial crisis, it has triggered a lot of debate, in particular regarding the costs associated with its complexity (e.g., Alvero *et al.* (2022), Hogan and Burns (2019)). As a result, several papers have computed different measures on the text of the DFA (in particular, Li *et al.* (2015) and McLaughlin *et al.* (2021)), and analyzing the same text makes it possible to compare different approaches.

The scale and scope of the DFA creates four new challenges compared to the limited example of Basel I:

First, a lot of operands in the DFA are "n-grams", expressions made of n distinct words. For instance, "Consumer Financial Protection Bureau" should be considered as one expression, not four distinct words. To take this into account, we made a list of all n-grams in the DFA (for details see the Online Appendix OA.3). We classified each n-gram into a category, and then removed it from further counts. That is, we made sure that "Consumer Financial Protection Bureau" is counted only once, as an operand.

Second, some words may appear several times with "variants", e.g., singular and plural of the same noun, or different forms of the same verb, for instance "disapprove", "disapproved", and "disapproving". When such a case arises, we manually record that these are variants of a unique "root" word, and when computing the number of unique words in a given category we only count the root word. We choose to not consider synonyms as variants, as the use of different words could be intentional (and, in any case, will be more complex for the reader than if a single word were used). Similarly, we do not count initials and abbreviations as variants, as the reader will have to keep track of what the initials stand for.

Third, some words in the text can be used sometimes as an operand and sometimes as an operator. The most prominent example is the word “is”. In principle, “is” could be a regulatory operator (e.g., “the risk-weight is 20%”). However, it could have a merely grammatical function to indicate the passive voice (e.g., “each report is submitted”). We classify such ambiguous words in the category “other”, and hence do not count them in our different measures.[16]

Fourth, the DFA uses a lot of external references. As an example, Section 201 (5) reads “The term “company” has the same meaning as in section 2(b) of the Bank Holding Company Act of 1956 (12 U.S.C. 1841(b)) [...]” A possible approach would be to include the text referenced in the example as being implicitly part of the DFA. However, with such an approach we would quickly run into the “dictionary paradox” (every reference refers to other texts). Instead, and more consistent with the Halstead approach, we consider that if a legal reference is mentioned it is part of the “vocabulary” one has to master in order to read the Act, similar to a program calling a pre-programmed function. The role of legal references is ambiguous, they are sometimes used as operators and sometimes as operands. Thus, we include them in the “other” category.

These difficulties required us to classify the words manually. To this end, we developed

[16] There is necessarily some judgment involved in this decision. One could consider other possibilities, such as estimating the fraction of occurrences in which “is” is a regulatory operator, an operand, etc., but we believe these estimates would not necessarily carry over to other regulatory texts, thus running against the objective of building a reusable dictionary.

a "dashboard" which displays the text of the Dodd-Frank Act in a browser and allows us to highlight every n-gram we identify manually and categorize it as an economic operand, regulatory operator, etc. Once we manually identify an n-gram, the dashboard will find it in the entire DFA and classify it accordingly. We provide more details on the dashboard in the Online Appendix OA.3. By classifying words in the 16 Titles of the DFA plus its introduction, we obtain a dictionary containing: 608 operators (276 logical operators, 221 regulatory operators, 111 mathematical operators), 5627 operands, as well as 2077 "other" words (1719 legal references, 189 function words, and 169 ambiguous words).[17] Table 3 shows the top 10 words in each category as well as the number of occurrences. We then compute our measures for the different titles of the DFA. Figure 2 shows a histogram of the distribution of the different measures across titles, and Table OA.7 in the Online Appendix reports the measures for each title.

[Insert Table 3 and Figure 2 here.]

2.3 Expanded dictionary: The Implementing Technical Standards

To validate our measures, in Section 3.2 we correlate them with the responses given by banks to a survey conducted by the European Banking Authority on the complexity of its "Implementing Technical Standards" (ITS). The ITS collect instructions on how all EU banks need to report information to their supervision authority in order to be compliant with the EU implementation of Basel III. Simply put, it is the most important text on banking

[17] We check in particular that our list of regulatory operators, used to compute *quantity*, includes the words used in Al-Ubaydli and McLaughlin (2017) and Calomiris *et al.* (2020), and that our list of logical operators includes the words used in Li *et al.* (2015) to compute cyclomatic complexity.

supervision in the EU.[18]

The ITS are organized in 87 different templates identified by a letter and a number.[19] "C03" is, for instance, the template for reporting capital ratios. Each template consists of a set of tables with empty cells indicating the information that banks have to report, and then additional explanations in an appendix. We group each template and its appendix together. Some of the templates are made of different subtemplates, such as C32.01, C32.02, C32.03, and C32.04. Finally, the survey that we use in Section 3.2 regroups some of these templates by topic. For example, "C12, C13, C14 - Credit risk: securitisations". In order to use the information from the survey, we compute our measures of complexity at the level of each template or group of templates covered in the survey. This gives us a cross section of 51 templates or groups of templates.[20]

As in Annex 2 of the Basel I Accords studied in Section 2.1, the main regulatory operators in the ITS are implicit: each empty cell is, in fact, an indication that banks should report a certain number. Hence, we first count the number of empty cells in each template and classify them as regulatory operators.[21] Then, we identify all the n-grams and words in the text. When they are not already present in the DFA dictionary, we classify them as operands and operators as before and add them to the dictionary. Of the 3083 operands and operators we identify in the ITS, 1867 or 61% were already in the DFA.

[18]More precisely, we consider the ITS as collected in the "Commission Implementing Regulation (EU) 2021/451" of 17 December 2020, published in the Official Journal of the European Union on 19 March 2021. This text, which is 1955 pages long, is a finalized version of the ITS for the application of Regulation (EU) No 575/2013, better known as the "Capital Requirements Regulation" or CRR.

[19]Templates start either with "C" for "Common Reporting", focused on capital, or "F" for "Financial Reporting", focused on financial variables.

[20]We exclude a few templates that are mentioned in the survey but do not appear in the version of the ITS we use. These are: C12 to C14, C30 and C31, C41 to C43, C60 and C61, F90 to F93.

[21]We do not count empty cells filled in gray or in black, as these colors indicate they do not have to be filled. We also do not count a small number of cells in columns marked "References" (accounting or legal) or "Breakdown in table" because we think these do not have to be filled either.

Using this updated dictionary, we compute our measures on the ITS as we did for the DFA. Figure 2 shows the histograms of the distributions of the different measures in the 51 templates, and Table OA.8 shows the measures for each template. We observe significant heterogeneity for all measures, with a significant right tail. In addition, the distribution of $\overline{quantity}$ is heavily shifted to the right in the ITS relative to the DFA, which is due to the ITS being reporting templates with a large amount of numbers to report and few explanations per reporting obligation. In contrast, *level* is shifted to the left, indicating that the ITS is more detailed and repetitive than the DFA.[22]

3 Testing the Measures

We can now test which of our text-based measures correlate with costs of complexity. The first test relies on experiments based on randomly generated regulations. The second test uses a survey conducted by the EBA on the compliance costs associated with the ITS templates. These two approaches are complementary. The experiments have the advantage of using artificial regulations that vary in complexity in an exogenous way. The advantage of the survey is that it relies on professional bankers working on compliance. At the end of this section we discuss how both approaches validate two of our measures in particular, and what they tell us about how banks perceive complexity.

[22] There are 42,814 empty cells in total in our sample, which may seem a very high number. An important function of the ITS is to ensure that different national authorities will enforce homogeneous standards across the EU, which requires a level of detail which may not be necessary in the US. A more skeptical view would also be that lobbying by different countries in favor of their "national champions" leads to more exceptions and special rules (Haselmann *et al.*, 2022), which then translates into more detailed reporting.

3.1 Experimental validation

The idea of using experiments to validate complexity measures originates in computer science. In that field, complexity measures are tested by asking different programmers to write the same code. One then checks whether the mistakes they make or the time they take to perform the task are correlated with a measure of algorithmic complexity. We apply this idea in the context of banking regulation.

Design. An ideal experiment requires a sample of regulations and rules with cross-sectional variation in complexity. Ideally, this variation should be exogenous, not correlated with unobservable variables or contaminated by the priors of the experimenter. The solution we propose is to generate a sample of random regulatory texts: we start with the Basel I rules studied in Section 2.1, and generate random variations on these rules. We obtain a number of artificial "Basel-I like" instructions to compute risk-weighted assets, where the instructions vary in complexity.

For instance, in our algorithmic version of Basel I (Online Appendix OA.1), the regulatory text *"Cash items in process of collection shall have a 20% risk weight"* translates into a conditional statement of the type `IF-X-AND-Y-THEN` with two conditions `X: ASSET_CLASS == "cash"` and `CASH_COLLECTION == "in progress"`. A random variation could for instance consist in changing the value of `ASSET_CLASS` to `"loans"`, and add new attributes such as `ISSUER`, `DENOMINATION`, and so on. Each randomly generated regulation consists of these building blocks, which are connected using a random number of `AND` and `OR` statements, at most 6 (the largest number of conditions in any `IF-THEN` clause in Annex 2).

As a last step, we manually check that the instructions make sense, i.e., they are complete and do not contain contradictory rules, and we make some minor manual changes to avoid ambiguities, grammar mistakes, etc.[23] At the end of this process, we obtain 38 regulations. As shown in Table 4 below, there is significant variation in all the complexity measures across the different regulations.[24]

[Insert Table 4 here.]

In order to find participants able to read the regulations and compute regulatory quantities, we asked the students of the MSc in International Finance of HEC Paris, class of 2020-2021, to volunteer for taking part in the experiment. Given the sanitary situation in early 2021, our experiment was conducted online.[25] After a registration page, an introductory page, and a test round to ensure that he/she understands the exercise, the participant starts the experiment. The main page is reproduced in Figure 3. The computer screen is split vertically in two. On the right-hand side, a series of instructions correspond to a random selection of one of our 38 artificially generated regulations. On the left-hand side, there is a simplified bank balance sheet with details about the different assets of a bank. The participant has to compute the risk-weighted assets of the bank following the instructions. We record the answer given by the participant (and hence whether it is correct), as well as the time taken to answer. The participant is then given a second randomly selected regulation, and continues until he or she has given 9 answers (correct or not). We detail in the Online

[23] Figure OA.2 shows the possible attributes for each asset class and the values these attributes can take, and the Appendix A.3 shows an example of such a randomly generated regulation. In addition, the replication files of this paper, available at https://cogeorg.github.io/MeasuringRegulatoryComplexity_Replication.zip, give: (i) the program used to generate the random regulations; (ii) the raw regulations generated by the program; (iii) the final regulations we used in the experiment.

[24] All measures are computed based on the actual texts seen by the participants to the experiment, not on the underlying code.

[25] The interested reader can try the experiment anonymously at https://regulatorycomplexity.org/ with the login "test_account" and password "test".

Appendix OA.5 how we selected and incentivized the students, and reproduce all the pages of the website.

[Insert Fig. 3 here.]

Mistakes. The answers form a balanced panel with 118 participants, indexed by i, answering 9 rounds each, indexed by t, for a total of 1,062 participant-round observations. The t-th round for each participant corresponds to regulation $R_{i,t} \in \{1...38\}$, which is randomly drawn from our 38 randomly generated regulations. For a given participant the 9 regulations are independently drawn without replacement. The draws are also independent across students. We denote $mistake_{i,t}$ a dummy variable equal to 1 if the participant i's answer to question t is incorrect. Following Definition 3, $mistake_{i,t}$ is an outcome-based measure of computational complexity for participant i and regulation $R_{i,t}$. There is substantial variation across regulations. Denoting $mistake_j$ the proportion of wrong answers for regulation $j \in \{1, ..., 38\}$, across all regulations the average of $mistake_j$ is 31.39%, the standard deviation is 16.78%, two regulations have the minimum of 0, the first quartile is 17.86%, median 35.32%, third quartile 41.38%, and the maximum 62.50%.

Denoting $\mu(R_{i,t})$ a text-based complexity measure for regulation $R_{i,t}$, in order to test whether μ is a useful measure we study its power to explain the variation in $mistake$. First, we estimate the probit model (7) below, at the participant-round level, using both participant and round fixed effects. $\Phi(.)$ denotes the cdf of the standardized normal distribution.[26]

$$\Pr(mistake_{i,t} = 1) = \Phi(\alpha + \beta\mu(R_{i,t}) + \gamma_i + \delta_t). \tag{7}$$

[26] The round fixed effect captures that participants may become better at this exercise after the first few rounds. The presence of this effect is in principle not an issue since regulations are randomized, and indeed not including this fixed effect does not alter our results (see Table OA.12 in the Online Appendix).

To test the text-based measures, we run the probit regression (7) on *length*, and then on *cyclomatic*, *quantity*, *potential*, *diversity*, and *level* separately, using *length* as a control variable. Since *length* is a natural measure of complexity, another measure is useful only if it can explain outcomes beyond *length*. Hence, we expect a good measure of complexity to be significantly positively associated with *mistake* even after controlling for *length*.

[Insert Table 5 here.]

We observe that *quantity* is the only measure that has a significant positive correlation with *mistake* once controlling for *length*.[27] In particular, controlling for *length*, *cyclomatic* and *diversity* are significantly negatively correlated with mistakes. Both capture the presence of conditions and logical statements in the different regulations. Many of these conditions actually simplify the computations: the more conditions an asset has to satisfy to have a certain risk weight, the more likely it is that the conditions are not met and that the participant does not have to perform the computation. We believe this is actually an interesting result that also applies in real world situations. A typical example in the real world would be exemptions for banks under a certain asset size. While such exemptions make understanding the entire regulation harder, they also clearly reduce the implementation costs for the exempted banks.

The measure *quantity* stands out in Table 5. Not only is it significantly positively correlated with *mistake* when controlling for *length*, but actually *length* loses any statistical significance. Note that while *cyclomatic* and *diversity* are measures of psychological complexity like *length*, *quantity* and *potential* are measures of intrinsic complexity. We expected

[27] Table OA.14 in the Online Appendix shows that all measures except level are significantly and positively correlated with *mistake* if one does not control for *length*.

these measures to potentially capture a dimension not already reflected in *length*, and we indeed find this is the case for *quantity*.

Time taken. For each participant i and round t, we record the time taken to answer, denoted $time_{i,t}$. We interpret *time* as the effort in Definition 3. Two issues complicate the measurement: (i) A participant may take an abnormally short time to answer because he or she misunderstood the regulation, or simply gave up, so that wrong answers given after a short time may not reflect a low complexity; (ii) A few correct answers were given after a long time (the maximum being 958 seconds, or about 16 minutes). These participants likely got distracted while completing the online experiment, in which case the actual effort exerted may be vastly overestimated.

To address these issues, we restrict the sample to answers that are the least likely to be affected. Starting with 1,062 observations, we keep only those 728 that correspond to correct answers, which solves (i). We then delete observations for which $time_{i,t}$ is above the 99th percentile of the initial distribution, which is 579 seconds (6 observations), alleviating (ii). There is still significant variation in $time_{i,t}$ in this restricted sample: the average is 132 seconds, the standard deviation 98, minimum 6, first quartile 59, median 107, third quartile 180, and maximum 561 seconds. We then run an OLS regression of $time_{i,t}$ on different measures of complexity, with participant and round fixed effects:

$$time_{i,t} = \alpha + \beta\mu(R_{i,t}) + \gamma_i + \delta_t + \epsilon_{i,t}. \tag{8}$$

[Insert Table 6 here.]

Table 6 shows that, when controlling for *length*, both *quantity* and *potential* are significantly and positively correlated with *time*. These are the two measures of intrinsic complexity, which were expected to potentially capture a different dimension from *length*.[28]

3.2 Empirical validation

Banker survey. To also test our measures in a real-world application, we use a survey conducted by the EBA on the cost of complying with the ITS (EBA (2021)). The EBA asked all banks in the European Economic Area (EEA) to complete a questionnaire on their reporting costs and the complexity of the reporting templates (the survey explicitly mentions complexity). They received responses from a total of 251 credit institutions, among which 39 are "large institutions", 49 "medium institutions", and 163 "small and non-complex institutions (SNCI)".[29]

We are mainly interested in the answers to Question 7, which reads: *"Please indicate, based on your experience, how costly the overall compliance with the reporting obligations of the ITS on Supervisory Reporting and the GLs on reporting on exposures subject to COVID-19 measures is."* Annex 5 of EBA (2021) reports, for each template or group of templates described in Section 2.3, the percentage of banks in each group (large, small, SNCI) that "classify the template or group of templates as associated with high or medium-high reporting cost". This gives us an outcome-based measure of complexity, in each group of banks.

We closely follow the approach of Section 3.1. An important difference is that the answers

[28]For robustness, Online Appendix OA.6 shows the results obtained when aggregating answers at the regulation level. Online Appendix OA.8 shows the results obtained with several other filtering choices. We obtain qualitatively similar results, with the exception that the impact of *quantity* becomes not significant at the 10% level in some specifications.

[29]Based on the answers, the study estimates the total ongoing reporting costs associated with the EBA templates at 13.6 bn EUR, to which one should add 6 bn EUR of implementation costs. Moreover, according to the study, these numbers can be compared to 389.8 bn EUR of total operational costs for EEA banks, so approximately 5%.

are aggregated at the group level, so our dependent variable $cost_{i,j}$ is the percentage of banks in group $i \in \{large, medium, small\}$ that report a high or medium-high cost for template j. We estimate the following model using OLS:

$$cost_{i,j} = \alpha + \beta\mu(R_j) + \gamma_2 medium_i + \gamma_3 small_i + \epsilon_{i,j}, \tag{9}$$

where $\mu(R_j)$ is a measure (or vector of measures) of regulatory complexity for template j, and $medium_i$ and $small_i$ are group-fixed effects (equal to 1 if group $i = medium$ or $i = small$, respectively). Thus, large banks serve as the baseline. Table 7 shows the results of the regressions following the specification (9), with the same regressors as in Table 5.

[Insert Table 7 here.]

In Table 7 we see that *quantity* is significantly positively correlated with *cost* after controlling for *length*, at the 1% level. *potential* is significant only at the 10% level. Interestingly, *quantity* was also the only variable to be significant in both Tables 5 and 6, so again the results in the experiment and in the empirical application are remarkably consistent.[30] We also check that the result on *quantity* is not only driven by empty cells, *n_cells*. To do this, we estimate (9) using as regressors ($length - quantity$), ($quantity - n\_cells$), and *n_cells*. We find that both ($quantity - n\_cells$) and *n_cells* are significant at the 1% level, and the coefficient on ($quantity - n\_cells$) is 5 times larger than the coefficient on *n_cells*. Note that ($quantity - n\_cells$) is not significant at the 10% level if we control for *length* but not for

[30] We report results from alternative specifications in the Online Appendix OA.9. In particular, a fractional probit regression. In fact, in principle, (9) is misspecified because the predicted value of *cost* could be below 0 or above 1. Indeed, for one template, we obtain a predicted value of 1.01 in two regressions. *quantity* becomes even more significant with the fractional probit specification, whereas *potential* becomes less significant (the p-value increases from 0.081 to 0.126).

n_cells.

Additional analysis. In addition to testing the validity of our measures, we can use this empirical setup and previous experiments to better understand how agents perceive complexity and test some hypotheses.

An important question is whether the economic costs of regulatory complexity stem from the difficulty of understanding the regulation or from the resources necessary to process it and perform the required computations. One question in the EBA survey suggests that for banks, especially small and medium ones, the cost of understanding regulation dominates.[31] The difficulty of understanding a regulation should be correlated with the occurrence of mistakes in Section 3.1 which, as we saw, is captured by *quantity*. Consequently, regulatory costs for banks are also mainly captured by *quantity* in Table 7. All these observations are consistent with the difficulty of understanding regulation as the main driver of regulatory costs for banks, and this difficulty is empirically captured by *quantity*.

A widespread view in banking is that smaller banks suffer disproportionately more from regulatory complexity, whereas larger ones may even benefit from it (e.g., through capture). In the EBA survey, medium and small banks are actually slightly *less* likely to report that a template has a high or medium cost.[32] This is probably due to medium and smaller banks being exempted from a number of reporting obligations, and is consistent with the evidence in Trebbi *et al.* (2023) (see also our discussion of exemptions and *cyclomatic* in Section 3.1).

