Measuring Regulatory Complexity *

Jean-Edouard Colliard† — HEC Paris

Co-Pierre Georg‡ — University of Cape Town

January 1, 2024

Abstract

Despite a heated debate on the complexity of financial regulation, a comprehensive framework to study regulatory complexity is lacking. We propose one inspired by the analysis of algorithmic complexity in computer science. We use this framework to distinguish different dimensions of complexity, classify existing measures, develop new ones, compute them on two examples—Basel I and the Dodd-Frank Act—and validate them using novel experiments. Our framework offers a quantitative approach to the policy trade-off between regulatory complexity and precision. Our toolkit is freely available and allows researchers to measure the complexity of any normative text and test alternative measures.

Keywords: Financial Regulation, Capital Regulation, Regulatory Complexity, Basel Accords.

JEL classification: G18, G28, G41.

---

*We are grateful to Tobias Berg, Bruno Biais, Claire Célérier, Stephen Karolyi, Igor Kozhanov, Iman van Lelyveld, Christoph Merkle, Sébastien Pouget, as well as to participants to the 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, OCC, University of Bonn, University of Exeter, and University of Luxembourg, for helpful comments and suggestions. We are grateful to 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/Risk Foundation "Regulation and Systemic Risk", the Natixis Chair "Business Analytics for Future Banking", and the Europlace Institute of Finance.

†E-Mail: colliard@hec.fr; Corresponding author.

‡E-Mail: co-pierre.georg@uct.ac.za

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, articulate the view that bank capital regulation has become so complex as to be counter-productive and likely to favor regulatory arbitrage. The Basel Committee on Banking Supervision is aware of the issue, and considers that there is a trade-off between the simplicity and the precision of regulation (Basel Committee on Banking Supervision (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]

While there is a widespread concern that regulation has become too complex, "regulatory complexity" remains an elusive concept. Debates about the complexity of different rules and contracts have come up in other contexts, such as structured 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 show that with these ingredients one can approach the trade-off studied by the Basel Committee on Banking Supervision (2013) in a quantitative manner.[2]

We hypothesize that a regulation can be seen as an algorithm: it is a sequence of instructions that are applied to an economic agent and return a regulatory action. Previous

[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://github.com/cogeorg/RegulatoryComplexity_Public`. The replications files for this paper are available here: `https://ufile.io/7al4cyzg`.

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 distinguish between different dimensions of complexity, derive six measures of regulatory complexity in a unified model of regulation, test the validity of these measures experimentally, compute them on a large scale regulatory text (the Dodd-Frank Act), and include them in a normative model of the trade-off between precision and complexity.

We first use our framework to formally define measures of regulatory complexity, and distinguish between the different dimensions of complexity that can be captured. In particular, we make a distinction between: (i) "problem complexity"—a regulation is complex because it aims at imposing many different rules on the regulated entities, independently of the language used; (ii) "psychological complexity"—a regulation is complex because it is difficult for a human reader to understand; and (iii) "computational complexity"—a regulation is complex because it is long to implement. We relate these different dimensions to empirically observable quantities, namely the occurrence of mistakes in the regulatory process, and the effort necessary to understand and apply the regulation. These quantities can be seen as "ex post" measures of regulatory complexity.

In many applications such ex post measures are not available, and as a proxy it is necessary to use "ex ante" measures based on the content of regulatory documents. Most measures that have been proposed rely on linguistics and only cover psychological complexity. To derive measures of problem complexity as well, we use the approach developed by Halstead (1977) for measuring 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 and the number of operands used in those operations. In the context of regulation, these measures can capture the number of different rules ("operations") 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 three measures of regulatory complexity that have already been proposed in the literature, and go on to define three new ones.

As a proof of concept, we show how to measure the complexity of a regulation in practice by considering the design of risk weights in the Basel I Accords. This regulation is a suitable testing ground because it is close to being an actual algorithm. We compare two different methods: (i) We write a computer code corresponding to the instructions of Basel I and measure the algorithmic complexity of this code directly; and (ii) We analyze the text of the regulation and classify words according to whether they correspond to what would be an operand or an operator in an algorithm, and compute the same measures, this time trying to adapt them from the realm of computer science to an actual text. In particular, we observe that the measures of "problem complexity", which in principle should not depend on the language used, are indeed very close in the text and the algorithm versions.

An important gap in the existing literature on regulatory complexity is the validation of complexity measures: how does one show that a proposed measure indeed captures some dimension of complexity? Here again the parallel with algorithms suggests an answer. The literature in computer science tests the validity of different measures of algorithmic complexity by testing their ability to forecast mistakes made by programmers or the time they need to code the program (see, e.g., Canfora *et al.* (2005)). We apply the same idea to the context of regulation. We give participants to an experiment different regulatory instructions consisting in (randomly generated) Basel-I type rules, and the balance sheet of a 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. In both cases we also test whether a given measure has explanatory power beyond the mere length of the regulation. We find that only two of the five measures we propose (other than length) pass this test, suggesting that our experimental design is a powerful touchstone to test the validity of new measures. Interestingly, these are the two measures meant to capture problem complexity, which validates the idea that this is indeed a dimension not captured by length alone. All the material is online and can be directly used to validate any measure of regulatory complexity based on the text of a regulation, not only ours, thus opening the path to comparing the performance of different measures within a unified framework.