[31] The EBA survey asks the respondents to name the main driver of compliance costs. Medium and small banks rank first (out of 37) the "Complexity of the underlying regulatory requirements", which the EBA questionnaire details as "The challenge to understand the concepts formulated in the underlying EU legislation, to understand how to apply them to the business transactions and to understand how to perform, for example, necessary calculations [...]". Large banks rank this driver third.

[32] The average of *cost* is 0.57 for large banks, 0.51 for medium banks, and 0.50 for SNCI.

If we extend the regressions of Table 7 to include complexity measure × group interaction terms, we observe that the effect of *quantity* comes from large and medium banks, not from the SNCI (see Table OA.26 in the Online Appendix). In contrast, SNCI are sensitive to *potential* (significance at the 1% level), whereas large and medium banks are not. This shows that large and small banks are not sensitive to the same forms of complexity. In this instance, a likely explanation is that the SNCI are exempted from many reporting obligations, presumably those associated with templates with a high *quantity*. Conversely, *potential* can be interpreted as a fixed cost which should indeed matter more for smaller institutions.

Regulator survey. EBA (2021) contains an additional survey with rare information on how regulators perceive the different templates. All authorities in the EEA were asked to what extent they consider different templates or groups of templates as important. Unfortunately, EBA (2021) only reports the results for the 34 templates identified as the most important and the 23 least important, with results not reported for an unknown number of intermediary templates. Of these 57 templates, we focus on the 43 that also appear in the bank questionnaire used in Section 3.2, and for which we computed the complexity measures. Despite these limitations, we think that this data can provide at least some suggestive evidence. We compute the measure $importance_j$ that reflects the average answer given by the 29 regulatory authorities surveyed.[33] For the 43 templates for which we have complexity measures, we can regress $importance_j$ on these measures.

The results are reported in Table 8. *potential* stands out in this table as the best measure

[33] We give a value of 3 to the answer "highly important", 2 to "important", 1 to "less important", and 0 to "not important", and then take the average. The results are very robust to alternative specifications, such as focusing on the answer "highly important" only, or simply comparing the 34 "most important" templates to the 23 "least important ones", see the Online Appendix OA.10.

to explain the variation in *importance*, suggesting that regulators ask for information about more variables when a topic is more important. *diversity* shows up significantly, but it is extremely correlated (> 90%) with *potential* in this particular sample. *cyclomatic* also shows up significantly (at the 10% level only), suggesting that regulators may introduce more detailed sets of conditions for more important topics. Finally, *quantity* is surprisingly (although weakly) negatively correlated with *importance*, but this result is driven by an outlier. Once it is removed, we obtain a non-significant coefficient for *quantity*.[34]

These observations support the idea (in line with Gabaix and Graeber (2024)) that more complex regulations are more complex for a reason, namely that they touch on issues that regulators deem more important. However, there seems to be also scope to make regulation more efficient by reducing *quantity*, which is the main driver of *cost* for banks, without being a significant driver of *importance*. In fact, EBA (2021) actually recommends reducing the number of values to report on a number of less important templates, to lower the cost to banks.

[Insert Table 8 here.]

3.3 Discussion

The first conclusion we draw from Sections 3.1 and 3.2 is a validation of two of our five text-based measures of complexity, namely *potential* and *quantity*. In fact, both can explain

[34] The outlier is template C70, in which banks have to report the roll-over of funding for each day of a one-month window, so that the same templates are repeated 31 times, generating a very large value for *quantity*. Interestingly, C70 is ranked n. 10 by large banks and n. 5 by medium banks in terms of costs, showing that this repetition is costly to banks, despite being of little interest to regulators. The result on *potential* is robust to removing C70 and more generally to removing up to the 7 (out of 43) templates with the largest values of *potential*. The result on *quantity* in Table 7 is robust to removing C70 and up to the 50% templates with the largest values of *quantity*.

the variation in *time* in the experiments, beyond the variable *length*. In addition, *quantity* can explain the variation in *mistakes* and *cost*, while *potential* can explain variation in *importance* and (to a lesser extent) *cost*. We think that the fact that the same two measures are picked up in four independent exercises and two independent datasets strongly supports the validity of these measures.[35] In these different exercises, *length* is already a very good proxy of complexity, and most measures do not have any extra explanatory power. This means that the type of regression we conduct is actually a powerful touchstone for testing novel measures of regulatory complexity, and it is unlikely that it will be a coincidence if the same two measures pass several independent tests on independent datasets.

Moreover, our results are consistent with the theoretical properties of these measures: Indeed, *potential* and *quantity* were expected to capture intrinsic complexity, whereas the other measures were expected to capture the same dimension as *length*, namely psychological complexity. As discussed in Section 3.2, it is consistent to find that *quantity* is correlated both with *mistakes* in the experiments and with *cost* in the survey, since the respondents to the survey declare *cost* to be primarily driven by the difficulty in understanding the regulation.

Having established that *potential* and *quantity* capture a different dimension of complexity, the second question is whether this different dimension is quantitatively meaningful. The answer depends on what we are using the measures for.

If one only needs an empirical proxy for regulatory complexity, all dimensions combined, then *length* on its own seems to be a very good proxy. This is because *length* is correlated

[35] Importantly, we conducted the analysis in Section 3.2 after the experiments, as evidenced by an earlier version of the paper that only contains the latter. In this sense, the results of *quantity* and *potential* in Section 3.2 are an external validation of the results of Section 3.1.

with many other measures and provides a good "catch-all" measure. As a result, adding *potential* or *quantity* to *length* in the regressions of Tables 5 to 7 leads to relatively small improvements in the adjusted-R^2, although it doubles in Table 8.

Another important use of complexity measures is normative: helping policymakers draft less costly regulations. A good proxy for regulatory complexity is then insufficient, as one needs to identify the drivers of regulatory complexity. Should a regulator with the aim of reducing complexity target *length*, *potential*, or *quantity*? To answer this question systematically, we revisit the regression results of Tables 5 to 8. For each dependent variable $y \in \{mistake, time, cost, importance\}$, we consider the regression of y over *length* and *quantity* or *potential*. Denoting $\hat{\beta}_\mu$ the estimated regression coefficient on complexity measure $\mu \in \{length, quantity, potential\}$, σ_y and σ_μ the sample standard deviations of y and μ, we calculate the standardized coefficient $\hat{\beta}_\mu \times \frac{\sigma_\mu}{\sigma_y}$. We then compute the p-value of the Student test of the null hypothesis $\beta_\mu = 0$. Finally, we use a variance decomposition to measure the contribution of *length* and the other measure to the R^2 of the regression.[36]

Table 9 summarizes the results. For each dependent variable y, we focus on the specification with the highest adjusted R^2, in bold in the table.[37] We find that, in this favorite

[36]Recall that $R^2 = ESS/TSS$, where ESS is the explained sum of squares and TSS the total sum of squares. If there are N observations and P regressors, $\hat{\beta}_p$ is the estimated coefficient on regressor p and $\sigma_{p,y}$ the sample covariance between this regressor and y, then we have $ESS = \sum_{p=1}^{P} ESS_p$, with $ESS_p = (N-1)\hat{\beta}_p \sigma_{p,y}$. We measure the contribution of regressor p to the R^2 by $\frac{1}{R^2}\frac{ESS_p}{TSS} = \frac{1}{R^2}\frac{\hat{\beta}_p \sigma_{p,y}}{\sigma_y^2}$. Differently from computing how much the R^2 improves by adding each variable to the regression separately, this decomposition takes into account the explanatory power of the variable while controlling for all the others. See, e.g., Hue *et al.* (2025) for a proof, as well as a discussion and a generalization beyond OLS regressions.

[37]Table 5 uses a probit model which is difficult to compare with the OLS models in Tables 6 to 8. For better comparison, we run a simple OLS model (a linear probability model) in Table OA.18, which gives very similar results to the probit. An additional complication is that the regressions in Tables OA.18 and 6 include participant and round fixed effects, and Table 7 bank size fixed effects, which mechanically decreases the contribution of complexity measures to the R^2. This is not an issue, as we are only interested in the relative contributions of *length* and *quantity* or *potential*.

specification (but not in others), *length* is never significant once controlling for *quantity* or *potential*, in line with the idea that *length* may be a proxy for complexity rather than its actual cause. The most striking result is the one on *cost*: a one-standard deviation increase in *quantity* leads to a 0.45-standard deviation increase in *cost*, with a p-value of 0.00, and *quantity* contributes to 89% of the R^2 of the regression explaining *cost*. In contrast, the effect of *length* is 0.00, with a p-value of 1.00, and a contribution to the R^2 of 0%. These results mean that a policymaker who would try to simplify regulation by targeting *length* but keeping the content constant (in particular *quantity* and *potential*) would presumably completely miss its objective.

4 Applying the Measures

Section 4.1 discusses how to use our measures in the context of a normative model of regulation, which policymakers could eventually use to quantify the trade-off between the complexity of regulation and other objectives. We identify that better linking text-based to outcome-based measures of complexity is a major gap in this research program. Section 4.2 then outlines how different streams of the empirical literature can fill this gap.

4.1 Towards "balancing risk-sensitivity and simplicity"

A natural ambition for a research program on regulatory complexity is to contribute to improving the production of regulation. In this section, we show how text-based measures, outcome-based measures, and economic modeling can be combined to formally estimate the regulator's trade-off between precision and complexity.

As seen in Section 1.1, a typical normative economic model assumes that the regulator has a utility function U over economic outcomes z, where the distribution μ of these outcomes depends on the state of the economy x and the regulatory instrument y. The regulator's program is then given by (1). To take into account the costs of regulatory complexity, one needs to subtract the total effort e (e.g., hours of work) spent drafting, understanding, and then implementing the regulation in each state of the world, which we denote $e^*(\varphi, x)$. We denote w the cost per unit of effort. We also need to account for the possibility of mistakes: in state x, the regulatory instrument will not necessarily be $\varphi(x)$, but instead will have a certain distribution denoted $\hat{\mu}(y, \varphi, x)$ due to the possibility of mistakes. This gives us the following program for the regulator:

$$\max_{\varphi(.)} \sum_{x \in X} \nu(x) \left[\sum_{\tilde{y} \in Y} \hat{\mu}(\tilde{y}, \varphi, x) \left(\sum_{\tilde{z} \in Z} \mu(\tilde{z}, x, \tilde{y}) U(\tilde{z}) \right) - we^*(\varphi, x) \right]. \tag{10}$$

Relative to the benchmark solution without complexity in (1), the solution to (10) will involve simplifying the regulation φ in order to reduce compliance and supervision costs and the occurrence of mistakes.

Is it possible to approach this trade-off quantitatively? To illustrate this question, we consider the intense policy debate on the complexity of capital requirements, which led the Basel Committee to publish a discussion paper on the trade-offs between "risk sensitivity, simplicity and comparability" (BCBS (2013)). The right trade-off remains elusive, in particular due to the lack of a normative framework to think about regulatory complexity. We sketch how our framework could eventually serve this normative purpose.

We first need an economic model mapping the characteristics x of a bank and a level of capital requirements y to a distribution over outcomes z. This model should also define a welfare criterion, which can be assumed to be the utility function U of the regulator. In Online Appendix OA.11, we derive a simple model of capital regulation for illustration, but in a real policy application one should use a quantitative model of the economy rich enough to accommodate different regulations.[38]

We then need a model of the regulatory text F. In the model of Online Appendix OA.11, we assume a bank invests in assets from a certain asset class $x \in \{1...N\}$, and capital regulation specifies a level y of capital requirements as follows:

`if` $x < \bar{x}_1$ `then` $y \equiv E_1^*$

`else if` $x < \bar{x}_2$ `then` $y \equiv E_2^*$

`...`

`else if` $x < \bar{x}_{I-1}$ `then` $y \equiv E_{I-1}^*$

`else` $y \equiv E_I^*$

where $\bar{x}_i$ are thresholds, E_i^* capital levels, and I the number of different risk buckets, all chosen by the regulator.

Finally, we need to map the different regulatory texts F into complexity costs (either effort or mistakes). This is where text-based and outcome-based measures should be used together. Outcome-based measures can quantify the costs, ideally in dollar terms, of regulations that have actually been implemented. If, moreover, these costs can be related to text-based measures, e.g., *quantity* and *potential*, then one can also estimate the costs of counterfactual regulations and thus solve (10). As an example, in Online Appendix OA.11 we compute

[38]This is precisely where the literature on bank capital requirements is heading. See, for instance, Begenau and Landvoigt (2021), or the BIS' "Financial Regulation Assessment: Meta Exercise" (`https://www.bis.org/frame/`) for a meta-analysis of the quantitative impact of capital requirements.

quantity and *potential* as a function of I in the regulatory text above. We then use the experimental results from Tables OA.14 and OA.17 to estimate how efforts and mistakes depend on I. We can then solve for the number of risk-buckets I that strikes the optimal trade-off between "risk-sensitivity" and "simplicity".[39]

The method we outline here is only a proof of concept, but to our knowledge this is the first proposal offering policymakers a quantitative approach to the trade-off between regulatory complexity and other policy objectives. The actual implementation of this approach for policy would require additional work, in particular a more thorough study of how the properties of a regulation captured by text-based measures, translate into actual costs, captured by outcome-based measures.

4.2 Tools for further research

To help the scientific community make progress on the measurement of regulatory complexity, we make as much material as possible available online, in particular the dictionaries we created, as well as the codes of the tools we used to help us build these dictionaries, so that they can be further enriched. To show that our dictionary is sufficiently rich to be reused on other financial regulation texts, in Online Appendix OA.3 we compute that if we analyzed each title of the DFA based on a dictionary built only from the other titles, we would find 88% of operands and 96% of regulatory operators on average. The second number is particularly encouraging, as Section 3 showed that the most reliable measure is *quantity*, which depends only on regulatory operators.

[39] In our simple framework all asset classes are symmetric, so that only the total number of intervals matters. Following the logic of Gabaix (2025), a more complete model of "behavioral mechanism design" would specify different levels of risk or importance for different asset classes, and the optimal regulation would presumably be more complex for more important asset classes.

As shown in the previous subsection, for a normative use of complexity measures one needs to study how text-based measures of regulatory complexity explain outcome-based measures. This requires to have a sample of regulatory texts that are sufficiently similar to each other, yet differ in complexity, and can also be associated to different outcome variables. This is challenging because all firms in a given industry are typically subject to the same regulations. A solution is then to ask firms which of the regulations they face are more complex, as the EBA did. A similar approach is followed in Amadxarif *et al.* (2019), who use the Q&A section of the EBA website to measure which regulations are more difficult to understand. Several papers also use comments on different regulations (e.g., Gissler *et al.* (2016), Haselmann *et al.* (2022), or Adams and Mosk (2023)), which it would be interesting to correlate with text-based measures.

Several outcome-based measures of complexity proposed in the literature vary in the cross-section of firms (e.g., Trebbi *et al.* (2023) and Singla (2023)), which allows to study their impact on firm performance and other variables of interest at the firm level. This is by definition harder to do for text-based measures, still for the reason that firms in a given industry share the same regulations. In some cases, there are regulatory texts at the firm level, such as the supervision letters studied in Goldsmith-Pinkham *et al.* (2016). Unfortunately, such texts are typically confidential. A possibility for future research would be to measure the exposure of a given firm to different regulatory texts, in the spirit of Kalmenovitz *et al.* (2024). In the case of banking regulation, this would require for instance matching detailed information on a bank's assets and liabilities to the texts determining the regulatory treatments of these items. This is again likely to require detailed supervisory information.

Note that these different approaches always rely on comparing different regulatory texts, so that necessarily part of the variation in complexity will be due to intrinsic complexity. A possibility to have a sample of texts with variation in complexity but intrinsic complexity constant is to use different versions and revisions of the same regulation. An example of this approach in a different context is Georg and Partida (2023), who compare the initial proposal of the European Digital Operational Resilience Act to the final regulation two years later. Interestingly, they find a significant increase in *length* and *quantity* between the two versions, but a decrease in *potential*, which suggests that there can be significant time-series variation in the complexity measures of a given regulation.

5 Conclusion

We propose a comprehensive framework, inspired by the computer science literature, to analyze regulatory complexity. Our framework allows us to distinguish different dimensions of regulatory complexity, derive measures that can be applied to large scale regulatory texts, and conduct validation tests that can be applied to any text-based measure. We find in particular that two measures capturing intrinsic complexity, *quantity* and *potential*, have explanatory power beyond the mere length of a text.

The present work is only a first step in applying this approach to financial regulation. We show that text-based measures of complexity like the ones we propose can help strike a better trade-off between the costs and benefits of additional complexity. However, fully realizing the potential of this approach will require additional empirical work to quantify the impact of different properties of a regulatory text on compliance costs borne by regulated actors, and other costs borne by society (e.g., risks of regulatory capture).

We make available a battery of tools to help the scientific community pursue this research program. Our dictionary allows researchers to compute complexity measures for other financial regulation texts. The experiments we conducted and the EBA survey we use allow interested researchers to test any text-based measure and compare it to the ones we consider in this study. We hope these tools will help the scientific community to build a richer database of regulations and complexity measures, with a view to testing some of the theories of regulatory complexity proposed in the literature, and to helping regulators draft more efficient rules.

References

Adams, R. and Mosk, T. (2023). *Dear Congress Member: Letters to Congress and the Shaping of (Financial) Legislation*. Working paper. 45

Al-Ubaydli, O. and McLaughlin, P. A. (2017). Regdata: A numerical database on industry-specific regulations for all united states industries and federal regulations, 1997–2012. *Regulation and Governance*, **11**, 109–123. 4, 16, 26, 52

Alvero, A., Ando, S. and Xiao, K. (2022). Watch What They Do, Not What They Say: Estimating Regulatory Costs from Revealed Preferences. *The Review of Financial Studies*, **36** (6), 2224–2273. 24

Amadxarif, Z., Brookes, J., Garbarino, N., Patel, R. and Walczak, E. (2019). *The language of rules: textual complexity in banking reforms*. Working paper. 19, 45, 52

Arora, S., Barak, B., Brunnermeier, M. and Ge, R. (2009). *Computational Complexity and Information Asymmetry in Financial Products*. Working paper. 7

Asriyan, V., Foarta, D. and Vanasco, V. (2023). The good, the bad, and the complex: Product design with imperfect information. *American Economic Journal: Microeconomics*, **15** (2), 187–226. 13

Basel Committee on Banking Supervision (1988). *International convergence of capital measurement and capital standards*. Discussion paper. 10

Basel Committee on Banking Supervision (2013). *The regulatory framework: balancing risk sensitivity, simplicity and comparability*. Discussion paper. 1, 6, 42, 114

Begenau, J. and Landvoigt, T. (2021). Financial Regulation in a Quantitative Model of the Modern Banking System. *The Review of Economic Studies*, **89** (4), 1748–1784. 43

Bennett, B., Garvey, G., Milbourn, T. and Wang, Z. (2019). *The Role of Stock Price Informativeness in Compensation Complexity*. Working paper. 1

Boulet, R., Mazzega, P. and Bourcier, D. (2011). A network approach to the french system of legal codes-part i: analysis of a dense network. *Artificial Intelligence and Law*, **19**. 19, 52

Brownlees, C. and Engle, R. F. (2016). SRISK: A Conditional Capital Shortfall Measure of Systemic Risk. *The Review of Financial Studies*, **30** (1), 48–79. 11

Calomiris, C. W. (2018). Has Financial Regulation Been a Flop? (or How to Reform Dodd-Frank). *Journal of Applied Corporate Finance*, **29** (4), 8–24. 1

—, Mamaysky, H. and Yang, R. (2020). *Measuring the Cost of Regulation: A Text-Based Approach*. Working paper. 19, 26, 52

Canfora, G., Garcia, F., Piattini, M., Ruiz, F. and Visaggio, C. (2005). A family of experiments to validate metrics for software process models. *Journal of Systems and Software*, **77** (2), 113 – 129. 4

Carlin, B. I. (2009). Strategic price complexity in retail financial markets. *Journal of Financial Economics*, **91** (3), 278–287. 7

—, Kogan, S. and Lowery, R. (2013). Trading complex assets. *The Journal of Finance*, **68** (5), 1937–1960. 7

Célérier, C. and Vallée, B. (2017). Catering to investors through security design: Headline rate and complexity. *The Quarterly Journal of Economics*, **132** (3), 1469–1508. 1

Cetorelli, N., McAndrews, J. J. and Traina, J. (2014). *Evolution in Bank Complexity*. Working paper. 14

Colliard, J.-E., Eden, L. and Georg, C.-P. (2021). Tax complexity and transfer pricing blueprints, guidelines, and manuals. *Tax Management International Journal*, **50** (2). 7

Ellison, S. F. (2016). Price search and obfuscation: an overview of the theory and empirics. In E. Basker (ed.), *Handbook on the Economics of Retailing and Distribution*, Edward Elgar Publishing. 7

European Banking Authority (2021). *Study of the Cost of Compliance with Supervisory Reporting Requirements*. Report. 34, 37, 38, 107, 108, 113

Foarta, D. and Morelli, M. (2022). *The Common Determinants of Legislative and Regulatory Complexity*. Working paper. 14

Franco, J. P., Yadav, N., Bossaerts, P. and Murawski, C. (2021). Generic properties of a computational task predict human effort and performance. *Journal of Mathematical Psychology*, **104**, 102592. 7

Gabaix, X. (2025). *A Theory of Complexity Aversion*. Working paper. 44

— and Graeber, T. (2024). *The Complexity of Economic Decisions*. Working paper. 14, 38

— and Laibson, D. (2006). Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets. *The Quarterly Journal of Economics*, **121** (2), 505–540. 7

Gai, P., Haldane, A. and Kapadia, S. (2011). Complexity, concentration and contagion. *Journal of Monetary Economics*, **58** (4). 14

—, Kemp, M., Sanchez Serrano, A. and Schnabel, I. (2019). *Regulatory complexity and the quest for robust regulation*. Report of the Advisory Scientific Committee 8, European Systemic Risk Board. 1

Ganglmair, B. and Wardlaw, M. (2017). *Complexity, Standardization, and the Design of Loan Agreements*. Working paper. 1

Georg, C.-P. and Partida, A. (2023). *How complex is the Digital Operational Resilience Act?* Working paper. 46

Ghent, A. C., Torous, W. N. and Valkanov, R. I. (2017). Complexity in Structured Finance. *The Review of Economic Studies*, **86** (2), 694–722. 1

GISSLER, S., OLDFATHER, J. and RUFFINO, D. (2016). Lending on hold: Regulatory uncertainty and bank lending standards. *Journal of Monetary Economics*, **81**, 89–101. 45

GOLDSMITH-PINKHAM, P., HIRTLE, B. and LUCCA, D. O. (2016). *Parsing the content of bank supervision*. Working paper. 45

GUTIÉRREZ, G. and PHILIPPON, T. (2019). *The Failure of Free Entry*. Working paper. 7

HAKENES, H. and SCHNABEL, I. (2012). *Regulatory Capture By Sophistication*. Working paper. 13

HALDANE, A. and MADOUROS, V. (2012). *The dog and the frisbee*. Proceedings, Jackson Hole Economic Policy Symposium. 1, 14, 16, 52

HALSTEAD, M. H. (1977). *Elements of Software Science*. Elsevier. 2, 15, 17, 18, 52

HASELMANN, R., SARKAR, A., SINGLA, S. and VIG, V. (2022). *The political economy of financial regulation*. Working paper. 28, 45

HERRING, R. J. (2018). The evolving complexity of capital regulation. *Journal of Financial Services Research*, **53** (2), 183–205. 16

HOGAN, T. L. and BURNS, S. (2019). Has Dodd-Frank affected bank expenses? *Journal of Regulatory Economics*, **55** (2), 214–236. 4, 24

HUE, S., HURLIN, C., PERIGNON, C. and SAURIN, S. (2025). *Measuring the Driving Forces of Predictive Performance: Application to Credit Scoring*. Working paper. 40

KALMENOVITZ, J. (2023). Regulatory Intensity and Firm-Specific Exposure. *The Review of Financial Studies*. 6, 19, 52

—, LOWRY, M. and VOLKOVA, E. (2024). Regulatory fragmentation. *Journal of Finance*, **Forthcoming**. 18, 45, 52

KATZ, D. and BOMMARITO, M. (2014). Measuring the complexity of the law: the united states code. *Artificial Intelligence and Law*, **22**. 19, 52

KENDALL, C. and OPREA, R. (2024). On the complexity of forming mental models. *Quantitative Economics*, **15**, 175–211. 8

KINCAID, J. P., FISHBURNE, R. P. J., ROGERS, R. L. and CHISSOM, B. S. (1975). *Derivation Of New Readability Formulas (Automated Readability Index, Fog Count And Flesch Reading Ease Formula) For Navy Enlisted Personnel*. Tech. Rep. 56, Institute for Simulation and Training. 19, 52

LI, W., AZAR, P., LAROCHELLE, D., HILL, P. and LO, A. W. (2015). Law is code: A software engineering approach to analyzing the united states code. *Journal of Business and Technology Law*, **10** (2). 1, 3, 16, 19, 24, 26, 52

LOUGHRAN, T. and MCDONALD, B. (2014). Measuring readability in financial disclosures. *The Journal of Finance*, **69** (4), 1643–1671. 7

MAAS, H.-D. (1972). Über den Zusammenhang zwischen Wortschatzumfang und Länge eines Textes. *Zeitschrift für Literaturwissenschaft und Linguistik*, **2** (8). 19, 52

McCabe, T. J. (1976). A complexity measure. *IEEE Transactions on Software Engineering*, **SE-2** (4), 308–320. 16

McLaughlin, P. A., Sherouse, O., Febrizio, M. and King, M. S. (2021). Is Dodd-Frank the Biggest Law Ever? *Journal of Financial Regulation*, **7** (1), 149–174. 16, 24

Murawski, C. and Bossaerts, P. (2016). How humans solve complex problems: The case of the knapsack problem. *Scientific Reports*, **6**. 7

Oehmke, M. and Zawadowski, A. (2019). *The Tragedy of Complexity*. Working paper. 13

Oprea, R. (2020). What makes a rule complex? *American Economic Review*, **110** (12), 3913–51. 7

Papke, L. E. and Wooldridge, J. M. (1996). Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Econometrics*, **11** (6), 619–632. 103

Rubinstein, A. (1986). Finite automata play the repeated prisoner's dilemma. *Journal of Economic Theory*, **39** (1), 83 – 96. 7

Singla, S. (2023). *Regulatory Costs and Market Power*. Working paper. 19, 45, 52

Trebbi, F., Zhang, M. B. and Simkovic, M. (2023). *The Cost of Regulatory Compliance in the United States*. Working paper. 4, 19, 36, 45, 52

Yu, S. and Zhou, S. (2010). A survey on metric of software complexity. In *2010 2nd IEEE International Conference on Information Management and Engineering*, pp. 352–356. 6

Zhang, M. and Baddoo, N. (2007). Performance comparison of software complexity metrics in an open source project. In *European Conference on Software Process Improvement*, pp. 160–174. 4

Zuse, H. (1990). *Software Complexity: Measures and Methods*. New York, NY, United States: de Gruyter, 1st edn. 6, 12

Zwick, E. (2021). The costs of corporate tax complexity. *American Economic Journal: Economic Policy*, **13** (2), 467–500. 1

A Appendix

A.1 Tables

Table 1: Summary of Measures of Regulatory Complexity

This table summarizes the different measures reviewed in Section 1.3 and organizes them according to whether they are text-based or outcome-based, and the dimension of complexity they capture.

Name	Source	Formula	Short description	Complexity Dimension
		Text-based measures		
Length	e.g., Haldane and Madouros (2012)	N	Number of words	Psychological
Cyclomatic complexity	Li *et al.* (2015)	N_L	Number of logical operators	Psychological
Quantity of regulations	Al-Ubaydli and McLaughlin (2017)	N_R	Number of regulatory operators	Intrinsic
Potential volume	This paper and Halstead (1977)	$2+\eta_{OD}$	2+ Number of unique operands	Intrinsic
Operator diversity	This paper	η_{OR}	Number of unique operators	Psychological
Level	This paper and Halstead (1977)	$\frac{2+\eta_{OD}}{N}$	Potential volume normalized by length	Psychological
RegFragmentation	Kalmenovitz *et al.* (2024)	-	Fragmentation of regulation across multiple agencies	Computational
Average word length	e.g., Amadxarif *et al.* (2019)	-	Average number of characters of words used	Psychological
Lexical diversity	Maas (1972)	-	Adjusted proportion of unique words	Psychological
Readability metric	Kincaid *et al.* (1975)	-	Metric based on the length of sentences and words	Psychological
Shannon's entropy	e.g., Katz and Bommarito (2014)	-	Dispersion of the distribution of words used	Psychological
PageRank	Amadxarif *et al.* (2019)	-	Centrality of a rule in a corpus of cross-citing rules	Psychological
Network Centralities	Boulet *et al.* (2011)	-	Centrality of a rule in a corpus of cross-citing rules	Psychological
		Outcome-based measures		
Regulation Index	Trebbi *et al.* (2023)	-	Share of labor costs related to regulation	Computational
RegIn	Kalmenovitz (2023)	-	Compliance costs, in hours of labor	Computational
NetReg	Calomiris *et al.* (2020)	-	Mentions of more minus less regulation in earnings calls	Computational
Regulatory costs	Singla (2023)	-	Costs estimated by regulatory agencies	Computational

Table 2: Correlation of the complexity measures between the algorithmic and the text versions of Basel I items

For each complexity measure, this table reports the correlation between the values obtained when computing the measures on the 19 items of Basel I in the algorithmic version, against the same 19 items in the text version. We compute both Pearson and Spearman (rank) correlation coefficients.

Reminder of the different measures: *length* is the total number of words, *cyclomatic* is the total number of logical operators, *quantity* is the total number of regulatory operators, *potential* is the number of unique operands, *diversity* is the number of unique operators, and *level* is the ratio *potential*/*length*.

	Pearson	Spearman
length	0.68	0.83
cyclomatic	0.41	0.64
quantity	1.00	1.00
potential	0.82	0.8
diversity	0.4	0.48
level	0.45	0.51

Table 3: Top 10 words per category in the DFA

This table reports the 10 most frequent operands, regulatory operators, logical operators, and mathematical operators found in the entire text of the DFA, together with the number of occurrences.

Operands		Operators					
		Regulatory		Logical		Mathematical	
commission	1487	shall	3596	and	9730	adding	267
1	1479	amended	651	or	8959	more	165
2	1222	required	605	any	4007	amount of	154
person	931	regulations	436	such	2849	additional	125
financial	877	regulation	308	as	2633	total	101
bureau	788	established	282	other	1547	minimum	86
corporation	772	determination	281	not	1128	the end of the	85
information	734	establish	247	after	907	more than	83
3	704	compliance	242	including	761	exceed	70
securities	672	require	223	each	697	over	69

Table 4: Summary statistics on complexity measures in the randomly generated regulations

We compute our complexity measures for each of the 38 randomly generated regulations we use in our experiments. This table reports summary statistics on the distribution of each measure across these 38 regulations: mean, standard deviation, minimum, first, second, and third quartile, and maximum.

Reminder of the different measures: *length* is the total number of words, $\overline{cyclomatic}$ is the total number of logical operators divided by *length*, $\overline{quantity}$ is the total number of regulatory operators divided by *length*, $\overline{diversity}$ is the number of unique operators divided by *length*, and *level* is the number of unique operands divided by *length*.

	length	$\overline{cyclomatic}$	$\overline{quantity}$	$\overline{diversity}$	*level*
Mean	32	0.15	0.16	0.15	0.55
Standard Deviation	12	0.06	0.03	0.05	0.09
Minimum	10	0.03	0.10	0.08	0.39
P25	21	0.10	0.14	0.12	0.48
P50	31	0.16	0.16	0.14	0.55
P75	41	0.19	0.18	0.17	0.62
Maximum	57	0.26	0.21	0.30	0.70

Table 5: Correlation of mistakes with measures of complexity, beyond *length*

We estimate the probit regression (7):

$$\Pr(mistake_{i,t} = 1) = \Phi(\alpha + \beta\mu(R_{i,t}) + \gamma_i + \delta_t),$$

where $mistake_{i,t}$ is a dummy variable equal to 1 if participant i gave an incorrect answer in round t, $\mu(R_{i,t})$ is a characteristic or vector of characteristics of the regulation given to participant i in round t, γ_i and δ_t are participant- and round-fixed effects, and Φ is the standardized Gaussian cdf. The table reports the coefficient(s) β when including different measures as regressors, t-statistics (in brackets) computed with robust standard errors, and Mc Fadden's Pseudo-R^2. $*$, $**$, and $***$ denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Reminder of the different measures: *length* is the total number of words, *cyclomatic* is the total number of logical operators, *quantity* is the total number of regulatory operators, *potential* is the number of unique operands, *diversity* is the number of unique operators, and $level = potential/length$.

	(1)	(2)	(3)	(4)	(5)	(6)
length	0.037***	0.055***	-0.005	0.030***	0.065***	0.031***
	(8.67)	(6.14)	(-0.64)	(2.85)	(8.41)	(5.55)
cyclomatic		-0.068**				
		(-2.25)				
quantity			0.554***			
			(6.27)			
potential				0.017		
				(0.73)		
diversity					-0.438***	
					(-4.49)	
level						-1.275
						(-1.60)
Pseudo-R^2	0.243	0.248	0.277	0.244	0.260	0.246

Table 6: Correlation of time taken with measures of complexity, beyond *length*

We estimate the OLS regression (8):

$$time_{i,t} = \alpha + \beta\mu(R_{i,t}) + \gamma_i + \delta_t + \epsilon_{i,t},$$

where $time_{i,t}$ is the time in seconds that participant i took to answer in round t, $\mu(R_{i,t})$ is a characteristic or vector of characteristics of the regulation given to participant i in round t, γ_i and δ_t are participant- and round-fixed effects, and $\epsilon_{i,t}$ is an error term. The table reports the coefficient(s) β when including different measures as regressors, t-statistics (in brackets) computed with robust standard errors, and adjusted R^2. The sample is restricted to correct answers with *time* below the 99th percentile. $*$, $**$, and $***$ denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Reminder of the different measures: *length* is the total number of words, *cyclomatic* is the total number of logical operators, *quantity* is the total number of regulatory operators, *potential* is the number of unique operands, *diversity* is the number of unique operators, and $level = potential/length$.

	(1)	(2)	(3)	(4)	(5)	(6)
length	3.388***	5.371***	2.234***	0.556	2.946***	4.183***
	(14.63)	(9.93)	(3.88)	(0.81)	(6.09)	(11.92)
cyclomatic		-8.084***				
		(-3.84)				
quantity			13.072**			
			(2.30)			
potential				6.807***		
				(4.41)		
diversity					6.899	
					(1.14)	
level						165.462***
						(3.29)
Adjusted-R^2	0.445	0.461	0.450	0.465	0.445	0.455

Table 7: Correlation of the fraction of banks reporting a high cost with measures of complexity, beyond *length*

We estimate the OLS regression (9):

$$cost_{i,j} = \alpha + \beta\mu(R_j) + \gamma_2 medium_i + \gamma_3 small_i + \epsilon_{i,j},$$

where $cost_{i,j}$ is the fraction of banks in group $i \in \{large, medium, small\}$ that report a high or medium-high cost for template j, $\mu(R_j)$ is a characteristic or vector of characteristics of template j, $medium_i$ and $small_i$ are dummy variables equal to 1 if $i = medium$ or $small$, respectively, and $\epsilon_{i,j}$ is an error term. The table reports the coefficient(s) β when including different measures as regressors, t-statistics (in brackets) computed with robust standard errors, and adjusted R^2. $*$, $**$, and $***$ denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Reminder of the different measures: *length* is the total number of words, *cyclomatic* is the total number of logical operators, *quantity* is the total number of regulatory operators, *potential* is the number of unique operands, *diversity* is the number of unique operators, and $level = potential/length$.

	(1)	(2)	(3)	(4)	(5)	(6)
length	0.00001***	0.00002***	-0.00000	0.00001***	0.00001***	0.00001***
	(6.64)	(2.93)	(-0.00)	(5.48)	(5.13)	(5.58)
cyclomatic		-0.00009				
		(-1.28)				
quantity			0.00006***			
			(3.83)			
potential				0.00026*		
				(1.76)		
diversity					0.00015	
					(0.38)	
level						-0.52748*
						(-1.78)
Adjusted-R^2	0.170	0.172	0.210	0.179	0.166	0.178

Table 8: Correlation of the importance given to templates by regulators with measures of complexity, beyond *length*

We estimate the OLS regression:

$$importance_j = \alpha + \beta\mu(R_j) + \epsilon_j,$$

where $importance_j$ is the average importance given by the surveyed regulatory authorities to template j, where 3 corresponds to the answer "highly important", 2 to "important", 1 to "less important", and 0 to "not important". $\mu(R_j)$ is a characteristic or vector of characteristics of template j and ϵ_j is an error term. The table reports the coefficient(s) β when including different measures as regressors, t-statistics (in brackets) computed with robust standard errors, and adjusted R^2. $*$, $**$, and $***$ denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Reminder of the different measures: *length* is the total number of words, *cyclomatic* is the total number of logical operators, *quantity* is the total number of regulatory operators, *potential* is the number of unique operands, *diversity* is the number of unique operators, and $level = potential/length$.

	(1)	(2)	(3)	(4)	(5)	(6)
length	0.00005**	-0.00006	0.00011***	0.00002	0.00002	0.00005**
	(2.48)	(-1.18)	(3.09)	(1.37)	(1.28)	(2.10)
cyclomatic		0.00125**				
		(2.54)				
quantity			-0.00026*			
			(-2.00)			
potential				0.00327***		
				(3.79)		
diversity					0.00744***	
					(3.23)	
level						-0.58279
						(-0.39)
Adjusted-R^2	0.090	0.141	0.111	0.195	0.176	0.071

Table 9: Synthesis - Explanatory power of *quantity* and *potential* against *length*, for all dependent variables

This table systematically compares the explanatory power of *quantity* and *potential* to *length*, for the four dependent variables *mistake*, *time*, *cost*, and *importance*. In each panel, each column corresponds to the regression of a dependent variable on *length* and *quantity* (Panel A) or *length* and *potential* (Panel B). The line "Table / Column" indicates the Table and Column with the results and details of the corresponding regression (e.g., detail of the fixed effects used). The characters are in bold if this regression is the one with the highest adjusted R^2 in the corresponding Table.

For each regression and variable, we report the standardized regression coefficient, the p-value of the Student test that the coefficient is zero, and the percentage of the R^2 attributed to the variable. For instance, Column 3 of Panel A considers the regression specification (3) of Table 7. In that regression, a one standard deviation increase in *quantity* is associated with a 0.45 standard deviation increase in *cost*. The null hypothesis that the coefficient on quantity is zero can be rejected at the 0.00 level, and 89% of the R^2 of the regression can be attributed to *quantity*.

Reminder of the dependent variables: *mistake* is a dummy variable equal to 1 if a participant in the experiment gave a wrong answer on a given regulation, *time* is the time that the participant needed, *cost* is the fraction of banks that report a high or medium-high cost for ITS templates, *importance* is the average importance given to ITS templates by regulatory authorities. See the corresponding tables for more detailed definitions.

Reminder of the different measures: *length* is the total number of words, *quantity* is the total number of regulatory operators, and *potential* is the number of unique operands.