To show that our approach can be adopted at scale, we apply our text analysis approach to the 2010 Dodd-Frank Act. Because the Dodd-Frank Act covers many different aspects of financial regulation, by doing so we created a large dictionary of 5,872 operands and 429 operators. We expect that a large fraction of words found in other texts will already be in our dictionary. To show this, we look at the fraction of words in each of the 16 titles of the Dodd-Frank Act that would have already been included in a dictionary obtained using only the other 15 titles. We find that, on average across all titles, 88% of operands and 96% of operators would have already been in this counterfactual dictionary. We make this dictionary available online, so that interested researchers can compute our measures on other regulatory texts and collaboratively enrich the dictionary. In addition, this dictionary can serve as a training sample for identifying operands and operators in longer texts using machine learning.

Finally, we show how building on our approach could eventually lead to a quantitative model of the trade-off between the precision and the complexity of regulation mentioned in Basel Committee on Banking Supervision (2013). To explore this possibility, we build a simple model of a bank capital regulation relying on risk buckets, as in Basel I. We can use our measures and the experimental estimates to compute the complexity cost of additional buckets, and hence study the optimal trade-off between these costs and the benefits of additional precision. More generally, this example shows that our measure can be used in normative models of regulation. For instance, in the context of a model this allows us to compare a complex regulation achieving the first-best to a simpler one that still achieves a high level of welfare.

We review the literature on measures of regulatory complexity in the next section, where we show how different measures fit into our framework, or explain why they do not.[3] As mentioned above, a growing number of papers have studied the complexity of various financial products and contracts more generally. We provide a unifying framework for these different applications, to the extent that they consider rules describing how to perform a certain operation.[4]

A growing number of recent theory papers have implications for the complexity of regulation. Hakenes and Schnabel (2012) develop a model of "capture by sophistication" in which some regulators cannot understand complex arguments and "rubber-stamp" some

[3]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.

[4]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)).

claims made by the industry so as not to reveal their lack of sophistication. 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 Asriyan *et al.* (2021), a policymaker proposes a regulation that then needs to be accepted, e.g., by 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 of the proposed regulation. Foarta and Morelli (2022) also model the dynamics of legal complexity over time, and make predictions regarding these dynamics. We hope that by proposing new measures of regulatory complexity our paper will make it possible to test these theories, which to our knowledge has not been done yet.[5]

There is a broader theoretical literature on complexity in product markets, developing the idea that complexity can be used by firms to "obfuscate" and gain market power (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, although there is a similarity with the idea of "capture by sophistication". 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, an interesting example being the sale of derivatives. Carlin *et al.* (2013) find support for this idea in a trading experiment, with adverse selection being larger for more

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

complex assets.

Further from finance applications, the experimental approach we use in Section 3 is related to a 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, and study how their performance correlates with measures of the complexity of the problem and measures of the complexity of different algorithms used to solve it. Our 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 in behavioral economics dating back to Rubinstein (1986) models the strategies and decision procedures of economic agents as automata, and associates measures of the complexity of these automata (in particular, the number of states involved) to the cognitive costs that following these strategies imposes on agents. Oprea (2020) uses an experimental approach to measure the cognitive costs of following different procedures ("implementation complexity"), and shows that these costs correlate well with complexity measures of the associated automata.[6] Our approach differs in that we do not represent regulation as an automaton. This is in principle possible but extremely costly to do on a large scale text, so that we believe the Halstead representation of an algorithm is a more promising approach for the study of regulatory complexity.

[6]See also Kendall and Oprea (2021) who study experimentally the computational complexity of inferring the process that generated a particular data sequence.

1 A unifying framework

Because the term "complexity" is used somewhat vaguely in the social sciences, different authors, policymakers, and industry participants have different concepts in mind when referring to "regulatory complexity". To clarify this issue, we first propose a model of the regulatory process that emphasizes the parallel between regulation and code. We then use this model to define the different dimensions of complexity and discuss how to measure them empirically. Finally, we propose measures based on an extension of Halstead (1977) and review how the different measures proposed in the literature fit in our framework.

1.1 A model of the regulatory process

We start by formalizing the analogy between regulations and algorithms which then allows us to draw from the computer science literature studying algorithmic and software complexity.[7] Knuth (1973) describes an algorithm as: *"a finite set of rules that gives a sequence of operations for solving a specific type of problem"* and identifies five features an algorithm must satisfy. First, an algorithm must terminate after a finite number of steps. Second, each step of the algorithm must be precisely defined—be it verbally or through formal use of mathematics or a programming language. Third, an algorithm has zero or more inputs, taken from a well specified set of objects. Fourth, it has one or more outputs—quantities that have a specified relationship to the inputs. Lastly, an algorithm should use sufficiently simple operations so that it can be computed, in principle, *"by someone using pencil and*

[7] Software complexity is typically defined in reference to psychological complexity and Zuse (1990), for example, defines it as the *"psychological complexity of programs"*. This is echoed by the IEEE (1990), who define software complexity as *"the degree to which a system or component has a design or implementation that is difficult to understand and verify"*.

paper."