Panel A - length and quantity

	mistake		time		**cost**		importance	
Table / Column	OA.18 / (3)		6 / (3)		7 / (3)		8 / (3)	
Variable	**length**	**quantity**	length	quantity	**length**	**quantity**	length	quantity
Standardized coeff.	**-0.03**	**0.28**	0.29	0.17	**-0.00**	**0.45**	0.71	-0.43
p-value	**0.53**	**0.00**	0.00	0.02	**1.00**	**0.00**	0.00	0.05
Contribution to R^2	**-2%**	**17%**	24%	14%	**-0%**	**89%**	154%	-54%

Panel B - length and potential

	mistake		**time**		cost		**importance**	
Table / Column	OA.18 / (4)		6 / (4)		7 / (4)		8 / (4)	
Variable	length	potential	**length**	**potential**	length	potential	**length**	**potential**
Standardized coeff.	0.16	0.05	**0.07**	**0.40**	0.32	0.14	**0.11**	**0.42**
p-value	0.03	0.47	**0.42**	**0.00**	0.00	0.08	**0.18**	**0.00**
Contribution to R^2	8%	3%	**6%**	**33%**	65%	23%	**15%**	**85%**

A.2 Figures

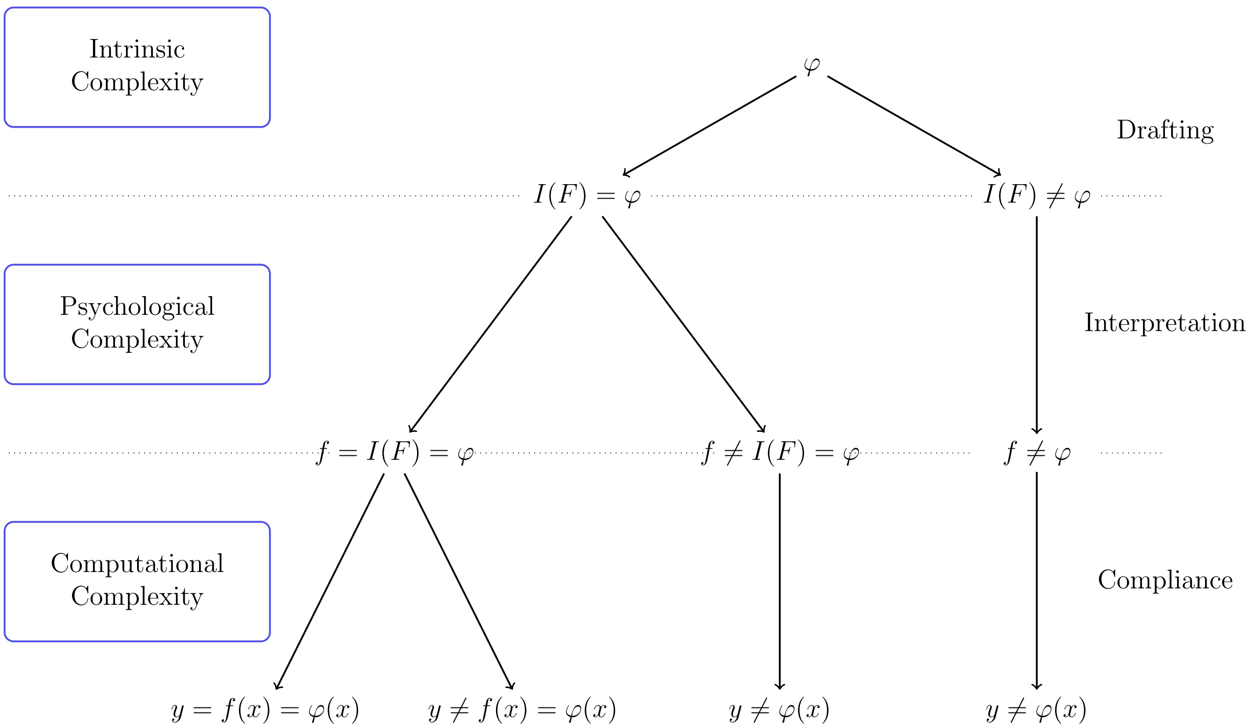


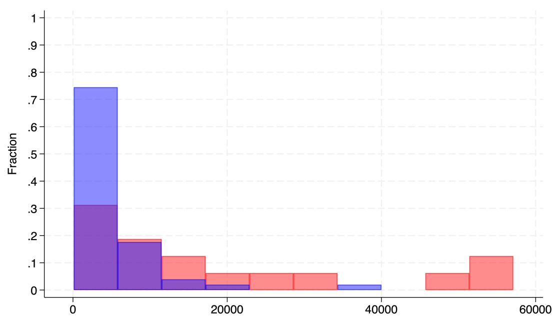
Figure 1: Regulatory process with complexity

This figure summarizes the model of the regulatory process described in Section 1. It shows how intrinsic, psychological, and computational complexity affect the likelihood of implementing the desired regulation. If the regulation is drafted incorrectly, then $I(F) \neq \varphi$. If a correct regulation is interpreted incorrectly, then $f \neq I(F)$. Lastly, if a mistake is made at the compliance stage then $y \neq f(x)$.

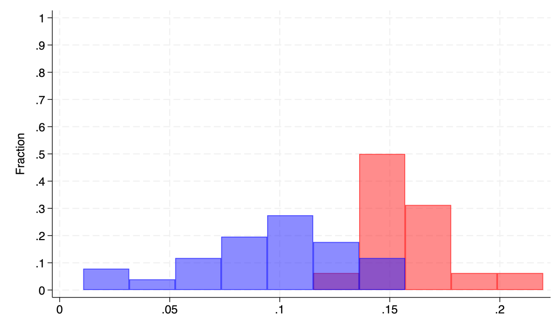
Figure 2: Histogram of complexity measures for the DFA and the ITS

This figure shows histograms of the values of our complexity measures for each of the 16 Titles of the DFA (red) and the 51 templates or groups of templates of the ITS (blue).

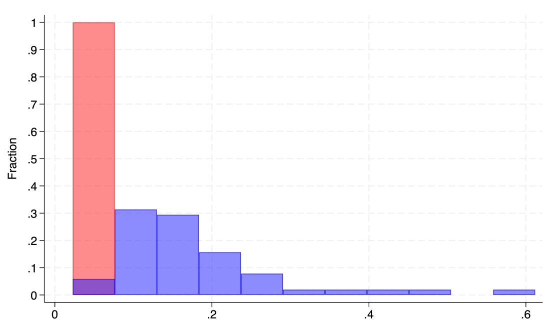
Reminder of the different measures: *length* is the total number of words, $\overline{cyclomatic}$ is the total number of logical operators divided by *length*, $\overline{quantity}$ is the total number of regulatory operators divided by *length*, $\overline{diversity}$ is the number of unique operators divided by *length*, and *level* is the number of unique operands divided by *length*.



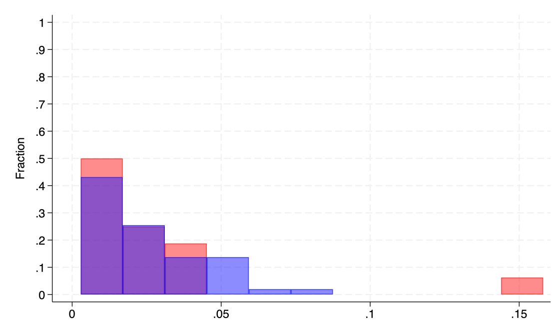
(a) *length*



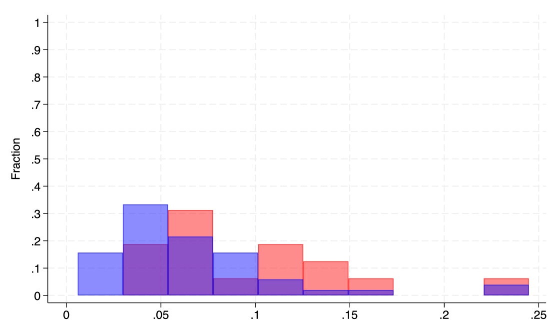
(b) $\overline{cyclomatic}$



(c) $\overline{quantity}$



(d) $\overline{diversity}$



(e) *level*

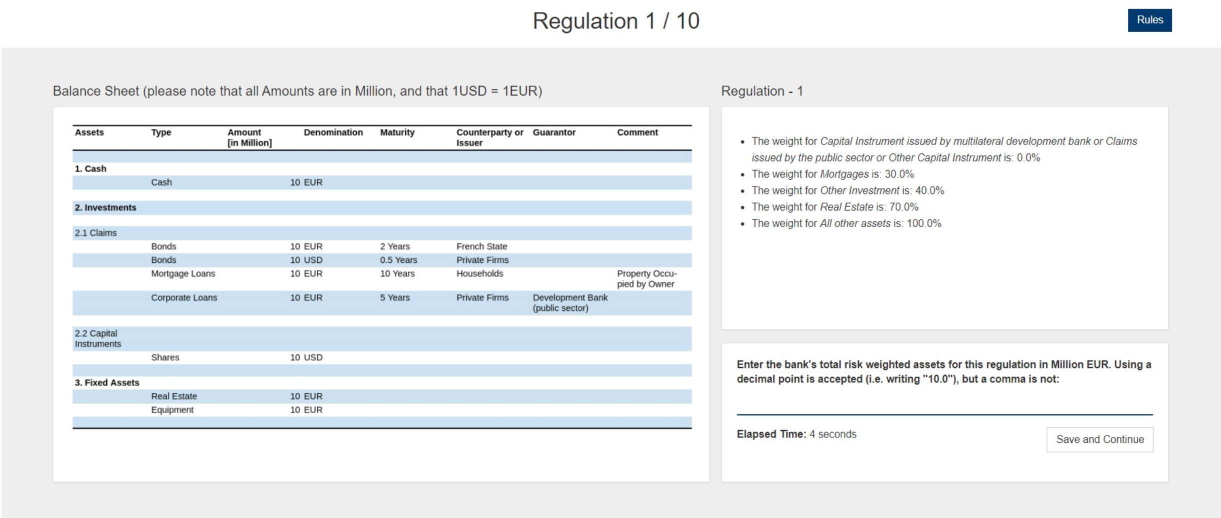


Figure 3: Online experiment - Test round

This figure is a screenshot of the main page of the website we use for the experiments. It corresponds to the test round, which is common to all participants. Every subsequent round uses the same layout, with a different regulation on the right-hand side.

A.3 Example of a Randomly Generated Regulation

We report here one of the random regulations generated by our algorithm. We first report the raw output and then the "translated" text that students saw in the experiment.

`IF (`
`    ASSET_CLASS == "capital_instruments"`
`AND (`
`        (ISSUER == "banks)`
`) THEN:`
`  risk_weight = 0.7;`
	The weight for *Capital Instrument issued by a bank* is: 70.0%

`IF (`
`    (ASSET_CLASS == "real_estate" OR`
`ASSET_CLASS == "other_investments" OR`
`ASSET_CLASS == "other_cap_inst")`
`) THEN:`
`  risk_weight = 0.0;`
	The weight for *Real Estate or Other Fixed Asset or Other Capital Instrument* is: 0.0%

`IF (`
`    (ASSET_CLASS == "loans" AND MATURITY`
`<= 1) OR (ASSET_CLASS == "claims" and`
`GUARANTEED == "central_government") OR`
`(ASSET_CLASS == "cash") OR (ASSET_CLASS`
`== "claims" AND ISSUER ==`
`"central_government)`
`) THEN:`
`  risk_weight = 0.1;`
	The weight for *Loans with asset maturity less than one year or Claims guaranteed by central government or Cash orClaims issued by central government* is: 10.0%

`IF (`
`    (ASSET_CLASS == "other_loans" OR`
`ASSET_CLASS == "other_claims" OR`
`ASSET_CLASS == "other_investment")`
`) THEN:`
`  risk_weight = 0.6;`
	The weight for *Other Loans or Other Claims or Other Investment* is: 60.0%

`ELSE THEN:` `  risk_weight = 1.0;`	The weight for *All other assets* is: 100.0%

B Online Appendix to "Measuring Regulatory Complexity"

This Online Appendix provides additional material omitted from the main text.

OA.1 Two Representations of Basel I Risk-Weights

In the following, we provide a description of the Basel I regulation in the form of a stylized algorithm and compare it side by side with the actual text of the regulation. We use pseudo code that simply captures the logical flow of the instructions in Basel I. To compute the Halstead measures for each item we consider the code contained between two "ASSET_CLASS ==" (excluding this expression). This section reports the text we used to compute the complexity measures in Table OA.4.

Basel I Algorithm	Basel I Text

`IF (`
`    ASSET_CLASS == "cash"`
`) THEN:`
`    risk_weight = 0.0;`
	Cash shall have a 0% risk weight

`IF (`
`    ASSET_CLASS == "claims" AND`
`        (ISSUER == "`
`central_governments" OR ISSUER == "`
`central_banks") AND`
`        DENOMINATION == "national" AND`
`        FUNDING_CURRENCY == "national"`
`) THEN:`
`    risk_weight = 0.0;`
	Claims on central governments and central banks denominated in national currency and funded in that currency shall have a 0% risk weight

`IF (`
`    ASSET_CLASS == "claims" AND`
`        (ISSUER == "`
`central_governments" OR ISSUER == "`
`central_banks") AND`
`        ISSUER_COUNTRY == "oecd"`
`) THEN:`
`    risk_weight = 0.0;`
	Other claims on OECD central governments and central banks shall have a 0% risk weight

`IF (`
`    ASSET_CLASS == "claims" AND`
`        (COLLATERALIZED == "oecd" OR`
`GUARANTEED == "oecd")`
`) THEN:`
`    risk_weight = 0.0;`
	Claims collateralised by cash of OECD central-government securities or guaranteed by OECD central governments shall have a 0% risk weight

`IF (`
`    ASSET_CLASS == "claims" AND (`
`        (ISSUER ==`
`"public-sector_entities" AND`
`ISSUER_COUNTRY == "domestic") AND`
`        (ISSUER != "central_governments"`
`AND ISSUER_COUNTRY == "domestic")`
`    ) OR`
`    ASSET_CLASS == "loans" AND (`
`        (GUARANTEED ==`
`"public-sector_entities" AND`
`GUARANTEED_COUNTRY == "domestic") AND`
`        (GUARANTEED !=`
`"central_governments" AND`
`GUARANTEED_COUNTRY == "domestic")`
`    )`
`) THEN:`
`    risk_weight = national_discretion;`
	Claims on domestic public-sector entities, excluding central government, and loans guaranteed by such entities shall have a 0%, 10%, 20%, or 50% risk weight (at national discretion)

```
IF (
    ASSET_CLASS == "claims" AND (
        (ISSUER == "IBRD" OR ISSUER ==
"IADB" OR ISSUER == "AsDB" OR ISSUER ==
"AfDB" OR
            ISSUER == "EIB") OR
        (GUARANTEED == "IBRD" OR
GUARANTEED == "IADB"  OR GUARANTEED ==
"AsDB" OR
            GUARANTEED == "AfDB" OR
GUARANTEED == "EIB") OR
        (COLLATERALIZED == "IBRD" OR
COLLATERALIZED == "IADB"  OR
COLLATERALIZED == "AsDB" OR
            COLLATERALIZED == "AfDB" OR
COLLATERALIZED == "EIB")
) THEN:
    risk_weight = 0.2;
```

Claims on multilateral development banks (IBRD, IADB, AsDB, AfDB, EIB) and claims guaranteed by, or collateralized by securities issued by such banks shall have a 20% risk weight

```
IF (
    ASSET_CLASS == "claims" AND (
        (ISSUER == "bank" AND
ISSUER_COUNTRY == "oecd")
    ) OR
    ASSET_CLASS == "loans" AND
        (GUARANTEED == "bank" AND
GUARANTEED_COUNTRY == "oecd")
) THEN:
    risk_weight = 0.2;
```

Claims on banks incorporated in the OECD and loans guaranteed by OECD incorporated banks shall have a 20% risk weight

```
IF (
    ASSET_CLASS == "claims" AND (
        (ISSUER == "bank" AND
ISSUER_COUNTRY != "oecd" AND
ASSET_MATURITY <= 1)
    ) OR
    ASSET_CLASS == "loans" AND
        (GUARANTEED == "bank" AND
GUARANTEED_COUNTRY != "oecd" AND
ASSET_MATURITY <= 1)
) THEN:
    risk_weight = 0.2;
```

Claims on banks incorporated in countries outside the OECD with a residual maturity of up to one year and loans with a residual maturity of up to one year guaranteed by banks incorporated in countries outside the OECD shall have a 20% risk weight

```
IF (
    ASSET_CLASS == "claims" AND (
        (ISSUER ==
"public_sector_entities" AND ISSUER !=
"central_governments" AND
            ISSUER_COUNTRY == "oecd" AND
ISSUER_COUNTRY != "domestic")
    ) OR
    ASSET_CLASS == "loans" AND
        (GUARANTEED ==
"public_sector_entities" AND GUARANTEED
!= "central_governments" AND
            GUARANTEED_COUNTRY == "oecd"
AND GUARANTEED_COUNTRY != "domestic")
) THEN:
    risk_weight = 0.2;
```

Claims on non-domestic OECD public-sector entities, excluding central government, and loans guaranteed by such entities shall have a 20% risk weight

```
IF (
    ASSET_CLASS == "cash" AND (
        CASH_COLLECTION == "in_process"
    )
) THEN:
    risk_weight = 0.2;
```

Cash items in process of collection shall have a 20% risk weight

```
IF (
    ASSET_CLASS == "loans" AND
        (LOAN_SECURITY == "mortgage" AND
(PROPERTY_OCCUPIED == "owner" OR
            PROPERTY_OCCUPIED ==
"rented"))
) THEN:
    risk_weight = 0.5;
```

Loans fully secured by mortgage on residential property that is or will be occupied by the borrower or that is rented shall have a 50% risk weight

```
IF (
    ASSET_CLASS == "claims" AND (
        ISSUER == "private_sector"
) THEN:
  risk_weight = 1.0;
```

Claims on the private sector shall have a 100% risk weight

```
IF (
    ASSET_CLASS == "claims" AND (
        (ISSUER == "banks" AND
ISSUER_COUNTRY != "oecd" AND
ASSET_MATURITY > 1)
) THEN:
  risk_weight = 1.0;
```

Claims on banks incorporated outside the OECD with a residual maturity of over one year shall have a 100% risk weight

```
IF (
    ASSET_CLASS == "claims" AND (
        (ISSUER == "central_governments"
AND ISSUER_COUNTRY != "oecd" AND
            DENOMINATION != "national"
AND FUNDING_CURRENCY != "national")
) THEN:
  risk_weight = 1.0;
```

Claims on central governments outside the OECD (unless denominated in national currency - and funded in that currency - see above) shall have a 100% risk weight

```
IF (
    ASSET_CLASS == "claims" AND (
        (ISSUER ==
"commercial_companies" AND ISSUER_OWNER
== "public_sector")
) THEN:
  risk_weight = 1.0;
```

Claims on commercial companies owned by the public sector shall have a 100% risk weight

```
IF (
    (ASSET_CLASS == "premises" OR
ASSET_CLASS == "plant" OR ASSET_CLASS ==
"equipment" OR
        ASSET_CLASS ==
"other_fixed_assets") OR
) THEN:
  risk_weight = 1.0;
```

Premises, plant and equipment and other fixed assets shall have a 100% risk weight

```
IF (
    (ASSET_CLASS == "real_estate" OR
ASSET_CLASS == "other_investments") OR
) THEN:
  risk_weight = 1.0;
```

Real estate and other investments (including non-consolidated investment participations in other companies) shall have a 100% risk weight

```
IF (
    ASSET_CLASS == "capital_instruments"
AND (
        (ISSUER == "banks" AND
DEDUCTED_FROM != "capital")
) THEN:
  risk_weight = 1.0;
```

Capital instruments issued by other banks (unless deducted from capital) shall have a 100% risk weight

```
ELSE THEN:
  risk_weight = 1.0;
```

All other assets shall have a 100% risk weight

OA.2 Complexity of Basel I - Descriptive Statistics

This section gives additional descriptions of the measures we computed on the Basel I rules, both the algorithmic version and the text version.

We report the measures computed on the algorithmic version of Basel I in Table OA.1. In addition, Table OA.2 gives the pair-wise correlation coefficients between the different measures, across the 19 regulatory instructions. We report both the Pearson and Spearman correlation coefficients. Since each item between (1a) and (5h) contains by construction exactly one regulatory instruction, the measure *quantity* is always equal to 1 and its correlation with other measures is undefined. The measures *length*, *cyclomatic*, and *level* are highly correlated with each other, while *potential* and *diversity* are less correlated and thus potentially bring information not captured before.

Turning to the text version, we first report the top 5 words in each category in Table OA.3. We then report the measures computed for each item in Table OA.4, and the correlations between the different measures across items in Table OA.5.

Table OA.1: Complexity measures for Basel I (algorithmic approach)

This table reports the value of each complexity measure for each of the 19 items in Annex 2 of Basel I, as well as the entire text, using the algorithmic approach explained in Section 2.1.

Reminder of the different measures: *length* is the total number of words, *cyclomatic* is the total number of logical operators, *quantity* is the total number of regulatory operators, *potential* is the number of unique operands, *diversity* is the number of unique operators, and *level* = *potential*/*length*.

Regulation	*length*	*cyclomatic*	*quantity*	*potential*	*diversity*	*level*
1a	8	2	1	6	4	0.75
1b	24	6	1	12	6	0.5
1c	20	5	1	11	6	0.55
1d	16	4	1	9	6	0.56
2a	43	11	1	14	7	0.33
3a	68	17	1	14	6	0.21
3b	26	7	1	12	6	0.46
3c	34	9	1	14	8	0.41
3d	44	11	1	15	7	0.34
3e	12	3	1	8	5	0.67
4a	20	5	1	11	6	0.55
5a	12	3	1	8	5	0.67
5b	20	5	1	12	7	0.6
5c	22	6	1	12	6	0.55
5d	16	4	1	10	5	0.63
5e	21	6	1	9	5	0.43
5f	13	4	1	7	5	0.54
5g	16	4	1	10	6	0.63
5h	5	2	1	4	3	0.8
Total	440	114	19	54	10	0.12

Table OA.2: Pairwise correlations between complexity measures in Basel I (algorithmic approach)

For each pair of complexity measures, this table reports their correlation across the 19 items in Annex 2 of Basel I. The measures are computed using the algorithmic approach. We compute both Pearson and Spearman (rank) correlation coefficients.

Reminder of the different measures: *length* is the total number of words, *cyclomatic* is the total number of logical operators, *potential* is the number of unique operands, *diversity* is the number of unique operators, and $level = potential/length$. *quantity* is not included, as it is constant across items.

Panel A: Pearson Correlation Coefficients

	length	*cyclomatic*	*potential*	*diversity*	*level*
length	1	1	0.81	0.6	-0.93
cyclomatic	1	1	0.8	0.58	-0.94
potential	0.81	0.8	1	0.9	-0.83
diversity	0.6	0.58	0.9	1	-0.67
level	-0.93	-0.94	-0.83	-0.67	1

Panel B: Spearman Rank Correlation Coefficients

	length	*cyclomatic*	*potential*	*diversity*	*level*
length	1	0.99	0.94	0.78	-0.93
cyclomatic	0.99	1	0.92	0.76	-0.95
potential	0.94	0.92	1	0.89	-0.79
diversity	0.78	0.76	0.89	1	-0.65
level	-0.93	-0.95	-0.79	-0.65	1

Table OA.3: Top 5 words per category in Annex 2 of Basel I (textual approach)

This table reports up to the 5 most frequent operands, regulatory operators, logical operators, and mathematical operators found in Annex 2 of Basel I (textual approach), together with the number of occurrences.

Operands		Operators: Regulatory		Logical		Mathematical	
risk weight	19	shall have	19	and	12	up to	2
claims	15			other	6	above	1
banks	10			or	5	all	1
OECD	10			outside	4	over	1
central	9			excluding	2		

Table OA.4: Complexity measures for Basel I (textual approach)

This table reports the value of each complexity measure for each of the 19 items in Annex 2 of Basel I, as well as the entire text, using the textual approach explained in Section 2.1.

Reminder of the different measures: *length* is the total number of words, *cyclomatic* is the total number of logical operators, *quantity* is the total number of regulatory operators, *potential* is the number of unique operands, *diversity* is the number of unique operators, and $level = potential/length$.

Regulation	*length*	*cyclomatic*	*quantity*	*potential*	*diversity*	*level*
1a	5	0	1	5	1	1
1b	20	2	1	12	2	0.6
1c	13	2	1	9	3	0.69
1d	19	1	1	13	2	0.68
2a	26	3	1	18	4	0.69
3a	26	2	1	17	3	0.65
3b	18	1	1	9	2	0.5
3c	38	3	1	13	4	0.34
3d	21	3	1	14	4	0.67
3e	9	0	1	8	1	0.89
4a	25	2	1	13	2	0.52
5a	9	0	1	7	1	0.78
5b	18	1	1	11	3	0.61
5c	23	3	1	12	5	0.52
5d	13	0	1	10	1	0.77
5e	12	3	1	9	3	0.75
5f	16	5	1	10	5	0.63
5g	14	2	1	9	3	0.64
5h	7	1	1	5	3	0.71
Total	332	34	19	69	13	0.21

Table OA.5: Pairwise correlations between complexity measures in Basel I (textual approach)

For each pair of complexity measures, this table reports their correlation across the 19 items in Annex 2 of Basel I. The measures are computed using the textual approach. We compute both Pearson and Spearman (rank) correlation coefficients.

Reminder of the different measures: *length* is the total number of words, *cyclomatic* is the total number of logical operators, *potential* is the number of unique operands, *diversity* is the number of unique operators, and $level = potential/length$. *quantity* is not included, as it is constant across items.

Panel A: Pearson Correlation Coefficients

	length	*cyclomatic*	*potential*	*diversity*	*level*
length	1	0.53	0.82	0.52	-0.81
cyclomatic	0.53	1	0.48	0.89	-0.53
potential	0.82	0.48	1	0.44	-0.47
diversity	0.52	0.89	0.44	1	-0.55
level	-0.81	-0.53	-0.47	-0.55	1

Panel B: Spearman Rank Correlation Coefficients

	length	*cyclomatic*	*potential*	*diversity*	*level*
length	1	0.6	0.93	0.53	-0.74
cyclomatic	0.6	1	0.56	0.89	-0.51
potential	0.93	0.56	1	0.47	-0.52
diversity	0.53	0.89	0.47	1	-0.48
level	-0.74	-0.51	-0.52	-0.48	1

OA.3 Details on the Dictionaries

As discussed in Section 2.2, we have created a dictionary of all the n-grams that appear in the Dodd-Frank Act and the ITS. The detailed process is the following:

1. We start by manually classifying n-grams using the dashboard discussed in Section 2.2, and reproduced in Fig. OA.1 below. This results in 6,115 unique entries and a marked-up version of the DFA where each classified n-gram is enclosed in a `<span class="Category"> </span>` html tag. The `Category` of each n-gram is either Logical Operators, Regulatory Operators, Operands (Economics Operands or Attributes), or Other (Legal References, Function Words, or ambiguous words). We record all residual text that is not manually classified.

2. We then standardize the n-grams in our dictionary by stripping away all special characters such as `'",;:.()` and transforming each n-gram into uppercase. This leaves us with a standardized dictionary of 9,099 n-grams.

3. Next, we sort those n-grams from longest to shortest and iterate through the similarly standardized text of the DFA again, removing each identified n-gram from the remaining text. We do this for each n-gram and in turn are able to match virtually the entire text of the DFA.

This process, however, does not scale easily because it takes time to find n-grams manually. Therefore, we have automated this process further and implemented the below algorithm which we use to create the final dictionaries for the Dodd-Frank Act and the ITS:

1. First, we identify and record candidate 12-grams, i.e., sequences of 12 words and numbers that appear several times in the same order in the text. We require that each candidate n-gram appears at least 5 times in the text.

2. We require words to be preceded and followed by a stop character like a whitespace or punctuation. This ensures that we do not identify substrings of whole words at any stage of the process.

3. Next, we manually inspect whether these are true 12-grams, i.e., 12 words representing a single object or entity, for example a legal reference or an economic operand.

4. We record true 12-grams and remove them from the text.

5. Then we use the resulting text and start at step 1 using 11-grams instead of 12-grams.

6. We repeat this process until we find all candidate 2-grams, of which there can be many.

7. Once we have identified and removed all true 2-grams, any remaining words and numbers must be true 1-grams.

For the Dodd-Frank Act, we started with the manually constructed dictionary and then extended it using the above algorithm. The code implementing this algorithm and the list of manually identified n-grams can be found in our replication package.

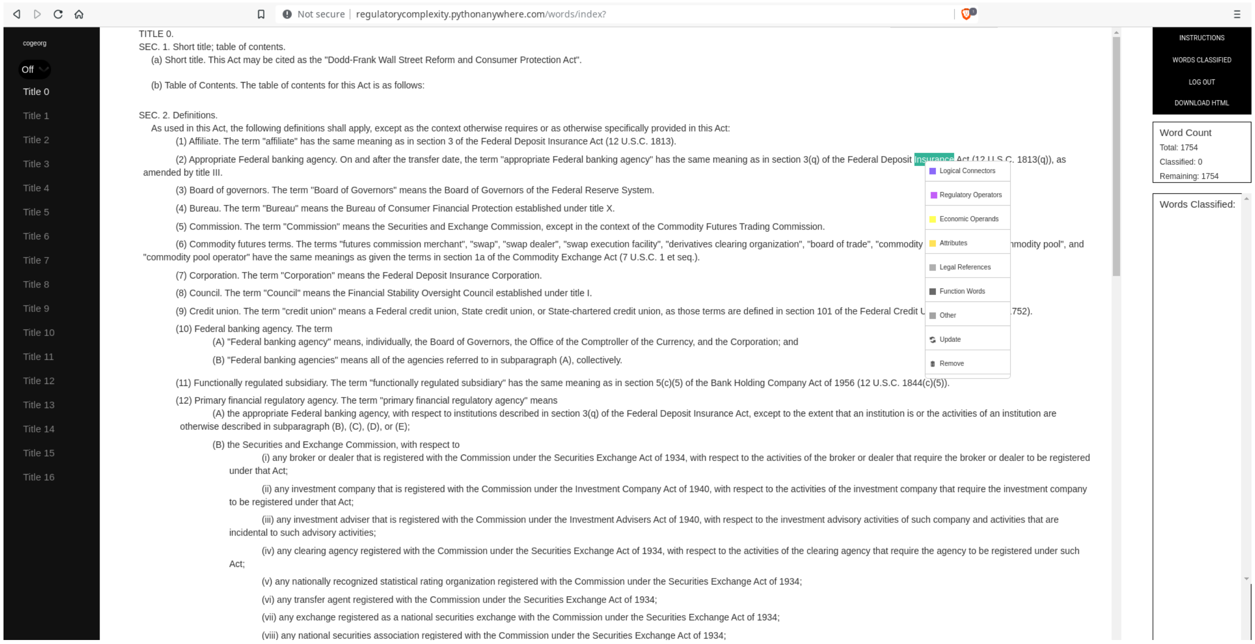


Figure OA.1: DFA Dashboard

This figure is a screenshot of the dashboard we developed to help us identify n-grams and classify words in the DFA into different operands and operators categories. Top: The plain text of the DFA. When highlighting a word or phrase, our dashboard displays a simple drop-down menu from which the category can be selected. The dashboard also provides some simple statistics on the right of the screen, and navigation on the left. Bottom: A mark-up of the DFA when all words and phrases have been classified.

We end this section with an exercise to quantify to what extent a dictionary based on one text also captures operands and operators in another text. For each title i between 1 and 16 of the DFA, we create an alternative dictionary based on all the words classified outside of title i. We then treat title i as a new regulation, and count what percentage of words we are able to classify based on the alternative dictionary. In addition, we also count this percentage separately for operands, operators of different types, and other words. As shown in Table OA.6, on average across all titles we are able to retrieve 88% of all words. Moreover, many of the words we cannot find are in the "Other" category and would not be used in the complexity measures anyway. We also find 97% of logical operators and 96% of regulatory operators, used to compute *cyclomatic* and *quantity*, respectively.

Table OA.6: Completeness of the DFA dictionary

This table shows, for each title of the DFA, the fraction of words in each category (all, operands, different operators, and other) that also appear in at least one other title. For example, 0.90 on line 1 and column "Operands" means that 90% of operands that appear in Title 1 would have been present in a dictionary built using Titles 2 to 16 only. The last line reports the average across the 16 titles.

Title	All	Operands	Operators			Other
			Logical	Regulatory	Mathematical	
1	0.90	0.90	0.97	0.99	0.88	0.84
2	0.86	0.86	0.97	0.89	0.79	0.84
3	0.85	0.93	1.00	0.95	1.00	0.68
4	0.94	0.92	0.99	1.00	1.00	0.96
5	0.89	0.87	0.98	0.96	1.00	0.89
6	0.88	0.91	0.98	0.98	0.96	0.77
7	0.83	0.84	0.94	0.95	0.92	0.74
8	0.96	0.96	1.00	0.98	1.00	0.91
9	0.81	0.82	0.92	0.94	0.88	0.73
10	0.78	0.81	0.92	0.93	0.91	0.63
11	0.91	0.91	0.98	0.96	1.00	0.88
12	0.96	0.94	1.00	1.00	1.00	0.98
13	0.89	0.92	1.00	1.00	1.00	0.77
14	0.80	0.80	0.92	0.93	0.87	0.72
15	0.87	0.85	0.98	0.97	0.93	0.87
16	0.91	0.87	1.00	1.00	.	0.91
Average	0.88	0.88	0.97	0.96	0.94	0.82

OA.4 Complexity Measures for the DFA and the ITS

Table OA.7: Complexity measures for the DFA

This table reports the value of each complexity measure for each of the 16 titles of the DFA, as well as the entire text.

Reminder of the different measures: *length* is the total number of words, $\overline{cyclomatic}$ is the total number of logical operators divided by *length*, $\overline{quantity}$ is the total number of regulatory operators divided by *length*, $\overline{diversity}$ is the number of unique operators divided by *length*, and *level* is the number of unique operands divided by *length*.

Title	*length*	$\overline{cyclomatic}$	$\overline{quantity}$	$\overline{diversity}$	*level*
1	18,641	0.14	0.05	0.01	0.07
2	29,076	0.18	0.03	0.01	0.06
3	15,229	0.15	0.03	0.01	0.06
4	3,626	0.16	0.03	0.04	0.12
5	6,048	0.16	0.04	0.02	0.12
6	14,672	0.16	0.04	0.01	0.07
7	56,703	0.17	0.05	0.01	0.04
8	6,616	0.15	0.05	0.02	0.09
9	48,275	0.14	0.04	0.01	0.05
10	57,095	0.16	0.04	0.01	0.05
11	6,023	0.15	0.04	0.02	0.11
12	1,337	0.15	0.05	0.04	0.17
13	1,009	0.17	0.04	0.04	0.14
14	28,456	0.15	0.04	0.01	0.07
15	3,826	0.13	0.03	0.03	0.14
16	132	0.22	0.02	0.14	0.24
Entire Act	296,764	0.16	0.04	0.00	0.02

Table OA.8: Complexity measures for the ITS

This table reports the value of each complexity measure for each of the 51 templates or groups of templates of the ITS.

Reminder of the different measures: *length* is the total number of words, $\overline{cyclomatic}$ is the total number of logical operators divided by *length*, $\overline{quantity}$ is the total number of regulatory operators divided by *length*, $\overline{diversity}$ is the number of unique operators divided by *length*, and *level* is the number of unique operands divided by *length*.

Template	*length*	$\overline{cyclomatic}$	$\overline{quantity}$	$\overline{diversity}$	*level*
C01-05	17,099	0.12	0.06	0.01	0.03
C06	2,780	0.13	0.08	0.03	0.07
C07	6,773	0.11	0.13	0.02	0.05
C08+C10	9,823	0.12	0.14	0.02	0.04
C09	3,700	0.10	0.20	0.03	0.05
C11	948	0.10	0.09	0.04	0.08
C15	95	0.05	0.12	0.05	0.22
C16	1,568	0.14	0.08	0.05	0.10
C17	5,160	0.11	0.17	0.02	0.04
C18-23	6,294	0.09	0.29	0.02	0.04
C24	988	0.10	0.08	0.06	0.10
C25	417	0.11	0.12	0.08	0.16
C26-29	3,272	0.12	0.09	0.04	0.07
C32.01	1,560	0.11	0.15	0.04	0.06
C32.02	3,774	0.11	0.09	0.03	0.05
C32.03+32.04	1,452	0.09	0.04	0.05	0.10
C33.00	4,747	0.12	0.17	0.02	0.04
C40+44+47	11,236	0.14	0.06	0.01	0.03
C66	5,688	0.02	0.45	0.00	0.02
C67+68	2,313	0.10	0.11	0.03	0.06
C69	1,215	0.14	0.13	0.06	0.11
C70	7,294	0.01	0.56	0.00	0.01
C71	946	0.09	0.15	0.05	0.15
C72-77	12,058	0.10	0.26	0.00	0.02

Table OA.9: Complexity measures for the ITS - Continued

Template	*length*	$\overline{cyclomatic}$	$\overline{quantity}$	$\overline{diversity}$	*level*
F01-03	8,579	0.11	0.09	0.01	0.04
F04-07	10,915	0.09	0.27	0.01	0.03
F08	2,338	0.03	0.18	0.02	0.05
F09	4,309	0.09	0.15	0.02	0.04
F10+11	8,540	0.12	0.18	0.01	0.03
F12+43	8,028	0.08	0.29	0.01	0.02
F13	1,999	0.12	0.18	0.03	0.08
F14+41	2,328	0.04	0.34	0.01	0.04
F15	2,575	0.07	0.21	0.01	0.04
F16+45	4,752	0.13	0.11	0.02	0.06
F17	1,894	0.06	0.11	0.01	0.06
F18+19	39,844	0.08	0.21	0.00	0.01
F20	4,894	0.09	0.22	0.01	0.05
F21+42	404	0.10	0.19	0.05	0.11
F22	2,279	0.14	0.08	0.03	0.09
F23	21,256	0.09	0.19	0.00	0.01
F24	3,932	0.13	0.17	0.02	0.03
F25	4,070	0.07	0.15	0.01	0.03
F26	1,924	0.10	0.13	0.03	0.06
F30	629	0.05	0.15	0.03	0.09
F31	1,498	0.15	0.16	0.03	0.07
F32	837	0.14	0.40	0.01	0.08
F33+34+36	2,164	0.07	0.45	0.01	0.03
F35	192	0.03	0.21	0.05	0.24
F40	1,185	0.08	0.10	0.04	0.09
F44	1,212	0.11	0.12	0.04	0.10
F46	1,951	0.06	0.23	0.01	0.04

OA.5 Details on the Experiment

Figure OA.2 below gives more details on how we generated the random regulations by showing all the possible attributes for each asset class and the values these attributes can take.

We then give more details on the selection of participants to the experiment and their incentives. The participants were all students of the MSc in International Finance of HEC Paris, class of 2020-2021, who volunteered to take part in the experiment. They had taken an 18-hour course on "Economics of Financial Regulation", which included in particular a description of the Basel I framework and a short example of how to compute risk-weighted capital requirements. Importantly, the course did not discuss how to measure regulatory complexity, so that there was no "priming" of the students.

Students were incentivized with bonus points in their course on "Economics of Financial Regulation": (i) 2 bonus points for completing the experiment, regardless of performance and (ii) 1/3 bonus point for each correct computation. Since there were 9 computations in total, students could obtain up to 5 bonus points, compared to 100 points for the final exam. This scheme served as an incentive to participate in the experiment and try to get a correct answer. As a result, 125 out of 191 students participated in the experiment. After excluding from the analysis 7 students who mistakenly took the test several times, and whose answers are potentially affected by a learning effect, we have a sample of 118 participants who give answers on 9 randomly selected regulations each, for a total of 1,062 participant-question observations.

Finally, we give more details on the website used to conduct the experiment, and reproduce all the pages below. After an introductory page (Figure OA.3), the participant registers and gives some background information (Figure OA.4). The participant is then shown a screen with explanations about the experiment and how to compute capital requirements (Figure OA.5). The next screen is a "test-round", which is the same for all participants (Figure 3 in the main text). The computer screen is split vertically in two. On the right-hand side, there is a series of instructions that mimick a Basel-I like capital regulation. On the left-hand side, there is a simplified bank balance sheet with details about the different assets of the bank that correspond to the regulation. The participant has to compute the risk-weighted assets of the bank following the instructions. We record the answer given by the participant (and hence whether it is correct), as well as the time taken to answer.

If the answer to the test-round is correct, the participant is notified that he/she found the correct answer (Figure OA.4). If the answer is wrong, the participant is told so (Figure OA.7). In both cases, the participant is given an explanation on how to compute the correct answer, and then moves to the second round. The second round is similar to the first one, except that the regulation is drawn randomly from our set of randomly generated regulations. Moreover, the participant doesn't receive any feedback on his/her answer. The experiment is then repeated for a total of 10 rounds (including the first training round). The balance sheet displayed on the left-hand side is constant across rounds and across participants.

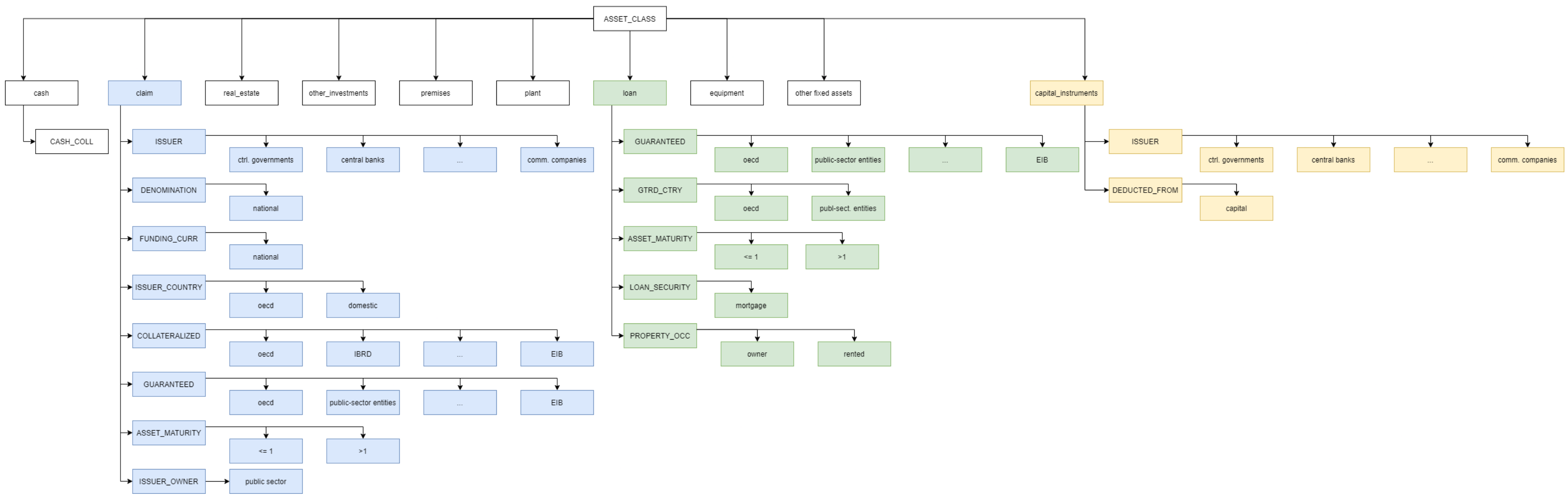


Figure OA.2: Decision tree for generating random regulations

This figure shows the decision tree with possible values for the "random regulation generator". Each box is an Economic Operand. Arrows indicate possible values of an Economic Operand in a bank's balance sheet. Colors are used for visual guidance only to make it easier to see which choices are part of the same branch in the tree.

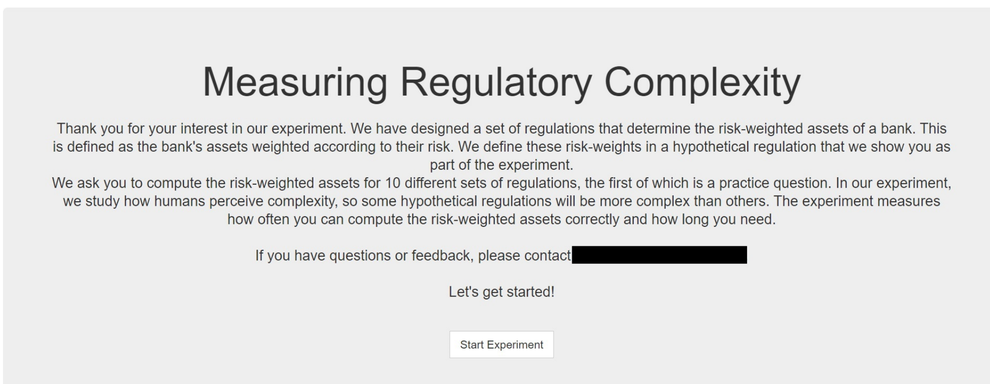


Figure OA.3: Online experiment - Welcome page

This figure is a screenshot of the welcome page of the website we use for the experiments.

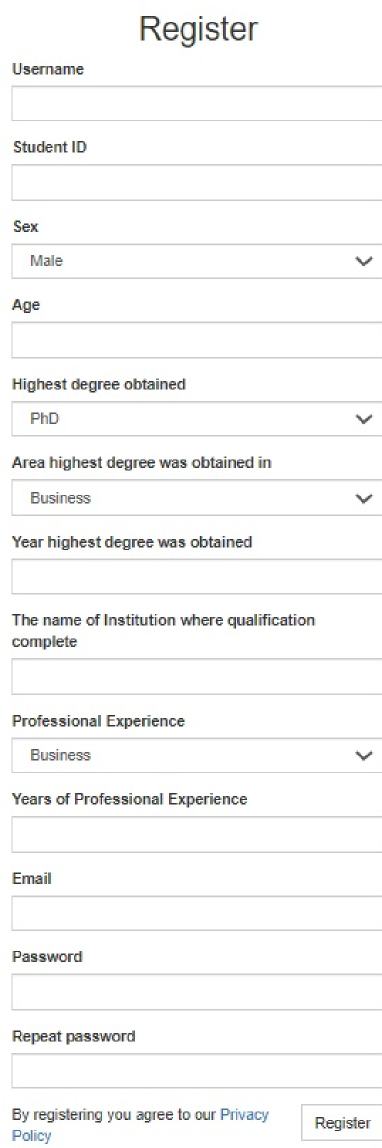


Figure OA.4: Online experiment - Registration page

This figure is a screenshot of the registration page of the website we use for the experiments.

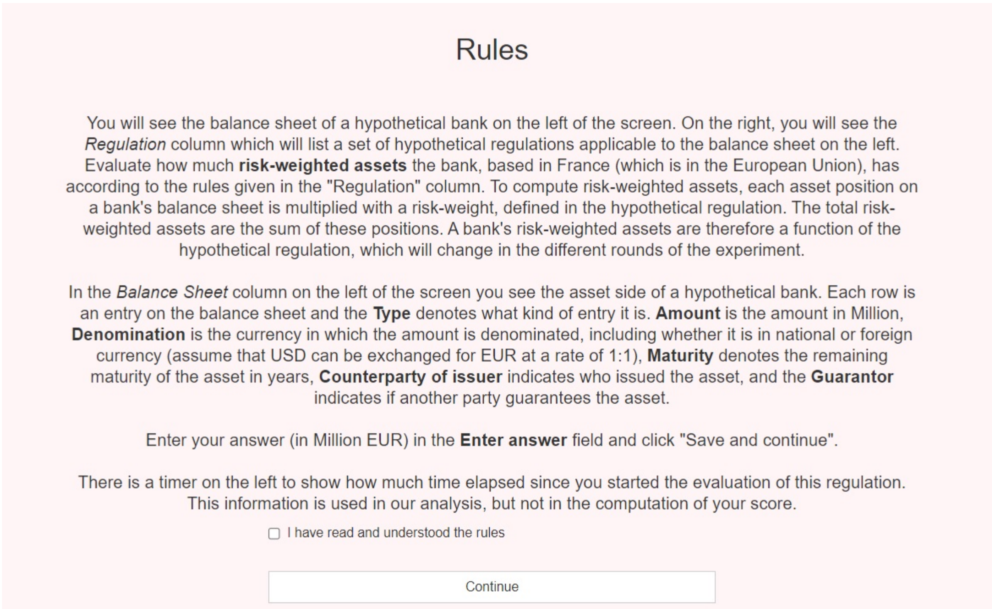


Figure OA.5: Online experiment - Instructions page

This figure is a screenshot of the instructions page of the website we use for the experiments.

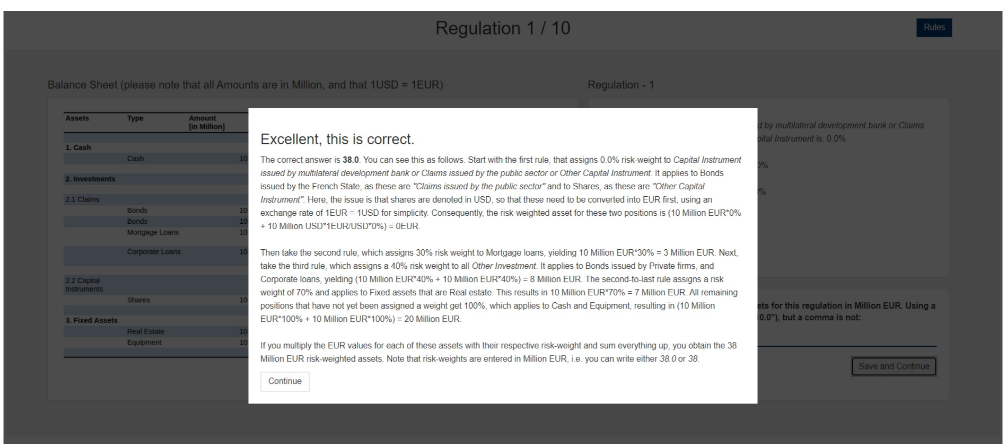


Figure OA.6: Online experiment - Feedback after correct answer in the test round

This figure is a screenshot of the feedback page of the website we use for the experiments: the participant is shown this page after a correct answer in the test round.

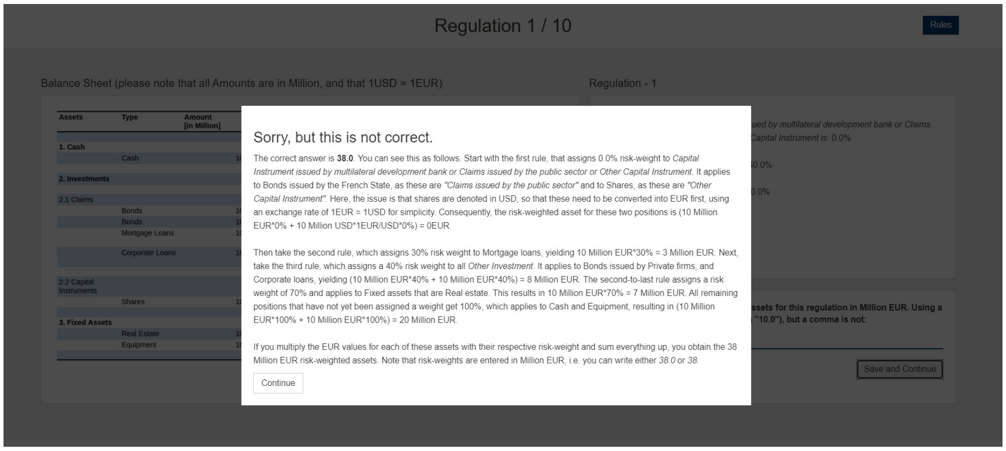


Figure OA.7: Online experiment - Feedback after wrong answer in the test round
This figure is a screenshot of the feedback page of the website we use for the experiments: the participant is shown this page after an incorrect answer in the test round.

OA.6 Experiments - Regulation-level Analysis

In this section we replicate the results of Section 3.1 after aggregating the answers of all participants at the regulation level. For each regulation $j \in \{1, 2..., 38\}$ we compute the average proportion $mistake_j$ of incorrect answers and the average time taken $time_j$ (excluding incorrect answers and times above 579 seconds, as in the main analysis of Section 3.1). This gives us a database with 38 observations, one for each regulation. We then run OLS regressions of $mistake_j$ and $time_j$ on the same measures of complexity as in Section 3.1. Table OA.10 below corresponds to Table 5, and Table OA.11 corresponds to 6. The results are qualitatively the same as in our preferred specification at the participant-question level.

Table OA.10: Correlation of mistakes with measures of complexity, beyond *length*, regulation-level

We estimate the OLS regression:

$$mistake_j = \alpha + \beta\mu(R_j) + \epsilon_j,$$

where $mistake_j$ is the fraction of incorrect answers for regulation j, $\mu(R_j)$ is a characteristic or vector of characteristics of regulation j, and ϵ_j is an error term. The table reports the coefficient(s) β when including different measures as regressors, t-statistics (in brackets) computed with robust standard errors, and adjusted R^2. $*$, $**$, and $***$ denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Reminder of the different measures: *length* is the total number of words, *cyclomatic* is the total number of logical operators, *quantity* is the total number of regulatory operators, *potential* is the number of unique operands, *diversity* is the number of unique operators, and $level = potential/length$.

	(1)	(2)	(3)	(4)	(5)	(6)
length	0.007***	0.011**	-0.002	0.005	0.012***	0.006**
	(4.34)	(2.69)	(-0.72)	(1.00)	(3.96)	(2.43)
cyclomatic		-0.015				
		(-1.05)				
quantity			0.111***			
			(3.41)			
potential				0.006		
				(0.50)		
diversity					-0.083*	
					(-1.97)	
level						-0.225
						(-0.55)
Adjusted-R^2	0.247	0.248	0.388	0.231	0.307	0.233

Table OA.11: Correlation of time taken with measures of complexity, beyond *length*, regulation-level

We estimate the OLS regression (8):

$$time_j = \alpha + \beta\mu(R_j) + \epsilon_j,$$

where $time_j$ is the average time in seconds that participants to answer regulation j, $\mu(R_j)$ is a characteristic or vector of characteristics of regulation j, and ϵ_j is an error term. The table reports the coefficient(s) β when including different measures as regressors, t-statistics (in brackets) computed with robust standard errors, and adjusted R^2. $time_j$ is computed on a sample restricted to correct answers with *time* below the 99th percentile. $*$, $**$, and $***$ denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Reminder of the different measures: *length* is the total number of words, *cyclomatic* is the total number of logical operators, *quantity* is the total number of regulatory operators, *potential* is the number of unique operands, *diversity* is the number of unique operators, and $level = potential/length$.

	(1)	(2)	(3)	(4)	(5)	(6)
length	3.256***	5.549***	1.899*	0.324	3.286***	4.177***
	(6.56)	(4.42)	(1.82)	(0.26)	(6.40)	(4.69)
cyclomatic		-8.940**				
		(-2.06)				
quantity			16.403*			
			(1.81)			
potential				7.254**		
				(2.37)		
diversity					-0.487	
					(-0.08)	
level						189.632*
						(1.71)
Adjusted-R^2	0.573	0.638	0.596	0.647	0.560	0.613

OA.7 Experiments - Alternative Specifications

In this section we report the results of alternative specifications for the regressions in Tables 5 and 6. Tables OA.12 and OA.15 reproduce the results without round-fixed effects. Tables OA.13 and OA.16 use as explanatory variables the number of (total and unique) operands and operators, and confirm that separating words into these two categories is useful to predict *mistake* and *time*, respectively. Tables OA.14 and OA.17 show the coefficients obtained on the different measures without controlling for length. Table OA.18 reports the results of estimating a linear probability model instead of the probit model of Table 5.

Table OA.12: Correlation of mistakes with measures of complexity, beyond *length*, no round-fixed effects

We estimate the probit regression (7):

$$\Pr(mistake_{i,t} = 1) = \Phi(\alpha + \beta\mu(R_{i,t}) + \gamma_i),$$

where $mistake_{i,t}$ is a dummy variable equal to 1 if participant i gave an incorrect answer in round t, $\mu(R_{i,t})$ is a characteristic or vector of characteristics of the regulation given to participant i in round t, γ_i is a participant-fixed effect, and Φ is the standardized Gaussian cdf. The table reports the coefficient(s) β when including different measures as regressors, t-statistics (in brackets) computed with robust standard errors, and Mc Fadden's Pseudo-R^2. $*$, $**$, and $***$ denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Reminder of the different measures: *length* is the total number of words, *cyclomatic* is the total number of logical operators, *quantity* is the total number of regulatory operators, *potential* is the number of unique operands, *diversity* is the number of unique operators, and $level = potential/length$.

	(1)	(2)	(3)	(4)	(5)	(6)
length	0.037***	0.055***	-0.004	0.027***	0.064***	0.032***
	(8.63)	(6.23)	(-0.53)	(2.59)	(8.37)	(5.77)
cyclomatic		-0.072**				
		(-2.39)				
quantity			0.533***			
			(6.10)			
potential				0.023		
				(0.99)		
diversity					-0.431***	
					(-4.49)	
level						-1.009
						(-1.28)
Pseudo-R^2	0.228	0.233	0.260	0.229	0.244	0.229

Table OA.13: Correlation of mistakes with operands and operators

We estimate the probit regression (7):

$$\Pr(mistake_{i,t} = 1) = \Phi(\alpha + \beta\mu(R_{i,t}) + \gamma_i + \delta_t),$$

where $mistake_{i,t}$ is a dummy variable equal to 1 if participant i gave an incorrect answer in round t, $\mu(R_{i,t})$ is a characteristic or vector of characteristics of the regulation given to participant i in round t, γ_i and δ_t are participant- and round-fixed effects, and Φ is the standardized Gaussian cdf. The table reports the coefficient(s) β when including different characteristics as regressors, t-statistics (in brackets) computed with robust standard errors, and Mc Fadden's Pseudo-R^2. $*$, $**$, and $***$ denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Reminder of the different measures: *length* is the total number of words, N_{OD} is the total number of operands, N_{OR} the total number of operators, η_{OD} the number of unique operands, and η_{OR} the number of unique operators.

	(1)	(2)	(3)	(4)	(5)
length	0.037***				0.044***
	(8.67)				(3.99)
N_{OD}		0.062***			
		(4.10)			
N_{OR}		-0.006			
		(-0.23)			
$\eta_{OD} + \eta_{OR}$			0.065***		
			(7.92)		
η_{OD}				0.144***	0.070***
				(8.31)	(2.74)
η_{OR}				-0.451***	-0.559***
				(-4.54)	(-5.14)
Pseudo-R^2	0.243	0.246	0.232	0.253	0.266

Table OA.14: Correlation of mistakes with measures of complexity

We estimate the probit regression (7):

$$\Pr(mistake_{i,t} = 1) = \Phi(\alpha + \beta\mu(R_{i,t}) + \gamma_i + \delta_t),$$

where $mistake_{i,t}$ is a dummy variable equal to 1 if participant i gave an incorrect answer in round t, $\mu(R_{i,t})$ is a characteristic or vector of characteristics of the regulation given to participant i in round t, γ_i and δ_t are participant- and round-fixed effects, and Φ is the standardized Gaussian cdf. The table reports the coefficient(s) β when including different measures as regressors, t-statistics (in brackets) computed with robust standard errors, and Mc Fadden's Pseudo-R^2. $*$, $**$, and $***$ denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Reminder of the different measures: *length* is the total number of words, *cyclomatic* is the total number of logical operators, *quantity* is the total number of regulatory operators, *potential* is the number of unique operands, *diversity* is the number of unique operators, and $level = potential/length$.

	(1)	(2)	(3)	(4)	(5)	(6)
length	0.037***					
	(8.67)					
cyclomatic		0.095***				
		(6.65)				
quantity			0.507***			
			(10.39)			
potential				0.078***		
				(8.30)		
diversity					0.243***	
					(4.53)	
level						-4.061***
						(-6.76)
Pseudo-R^2	0.243	0.217	0.277	0.237	0.198	0.221

Table OA.15: Correlation of time taken with measures of complexity, beyond *length*, no round-fixed effects

We estimate the OLS regression (8):

$$time_{i,t} = \alpha + \beta\mu(R_{i,t}) + \gamma_i + \epsilon_{i,t},$$

where $time_{i,t}$ is the time in seconds that participant i took to answer in round t, $\mu(R_{i,t})$ is a characteristic or vector of characteristics of the regulation given to participant i in round t, γ_i is a participant-fixed effect, and $\epsilon_{i,t}$ is an error term. The table reports the coefficient(s) β when including different measures as regressors, t-statistics (in brackets) computed with robust standard errors, and adjusted R^2. The sample is restricted to correct answers with *time* below the 99th percentile. $*$, $**$, and $***$ denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Reminder of the different measures: *length* is the total number of words, *cyclomatic* is the total number of logical operators, *quantity* is the total number of regulatory operators, *potential* is the number of unique operands, *diversity* is the number of unique operators, and $level = potential/length$.

	(1)	(2)	(3)	(4)	(5)	(6)
length	3.425***	5.563***	2.099***	0.311	3.089***	4.330***
	(13.73)	(9.21)	(3.26)	(0.42)	(5.82)	(11.42)
cyclomatic		-8.699***				
		(-3.61)				
quantity			15.036**			
			(2.40)			
potential				7.505***		
				(4.49)		
diversity					5.257	
					(0.78)	
level						187.402***
						(3.41)
Adjusted-R^2	0.360	0.379	0.367	0.384	0.360	0.373

Table OA.16: Correlation of time taken with operands and operators

We estimate the OLS regression (8):

$$time_{i,t} = \alpha + \beta\mu(R_{i,t}) + \gamma_i + \delta_t + \epsilon_{i,t},$$

where $time_{i,t}$ is the time in seconds that participant i took to answer in round t, $\mu(R_{i,t})$ is a characteristic or vector of characteristics of the regulation given to participant i in round t, γ_i and δ_t are participant- and round-fixed effects, and $\epsilon_{i,t}$ is an error term. The table reports the coefficient(s) β when including different characteristics as regressors, t-statistics (in brackets) computed with robust standard errors, and adjusted R^2. The sample is restricted to correct answers with *time* below the 99th percentile. $*$, $**$, and $***$ denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Reminder of the different measures: N_{OD} is the total number of operands, N_{OR} the total number of operators, η_{OD} the number of unique operands, and η_{OR} the number of unique operators.

	(1)	(2)	(3)	(4)	(5)
$length = N_{OD} + N_{OR}$	3.388***				0.654
	(14.63)				(0.93)
N_{OD}		7.102***			
		(7.42)			
N_{OR}		-3.193*			
		(-1.90)			
$\eta_{OD} + \eta_{OR}$			6.886***		
			(15.21)		
η_{OD}				8.452***	7.272***
				(7.38)	(4.32)
η_{OR}				-3.309	-4.550
				(-0.53)	(-0.70)
Adjusted-R^2	0.445	0.461	0.462	0.464	0.464

Table OA.17: Correlation of time taken with measures of complexity

We estimate the OLS regression (8):

$$time_{i,t} = \alpha + \beta\mu(R_{i,t}) + \gamma_i + \delta_t + \epsilon_{i,t},$$

where $time_{i,t}$ is the time in seconds that participant i took to answer in round t, $\mu(R_{i,t})$ is a characteristic or vector of characteristics of the regulation given to participant i in round t, γ_i and δ_t are participant- and round-fixed effects, and $\epsilon_{i,t}$ is an error term. The table reports the coefficient(s) β when including different measures as regressors, t-statistics (in brackets) computed with robust standard errors, and adjusted R^2. The sample is restricted to correct answers with *time* below the 99th percentile. $*$, $**$, and $***$ denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Reminder of the different measures: *length* is the total number of words, *cyclomatic* is the total number of logical operators, *quantity* is the total number of regulatory operators, *potential* is the number of unique operands, *diversity* is the number of unique operators, and $level = potential/length$.

	(1)	(2)	(3)	(4)	(5)	(6)
length	3.388***					
	(14.63)					
cyclomatic		9.338***				
		(9.64)				
quantity			32.996***			
			(14.55)			
potential				7.965***		
				(15.09)		
diversity					39.316***	
					(12.78)	
level						-265.538***
						(-6.90)
Adjusted-R^2	0.445	0.363	0.433	0.465	0.403	0.308

Table OA.18: Correlation of mistakes with measures of complexity, beyond *length*, linear probability model

We estimate the linear probability model:

$$mistake_{i,t} = \alpha + \beta\mu(R_{i,t}) + \gamma_i + \delta_t + \epsilon_{i,t},$$

where $mistake_{i,t}$ is a dummy variable equal to 1 if participant i gave an incorrect answer in round t, $\mu(R_{i,t})$ is a characteristic or vector of characteristics of the regulation given to participant i in round t, γ_i and δ_t are participant- and round-fixed effects, and $\epsilon_{i,t}$ is an error term. The table reports the coefficient(s) β when including different measures as regressors, t-statistics (in brackets) computed with robust standard errors, and adjusted R^2. $*$, $**$, and $***$ denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Reminder of the different measures: *length* is the total number of words, *cyclomatic* is the total number of logical operators, *quantity* is the total number of regulatory operators, *potential* is the number of unique operands, *diversity* is the number of unique operators, and $level = potential/length$.

	(1)	(2)	(3)	(4)	(5)	(6)
length	0.008***	0.012***	-0.001	0.006**	0.014***	0.007***
	(8.28)	(5.24)	(-0.62)	(2.24)	(8.21)	(4.80)
cyclomatic		-0.015*				
		(-1.86)				
quantity			0.110***			
			(5.30)			
potential				0.004		
				(0.72)		
diversity					-0.094***	
					(-4.09)	
level						-0.258
						(-1.24)
Adjusted-R^2	0.294	0.296	0.314	0.294	0.305	0.294

OA.8 Experiments - Alternative Filters

In this section we check that the results reported in Table 6 are robust to different ways of filtering out observations that are likely to be affected by measurement error. We report results on the following alternative specifications: (i) we winsorize outliers at 579 seconds instead of trimming these observations (Table OA.19) ; (ii) we keep the outliers but exclude incorrect answers (Table OA.20) ; (iii) we keep incorrect answers and exclude outliers with time above 579 seconds (Table OA.21) ; (iv) we keep all observations (Table OA.22).

The results of Table 6 on *potential* are robust across all specifications. However, the positive coefficient on *quantity* when controlling for *length* is no longer significant at the 10% level in specifications (i), (ii), and (iv). More precisely, the coefficient on *quantity* drops from 13.072 with a t-stat of 2.30 in Table 6 to 7.373 with a t-stat of 1.07 in the most adverse specification where all outliers are kept (specification (ii)). While we believe that observations with *time* > 579 are with a very high probability contaminated with measurement error and should be excluded, it is certainly the case that the impact of *quantity* on *time* is statistically weaker than the impact of *potential* and not always robust to how outliers are treated.

Table OA.19: Correlation of time taken with measures of complexity, beyond *length*, outliers winsorized

We estimate the OLS regression (8):

$$time_{i,t} = \alpha + \beta\mu(R_{i,t}) + \gamma_i + \delta_t + \epsilon_{i,t},$$

where $time_{i,t}$ is the time in seconds that participant i took to answer in round t, $\mu(R_{i,t})$ is a characteristic or vector of characteristics of the regulation given to participant i in round t, γ_i and δ_t are participant- and round-fixed effects, and $\epsilon_{i,t}$ is an error term. The table reports the coefficient(s) β when including different measures as regressors, t-statistics (in brackets) computed with robust standard errors, and adjusted R^2. The sample is restricted to correct answers. Observations with $time_{i,t}$ above the 99th percentile are winsorized. $*$, $**$, and $***$ denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Reminder of the different measures: *length* is the total number of words, *cyclomatic* is the total number of logical operators, *quantity* is the total number of regulatory operators, *potential* is the number of unique operands, *diversity* is the number of unique operators, and $level = potential/length$.

	(1)	(2)	(3)	(4)	(5)	(6)
length	3.713***	5.091***	2.952***	1.437*	3.585***	4.397***
	(13.81)	(8.91)	(4.37)	(1.87)	(5.80)	(11.72)
cyclomatic		-5.569**				
		(-2.36)				
quantity			8.692			
			(1.37)			
potential				5.497***		
				(3.33)		
diversity					2.018	
					(0.26)	
level						141.549***
						(2.68)
Adjusted-R^2	0.447	0.453	0.448	0.457	0.446	0.453

Table OA.20: Correlation of time taken with measures of complexity, beyond *length*, outliers included

We estimate the OLS regression (8):

$$time_{i,t} = \alpha + \beta\mu(R_{i,t}) + \gamma_i + \delta_t + \epsilon_{i,t},$$

where $time_{i,t}$ is the time in seconds that participant i took to answer in round t, $\mu(R_{i,t})$ is a characteristic or vector of characteristics of the regulation given to participant i in round t, γ_i and δ_t are participant- and round-fixed effects, and $\epsilon_{i,t}$ is an error term. The table reports the coefficient(s) β when including different measures as regressors, t-statistics (in brackets) computed with robust standard errors, and adjusted R^2. The sample is restricted to correct answers. Observations with $time_{i,t}$ above the 99th percentile are included. $*$, $**$, and $***$ denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Reminder of the different measures: *length* is the total number of words, *cyclomatic* is the total number of logical operators, *quantity* is the total number of regulatory operators, *potential* is the number of unique operands, *diversity* is the number of unique operators, and $level = potential/length$.

	(1)	(2)	(3)	(4)	(5)	(6)
length	3.806***	5.100***	3.160***	1.791**	3.759***	4.427***
	(13.25)	(8.41)	(4.28)	(2.09)	(5.73)	(11.55)
cyclomatic		-5.229**				
		(-2.07)				
quantity			7.373			
			(1.07)			
potential				4.867***		
				(2.71)		
diversity					0.739	
					(0.09)	
level						128.356**
						(2.31)
Adjusted-R^2	0.434	0.439	0.435	0.441	0.433	0.438

Table OA.21: Correlation of time taken with measures of complexity, beyond *length*, incorrect answers included

We estimate the OLS regression (8):

$$time_{i,t} = \alpha + \beta\mu(R_{i,t}) + \gamma_i + \delta_t + \epsilon_{i,t},$$

where $time_{i,t}$ is the time in seconds that participant i took to answer in round t, $\mu(R_{i,t})$ is a characteristic or vector of characteristics of the regulation given to participant i in round t, γ_i and δ_t are participant- and round-fixed effects, and $\epsilon_{i,t}$ is an error term. The table reports the coefficient(s) β when including different measures as regressors, t-statistics (in brackets) computed with robust standard errors, and adjusted R^2. The sample is restricted to observations with $time_{i,t}$ below the 99th percentile, but incorrect answers are included. $*$, $**$, and $***$ denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Reminder of the different measures: *length* is the total number of words, *cyclomatic* is the total number of logical operators, *quantity* is the total number of regulatory operators, *potential* is the number of unique operands, *diversity* is the number of unique operators, and $level = potential/length$.

	(1)	(2)	(3)	(4)	(5)	(6)
length	3.236***	4.631***	2.238***	1.611***	3.131***	3.592***
	(16.81)	(10.82)	(5.51)	(3.14)	(8.37)	(12.88)
cyclomatic		-5.499***				
		(-3.55)				
quantity			11.868***			
			(2.92)			
potential				3.972***		
				(3.41)		
diversity					1.707	
					(0.35)	
level						74.121*
						(1.92)
Adjusted-R^2	0.488	0.495	0.492	0.494	0.487	0.489

Table OA.22: Correlation of time taken with measures of complexity, beyond *length*, all observations included

We estimate the OLS regression (8):

$$time_{i,t} = \alpha + \beta\mu(R_{i,t}) + \gamma_i + \delta_t + \epsilon_{i,t},$$

where $time_{i,t}$ is the time in seconds that participant i took to answer in round t, $\mu(R_{i,t})$ is a characteristic or vector of characteristics of the regulation given to participant i in round t, γ_i and δ_t are participant- and round-fixed effects, and $\epsilon_{i,t}$ is an error term. The table reports the coefficient(s) β when including different measures as regressors, t-statistics (in brackets) computed with robust standard errors, and adjusted R^2. The sample includes all observations. $*$, $**$, and $***$ denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Reminder of the different measures: *length* is the total number of words, *cyclomatic* is the total number of logical operators, *quantity* is the total number of regulatory operators, *potential* is the number of unique operands, *diversity* is the number of unique operators, and $level = potential/length$.

	(1)	(2)	(3)	(4)	(5)	(6)
length	3.652***	4.507***	3.049***	2.520***	3.929***	3.916***
	(15.15)	(9.76)	(5.76)	(3.98)	(7.49)	(12.64)
cyclomatic		-3.345*				
		(-1.84)				
quantity			7.232			
			(1.47)			
potential				2.779**		
				(2.10)		
diversity					-4.589	
					(-0.70)	
level						54.825
						(1.33)
Adjusted-R^2	0.468	0.469	0.468	0.470	0.467	0.468

OA.9 Empirical validation - Alternative specifications

In this section we report the results of alternative specifications to explain *cost* (Table 7 in the paper). Table OA.23 estimates a fractional probit model instead of an OLS, and confirms that our results are robust to this specification. Table OA.24 uses as explanatory variables the number of (total and unique) operands and operators, and confirms that separating words into these two categories is useful to predict *cost*. Table OA.25 shows the coefficients obtained on the different measures without controlling for length. Table OA.26 reports the results of regression with measure × group interaction effects.

Table OA.23: Correlation of the fraction of banks reporting a high cost with measures of complexity, beyond *length*, fractional probit

We estimate the following fractional probit regression (Papke and Wooldridge, 1996):

$$\mathbb{E}(cost_{i,j}) = \Phi\left(\alpha + \beta\mu(R_j) + \gamma_2 medium_i + \gamma_3 small_i\right),$$

where $cost_{i,j}$ is the fraction of banks in group $i \in \{large, medium, small\}$ that report a high or medium-high cost for template j, $\mu(R_j)$ is a characteristic or vector of characteristics of template j, $medium_i$ and $small_i$ are dummy variables equal to 1 if $i = medium$ or $small$, respectively, Φ is the cdf. of the standardized normal distribution. The table reports the coefficient(s) β when including different measures as regressors, t-statistics (in brackets) computed with robust standard errors, and McFadden's pseudo-R^2. $*$, $**$, and $***$ denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Reminder of the different measures: *length* is the total number of words, *cyclomatic* is the total number of logical operators, *quantity* is the total number of regulatory operators, *potential* is the number of unique operands, *diversity* is the number of unique operators, and $level = potential/length$.

	(1)	(2)	(3)	(4)	(5)	(6)
length	0.00003***	0.00006***	0.00000	0.00003***	0.00003***	0.00003***
	(6.55)	(3.21)	(0.20)	(5.47)	(5.09)	(5.62)
cyclomatic		-0.00027				
		(-1.58)				
quantity			0.00016***			
			(4.05)			
potential				0.00059		
				(1.53)		
diversity					0.00020	
					(0.19)	
level						-1.20494
						(-1.62)
Pseudo-R^2	0.021	0.022	0.026	0.022	0.021	0.022

Table OA.24: Correlation of the fraction of banks reporting a high cost with operands and operators

We estimate the OLS regression (9):

$$cost_{i,j} = \alpha + \beta\mu(R_j) + \gamma_2 medium_i + \gamma_3 small_i + \epsilon_{i,j},$$

where $cost_{i,j}$ is the fraction of banks in group $i \in \{large, medium, small\}$ that report a high or medium-high cost for template j, $\mu(R_j)$ is a characteristic or vector of characteristics of template j, $medium_i$ and $small_i$ are dummy variables equal to 1 if $i = medium$ or $small$, respectively, and $\epsilon_{i,j}$ is an error term. The table reports the coefficient(s) β when including different measures as regressors, t-statistics (in brackets) computed with robust standard errors, and adjusted R^2. $*$, $**$, and $***$ denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Reminder of the different measures: N_{OD} is the total number of operands, N_{OR} the total number of operators, η_{OD} the number of unique operands, and η_{OR} the number of unique operators.

	(1)	(2)	(3)	(4)	(5)
$length$	0.00001***				0.00001***
	(6.64)				(5.85)
N_{OD}		-0.00002			
		(-1.44)			
N_{OR}		0.00006***			
		(4.03)			
$\eta_{OD} + \eta_{OR}$			0.00041***		
			(4.46)		
η_{OD}				0.00114***	0.00086***
				(3.21)	(2.63)
η_{OR}				-0.00151*	-0.00167**
				(-1.69)	(-2.00)
Adjusted-R^2	0.170	0.215	0.098	0.123	0.197

Table OA.25: Correlation of the fraction of banks reporting a high cost with measures of complexity

We estimate the OLS regression (9):

$$cost_{i,j} = \alpha + \beta\mu(R_j) + \gamma_2 medium_i + \gamma_3 small_i + \epsilon_{i,j},$$

where $cost_{i,j}$ is the fraction of banks in group $i \in \{large, medium, small\}$ that report a high or medium-high cost for template j, $\mu(R_j)$ is a characteristic or vector of characteristics of template j, $medium_i$ and $small_i$ are dummy variables equal to 1 if $i = medium$ or $small$, respectively, and $\epsilon_{i,j}$ is an error term. The table reports the coefficient(s) β when including different measures as regressors, t-statistics (in brackets) computed with robust standard errors, and adjusted R^2. $*$, $**$, and $***$ denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Reminder of the different measures: *length* is the total number of words, *cyclomatic* is the total number of logical operators, *quantity* is the total number of regulatory operators, *potential* is the number of unique operands, *diversity* is the number of unique operators, and $level = potential/length$.

	(1)	(2)	(3)	(4)	(5)	(6)
length	0.00001***					
	(6.64)					
cyclomatic		0.00011***				
		(6.92)				
quantity			0.00006***			
			(6.56)			
potential				0.00059***		
				(4.70)		
diversity					0.00108***	
					(3.30)	
level						-1.21227***
						(-3.96)
Adjusted-R^2	0.170	0.137	0.215	0.110	0.062	0.099

Table OA.26: Correlation of the fraction of banks reporting a high cost with measures of complexity, beyond *length*, with interaction effects

We estimate the OLS regression:

$$cost_{i,j} = \alpha + \gamma_2 medium_i + \gamma_3 small_i + \beta_1 \mu(R_j) \times large_i + \beta_2 \mu(R_j) \times medium_i + \beta_3 \mu(R_j) \times small_i + \epsilon_{i,j},$$

where $cost_{i,j}$ is the fraction of banks in group $i \in \{large, medium, small\}$ that report a high or medium-high cost for template j, $\mu(R_j)$ is a characteristic or vector of characteristics of template j, $large_i$, $medium_i$, and $small_i$ are dummy variables equal to 1 if $i = large$, $medium$, or $small$, respectively, and $\epsilon_{i,j}$ is an error term. The table reports the coefficients β_1, β_2, and β_3 when including different measures as regressors, t-statistics (in brackets) computed with robust standard errors, and adjusted R^2. $*$, $**$, and $***$ denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Reminder of the different measures: *length* is the total number of words, *cyclomatic* is the total number of logical operators, *quantity* is the total number of regulatory operators, *potential* is the number of unique operands, *diversity* is the number of unique operators, and $level = potential/length$.

	(1)	(2)	(3)
$length \times large$	0.00001***	-0.00000	0.00001***
	(4.18)	(-0.00)	(3.62)
$length \times medium$	0.00001***	-0.00001	0.00001***
	(3.51)	(-1.04)	(2.75)
$length \times small$	0.00001***	0.00001	0.00001***
	(3.89)	(0.86)	(3.40)
$quantity \times large$		0.00006***	
		(2.92)	
$quantity \times medium$		0.00009***	
		(3.69)	
$quantity \times small$		0.00004	
		(1.10)	
$potential \times large$			0.00013
			(0.58)
$potential \times medium$			0.00016
			(0.52)
$potential \times small$			0.00049**
			(2.12)
Adjusted-R^2	0.160	0.194	0.162

OA.10 Regulatory importance of ITS templates

In this section we give more details on our analysis in Section 3.3 of the importance given by regulators to different ITS templates. We also report alternative specifications of the regressions in Table 8, and shows that our results are robust.

Our analysis is based on Figures 12 and 13 in EBA (2021). Figure 12 reports for 34 templates or groups of templates how many of the 29 regulatory authorities consider the template "highly important", "important", "less important", "not important", or do not express any view. The figure reports the numbers in each categories as a bar chart, so our first step is to manually translate the bars into numbers and save them in a dataset. We follow the same process for Figure 13, which contains 23 templates.

We then need to match these 57 templates with the templates used in Section 3.2 for which we have complexity measures. First, we observe that some templates appear several times. For instance F44 appears 3 times, as "Defined benefit plans (F 44)", "Staff expenses by type of benefits (F 44)", and "Staff expenses by category of staff (F 44)". We merge these duplicates and keep the average answers across these three lines (in all such cases the answers are actually very close to each other). Second, we keep only templates that also appear in the bank questionnaire used in Section 3.2. This gives us a dataset with 43 templates or groups of templates, for which we have the measures *length*, *cyclomatic*, *quantity*, *potential*, *diversity*, and *level*.

We then compute the average importance attributed by regulators to each of the 43 templates. We turn each answer "highly important" into a numerical score of 3, "important" into a score of 2, "less important" into a score of 1, and "not important" into a score of 0. We compute $importance_j$ as the average score for template j among regulators who expressed a view. As alternative measures, we also compute:

- $important1_j$ is the proportion, among regulators who expressed a view, of answers "highly important".

- $important2_j$ is the proportion, among regulators who expressed a view, of answers "highly important" or "important".

- $notimportant1_j$ is the proportion, among regulators who expressed a view, of answers "not important".

- $notimportant2_j$ is the proportion, among regulators who expressed a view, of answers "not important" or "less important".

- top_j is a dummy variable equal to 1 if template j is mentioned on Figure 12 of EBA (2021) rather than on Figure 13.

In the following we report the results of regressions using these different measures of importance as dependent variables instead of $importance_j$, as in Table 8. The results are extremely similar, with of course the signs inverted when using $notimportant1_j$ or $notimportant2_j$.

Table OA.27: Correlation of the importance given to templates by regulators with measures of complexity, beyond *length*, only answers "highly important"

We estimate the OLS regression:

$$important1_j = \alpha + \beta\mu(R_j) + \epsilon_j,$$

where $important1_j$ is the proportion for template j, among regulators who expressed a view, of answers "highly important". $\mu(R_j)$ is a characteristic or vector of characteristics of template j, and ϵ_j is an error term. The table reports the coefficient(s) β when including different measures as regressors, t-statistics (in brackets) computed with robust standard errors, and adjusted R^2. $*$, $**$, and $***$ denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Reminder of the different measures: *length* is the total number of words, *cyclomatic* is the total number of logical operators, *quantity* is the total number of regulatory operators, *potential* is the number of unique operands, *diversity* is the number of unique operators, and $level = potential/length$.

	(1)	(2)	(3)	(4)	(5)	(6)
length	0.00003***	-0.00002	0.00005***	0.00001*	0.00001	0.00003**
	(2.70)	(-0.96)	(3.17)	(1.72)	(1.67)	(2.33)
cyclomatic		0.00057**				
		(2.34)				
quantity			-0.00013*			
			(-1.99)			
potential				0.00157***		
				(3.36)		
diversity					0.00347***	
					(2.93)	
level						-0.15953
						(-0.22)
Adjusted-R^2	0.103	0.144	0.123	0.201	0.177	0.082

Table OA.28: Correlation of the importance given to templates by regulators with measures of complexity, beyond *length*, only answers "highly important" or "important"

We estimate the OLS regression:

$$important2_j = \alpha + \beta\mu(R_j) + \epsilon_j,$$

where $important2_j$ is the proportion for template j, among regulators who expressed a view, of answers "highly important" or "important". $\mu(R_j)$ is a characteristic or vector of characteristics of template j, and ϵ_j is an error term. The table reports the coefficient(s) β when including different measures as regressors, t-statistics (in brackets) computed with robust standard errors, and adjusted R^2. $*$, $**$, and $***$ denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Reminder of the different measures: *length* is the total number of words, *cyclomatic* is the total number of logical operators, *quantity* is the total number of regulatory operators, *potential* is the number of unique operands, *diversity* is the number of unique operators, and $level = potential/length$.

	(1)	(2)	(3)	(4)	(5)	(6)
length	0.00002**	-0.00002	0.00004***	0.00000	0.00000	0.00002*
	(2.29)	(-1.32)	(3.02)	(1.09)	(0.94)	(1.90)
cyclomatic		0.00047***				
		(2.71)				
quantity			-0.00009**			
			(-2.04)			
potential				0.00126***		
				(4.29)		
diversity					0.00298***	
					(3.66)	
level						-0.27862
						(-0.54)
Adjusted-R^2	0.083	0.142	0.104	0.206	0.195	0.068

Table OA.29: Correlation of the importance given to templates by regulators with measures of complexity, beyond *length*, only answers "not important"

We estimate the OLS regression:

$$notimportant1_j = \alpha + \beta\mu(R_j) + \epsilon_j,$$

where $notimportant1_j$ is the proportion for template j, among regulators who expressed a view, of answers "not important". $\mu(R_j)$ is a characteristic or vector of characteristics of template j, and ϵ_j is an error term. The table reports the coefficient(s) β when including different measures as regressors, t-statistics (in brackets) computed with robust standard errors, and adjusted R^2. $*$, $**$, and $***$ denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Reminder of the different measures: *length* is the total number of words, *cyclomatic* is the total number of logical operators, *quantity* is the total number of regulatory operators, *potential* is the number of unique operands, *diversity* is the number of unique operators, and $level = potential/length$.

	(1)	(2)	(3)	(4)	(5)	(6)
length	-0.00001*	0.00001	-0.00001**	-0.00000	-0.00000	-0.00000
	(-2.00)	(1.43)	(-2.59)	(-0.80)	(-0.76)	(-1.62)
cyclomatic		-0.00020**				
		(-2.46)				
quantity			0.00004*			
			(1.69)			
potential				-0.00044***		
				(-3.22)		
diversity					-0.00098**	
					(-2.62)	
level						0.14808
						(0.58)
Adjusted-R^2	0.038	0.084	0.051	0.097	0.081	0.024

Table OA.30: Correlation of the importance given to templates by regulators with measures of complexity, beyond *length*, only answers "not important" and "less important"

We estimate the OLS regression:

$$notimportant2_j = \alpha + \beta\mu(R_j) + \epsilon_j,$$

where $notimportant2_j$ is the proportion for template j, among regulators who expressed a view, of answers "not important" or "less important". $\mu(R_j)$ is a characteristic or vector of characteristics of template j, and ϵ_j is an error term. The table reports the coefficient(s) β when including different measures as regressors, t-statistics (in brackets) computed with robust standard errors, and adjusted R^2. $*$, $**$, and $***$ denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Reminder of the different measures: *length* is the total number of words, *cyclomatic* is the total number of logical operators, *quantity* is the total number of regulatory operators, *potential* is the number of unique operands, *diversity* is the number of unique operators, and $level = potential/length$.

	(1)	(2)	(3)	(4)	(5)	(6)
length	-0.00002**	0.00002	-0.00004***	-0.00000	-0.00000	-0.00002*
	(-2.29)	(1.33)	(-3.04)	(-1.08)	(-0.93)	(-1.90)
cyclomatic		-0.00047***				
		(-2.72)				
quantity			0.00010**			
			(2.07)			
potential				-0.00126***		
				(-4.30)		
diversity					-0.00298***	
					(-3.67)	
level						0.27648
						(0.54)
Adjusted-R^2	0.083	0.142	0.104	0.206	0.194	0.067

:

Table OA.31: Correlation of the importance given to templates by regulators with measures of complexity, beyond *length*, probit model

We estimate the probit regression:

$$\Pr(top_j = 1) = \Phi(\alpha + \beta\mu(R_j)),$$

where top_j is a dummy variable equal to 1 if template j is mentioned on Figure 12 of EBA (2021). $\mu(R_j)$ is a characteristic or vector of characteristics of template j, and Φ is the standardized Gaussian cdf. The table reports the coefficient(s) β when including different measures as regressors, t-statistics (in brackets) computed with robust standard errors, and Mc Fadden's Pseudo-R^2. $*$, $**$, and $***$ denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Reminder of the different measures: *length* is the total number of words, *cyclomatic* is the total number of logical operators, *quantity* is the total number of regulatory operators, *potential* is the number of unique operands, *diversity* is the number of unique operators, and $level = potential/length$.

	(1)	(2)	(3)	(4)	(5)	(6)
length	0.00023**	-0.00008	0.00050***	0.00006	0.00009	0.00028**
	(2.30)	(-0.54)	(2.96)	(0.53)	(0.76)	(2.22)
cyclomatic		0.00495***				
		(2.83)				
quantity			-0.00087**			
			(-2.05)			
potential				0.00947**		
				(2.30)		
diversity					0.02211**	
					(2.14)	
level						2.24447
						(0.89)
Pseudo-R^2	0.156	0.276	0.229	0.242	0.237	0.167

OA.11 A model of Risk-sensitivity and Simplicity

This Section proposes a simplified model to think about the trade-off between risk-sensitivity and simplicity mentioned in BCBS (2013). While the model is still too simple to be applied for policy, it shows that a quantitative analysis of this trade-off is possible, and which ingredients are required. Unless mentioned otherwise, the notations of the model follow Section 1.1.

We consider a bank with 1 in assets, that can be financed either with deposits D or equity E. In case the bank fails, depositors are reimbursed by the government using public funds, which have a marginal cost of $1 + \lambda$. These losses can be mitigated by asking the bank to use more equity, but we take as given that equity has a marginal social cost of $1 + \delta$.

There are different asset classes $x \in \{1, ..., N\}$. The bank starts with an asset of type x with probability $\nu(x)$. With probability p, the economy is growing and asset x pays $r(x)$. With probability $1 - p$, the economy enters a recession and the asset pays only $1 - \ell(x)$, i.e., the bank makes a loss of $\ell(x)$ on its investment. If $E < \ell(x)$ the bank defaults, and the government has to repay $D - (1 - \ell(x)) = \ell(x) - E$ to the depositors.

The regulator determines the minimum capital requirements y for the bank. As will become clear below, it is in the bank's interest to bind these capital requirements, so that $E = y$ and $D = 1 - y$.

We assume that the social cost of capital is lower than the expected gain of reducing losses to the public sector:

$$\lambda(1 - p) > \delta. \tag{OA.1}$$

For a given level of capital requirements y and an asset type x, total welfare writes as:

$$pr(x) + (1-p)[1 - y - \lambda \min(\ell(x) - y, 0)] - \delta y. \tag{OA.2}$$

We want to derive an objective function for the regulator. As $pr(x) + (1-p)(1-y)$ is exogenously given, we can consider the following objective function:

$$\mathcal{W}(x, y) = -\lambda(1-p) \min(\ell(x) - y, 0) - \delta y. \tag{OA.3}$$

As explained in Section 4.1, because of the possibility of mistakes, for a given regulation φ and state of the world x, the actual capital requirements will not necessarily be $\varphi(x)$ but follow some distribution $\hat{\mu}$. To keep the model simple, we assume that with a probability denoted $\hat{\mu}(\varphi)$ the regulation is correctly implemented and $y = \varphi(x)$, while with the converse probability y is uniformly distributed between 0 and 1. If there is no mistake, the expected welfare for the regulator is:

$$\mathcal{W}^*(\varphi) = \sum_{x=1}^{N} \nu(x) \left[-\lambda(1-p) \min(\ell(x) - \varphi(x), 0) - \delta\varphi(x)\right]. \tag{OA.4}$$

Instead, conditionally on a mistake, the expected welfare is:

$$\mathcal{W}_0(\varphi) = \sum_{x=1}^{N} \nu(x) \int_0^1 \left[-\lambda(1-p) \min(\ell(x) - y, 0) - \delta y\right] dy. \tag{OA.5}$$

Taking into account the costs of regulation, we can write the regulator's problem as:

$$\max_{\varphi} \hat{\mu}(\varphi)\mathcal{W}^*(\varphi) + (1 - \hat{\mu}(\varphi))\mathcal{W}_0(\varphi) - w \sum_{x=1}^{N} \nu(x) e^*(\varphi, x). \tag{OA.6}$$

Without regulatory complexity, $\hat{\mu}(\varphi)$ would be equal to 1 and $e^*(\varphi, x)$ to zero. Then it would clearly be optimal to set $\varphi(x) = \ell(x)$. This would require a regulation specifying N different "risk buckets". Using the regulatory text F given in Section 4.1, we would have $I = N$ different intervals.

Solving the problem with regulatory complexity is more intricate. To get analytical results instead of resorting to a numerical solution, we simplify the economic environment by making it continuous: x is now uniformly distributed over $[0, 1]$, and moreover $\ell(x) = x$. We can then easily compute the expected welfare $\mathcal{W}^*_{a,b}(y)$ for x in a given interval $[a, b]$ and a given level of capital requirements y, conditionally on the regulation being correctly implemented:

$$\begin{aligned}
\mathcal{W}^*_{a,b}(y) &= \int_a^b [-\lambda(1-p)\min(x-y,0) - \delta y]dx && \text{(OA.7)}\\
&= -\lambda(1-p)\int_y^b (x-y)dy - \delta y(b-a) && \text{(OA.8)}\\
&= -\lambda(1-p)\frac{(b-y)^2}{2} - \delta y(b-a). && \text{(OA.9)}
\end{aligned}$$

Maximizing this quantity in y, we obtain the optimal capital requirement $y^*_{a,b}$ for interval $[a, b]$:

$$y^*_{a,b} = b - \delta \frac{b-a}{\lambda(1-p)}. \tag{OA.10}$$

Note that we have $a \le y^*_{a,b} \le b$. This means that banks with assets x close to a will be overcapitalized (they have more capital than what is necessary to sustain the losses $\ell(x) = x$), while banks with assets x close to b will be undercapitalized (they default with probability $1-p$).

We obtain that the optimal welfare over interval $[a,b]$ is given by:

$$\mathcal{W}^*_{a,b}(y^*_{a,b}) = \delta(b-a)\left[\frac{\delta(b-a)}{2\lambda(1-p)} - b\right]. \tag{OA.11}$$

Using this expression, we can determine the optimal intervals chosen by the regulator. If the regulator uses I intervals it is actually optimal to split $[0,1]$ into I intervals of equal length. To see why, consider the case of two intervals, $[0,\bar{x}]$ and $[\bar{x},1]$. Total expected welfare is then given by:

$$\begin{aligned}\mathcal{W}^*_{0,\bar{x}}(y^*_{0,\bar{x}}) + W_{\bar{x},1}(y^*_{\bar{x},1}) &= \delta\bar{x}\left[\frac{\delta\bar{x}}{2\lambda(1-p)} - \bar{x}\right] + \delta(1-\bar{x})\left[\frac{\delta(1-\bar{x})}{2\lambda(1-p)} - 1\right] && \text{(OA.12)}\\ &= \delta\bar{x}(1-\bar{x})\frac{\lambda(1-p)-\delta}{\lambda(1-p)} - \frac{\delta}{2\lambda(1-p)}[\delta - 2\lambda(1-p)] && \text{(OA.13)}\end{aligned}$$

We immediately see that the optimal $\bar{x}$ is equal to $1/2$, that is, the two intervals are symmetric.

Consider now any number I of intervals. Following the same approach it is easily proved that all intervals must have the same length, so that the I intervals are $[0,1/I]$, $[1/I,2/I]$...$[(I-$

1)/I, 1]. The $i+1$-th interval has a welfare of:

$$\mathcal{W}^*_{i/I,(i+1)/I}(y^*_{i/I,(i+1)/I}) = \frac{\delta}{I}\left[\frac{\delta}{2I\lambda(1-p)} - \frac{i+1}{I}\right] \quad \text{(OA.14)}$$

$$= \frac{\delta}{I^2}\left[\frac{\delta - 2\lambda(1-p)}{2\lambda(1-p)} - i\right]. \quad \text{(OA.15)}$$

We use this lest expression to compute total welfare conditionally on the regulation being well implemented, aggregating over all the intervals:

$$\mathcal{W}^*(I) = \sum_{i=0}^{I-1} W_{i/I,(i+1)/I}(y^*_{i/I,(i+1)/I}) = -\frac{\delta}{2} - \frac{\delta}{2I\lambda(1-p)}[\lambda(1-p) - \delta]. \quad \text{(OA.16)}$$

Total welfare is thus increasing in I, and converges to $-\delta/2$ as $I \to +\infty$.

We then compute $\mathcal{W}_0$, which is expected welfare conditionally on regulation not being well implemented, in which case the capital requirements y are just random. We have:

$$\mathcal{W}_0 = \int_0^1 \int_0^1 [-\lambda(1-p)\min(x-y,0) - \delta y] dy dx = -\frac{\lambda(1-p)}{6} - \frac{\delta}{2}. \quad \text{(OA.17)}$$

For illustration, we assume that the probability $\hat{\mu}$ of implementing the regulation correctly and the total effort e^* only depend on the *quantity* and *potential* of F, according to the estimates from specification (3) in Table OA.14 and (4) in Table OA.17. This gives us, for every I:[40]

[40] In each case the constant term is the sum of the constant in the regression, the average participant fixed effect, and the average question fixed effect.

$$\hat{\mu}(I) = 1 - \Phi(-2.877 + 0.507 quantity(I)) \quad \text{(OA.18)}$$

$$e^*(I) = 4.055 + 7.965 potential(I). \quad \text{(OA.19)}$$

In the regulatory text we assumed, the logical operators are "if", "else", and "then", $\equiv$ is a regulatory operator, and $<$ is a mathematical operator. The operands are x, y, the $\bar{x}_i$, and the E_i^*. We have $\eta_R = 1$ and $N_R = I$, $\eta_L = 3$ and $N_L = 3(I-1)$, $\eta_M = 1$ and $N_M = I - 1$, $\eta_{OD} = 2I + 1$ and $N_{OD} = 4I - 2$. Given the number I of intervals used, we can then easily compute the measures using the formulas in Table 1 and see how they vary with the number of asset classes I. On Fig. OA.8, we plot $\hat{\mu}(I)$, $e^*(I)$, $W^*(I)$ from equation (OA.16), and the regulator's objective (OA.6) as a function of I.[41] In this example, we obtain that expected welfare as a function of I is bell-shaped, and for more than $I = 3$ risk-buckets the costs of complexity outweigh the benefits. Hence, the policymaker can compute that the optimal trade-off between "risk-sensitivity" and "simplicity" is achieved for $I = 3$ intervals.

[41] We use $\lambda = 0.05$, $\delta = 0.01$, $p = 0.05$, and $w = 10^{-6}$. These parameters are meant for illustration only.

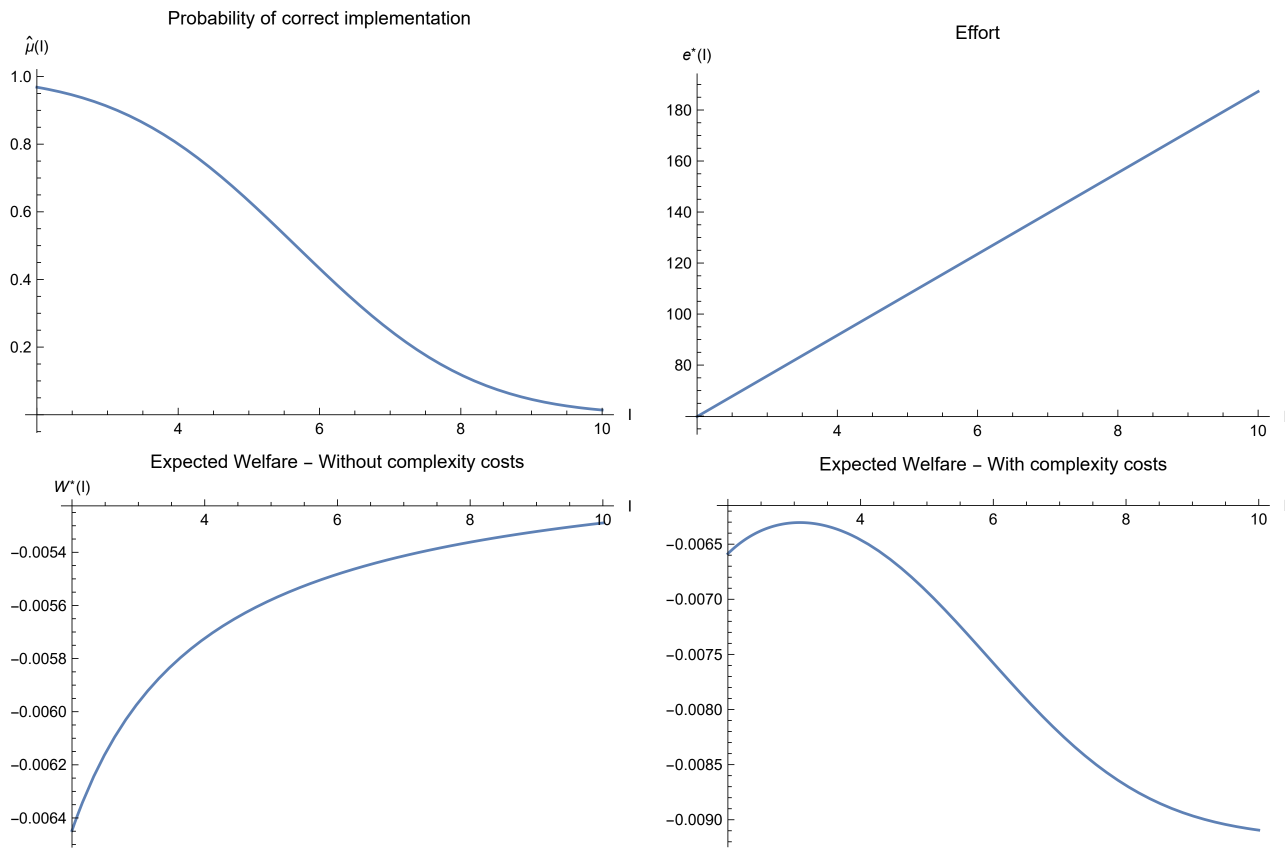


Figure OA.8: Costs of complexity and welfare as a function of the number of risk buckets

This figure plots the probability of correct implementation $\hat{\mu}(I)$ and the effort $e^*(I)$ as a function of the number fo intervals or "risk-buckets" I, as well as the expected welfare without complexity costs $\mathcal{W}^*(I)$, and the welfare with complexity costs given by (OA.6).