Surprisingly, a formal definition of an algorithm beyond the informal characterization provided above is not without difficulty, but for the purpose of our paper, the informal description of an algorithm is sufficient. In the case of regulation, the "inputs" are the characteristics of regulated entities (e.g., an individual financial institution, a market, or the entire financial system), and the "output" a certain regulatory action (e.g., imposing a fine on a bank, imposing higher capital requirements, or simply allowing the bank to continue operating). The regulation is a list of rules describing how to map a regulated entity with certain characteristics to a certain regulatory action.

It is important to make a distinction between an algorithm and the problem it tries to solve. As an example, consider the problem of sorting a deck of cards. There are different algorithms, or lists of operations, to solve this problem, e.g., the `InsertionSort` or `QuickSort` algorithms. They all solve the same problem, take as input the deck of cards that is to be sorted, and return as output the sorted deck of cards, but they can differ in their complexity. In the case of regulation, the problem is the mapping between regulated entities and regulatory actions, and the regulation is a description of this mapping:

Definition 1. *A regulatory problem is a mapping $\varphi : x \mapsto y$ from a set of regulated entities $\mathcal{X}$ to a set of regulatory actions $\mathcal{A}$.*

In a world of unlimited cognitive ability of regulators and supervisors, the mapping φ would be sufficient to describe the regulatory process. The traditional literature on banking regulation implicitly makes this assumption, and derives φ as the optimal solution to a contract theory or mechanism design problem. Similarly, in the realm of coding, one could posit a mathematical problem, show it has a solution, and simply assume that some code exists to compute this solution. We depart from this assumption by considering that writing down the actual rules or "code" and then implementing them is not trivial, and is actually costly and error-prone. In the following, we write down a simple model of drafting and implementing regulations (or code), that allows for errors at different steps.

The first step of the regulatory process is the *drafting* of a regulation. We define a "regulation" as a list of written rules that aim at solving the regulatory problem φ:

Definition 2. *A regulation F is a finite sequence of elements taken in a vocabulary $\mathcal{V}$. This sequence of elements is interpreted through a language $\mathcal{L}$.*

We assume that a *regulator* needs to draft the regulation F. This regulator is endowed with a certain skill θ_D at drafting regulations, and may exert a certain quantity of costly effort e_D when drafting. The regulation is then a function of the regulatory problem the regulator tries to solve, her skill, and her effort:

$$F = D(\varphi, \theta_D, e_D, \mathcal{L}, \mathcal{V}). \tag{1}$$

The second step of the regulatory process is *interpretation*: one needs to read and interpret the regulation F in order to apply it to a particular entity. Note that F is not a mapping but simply a list of elements of $\mathcal{V}$, i.e., words. If this list is perfectly interpreted according to the rules of language $\mathcal{L}$, it describes a mapping from $\mathcal{X}$ to $\mathcal{A}$. We denote f this mapping corresponding to a perfect interpretation of F.

In the context of coding, the code is interpreted by a computer and this interpretation is usually correct. In the context of regulation, the interpretation is done by a human *reader* who may misunderstand the regulation. Such a reader may have a certain skill θ_I at interpreting the text, and exerts a costly effort e_I. The reader's interpretation of the regulatory text is a new mapping from $\mathcal{X}$ to $\mathcal{A}$ denoted $\tilde{f}$:

$$\tilde{f} = I(F, \theta_I, e_I, \mathcal{L}, \mathcal{V}). \tag{2}$$

The last step of the regulatory process is to apply the mapping $\tilde{f}$ to a given entity x and determine the regulatory action to take, a step we call *supervision*. In the context of coding, a computer would always, mechanically, associate the output $f(x)$ to the input x. In the context of regulation, however, a human supervisor whose understanding of regulation is represented by $\tilde{f}$ may not necessarily associate $\tilde{f}(x)$ to an entity x. Instead, we denote $\hat{f}(x)$ the action taken by the supervisor. We allow this action to depend on the supervisor's skill θ_S and effort e_S.

$$\hat{f}(x) = S(x, \tilde{f}, \theta_S, e_S). \tag{3}$$

At the end of this regulatory process, we reach a regulatory action $\hat{f}(x)$. In the absence of mistakes at the drafting, interpreting, and supervision stages, this action is equal to $\varphi(x)$. However, at each stage there is room for mistakes, with the consequence that ultimately the wrong regulatory action may be taken, $\hat{f}(x) \neq \varphi(x)$. Figure 1 summarizes our model of the regulatory process and which mistakes can occur during this process.

[Insert Fig. 1 here.]

1.2 Dimensions of complexity

We now define different dimensions of complexity based on how costly and prone to mistakes are the different stages of the regulatory process described in Section 1.1. An advantage of this approach is that the different dimensions of complexity we define are measurable empirically, based on the implementation of a certain regulation, and we call these measures *ex-post* measures of regulatory complexity. The experiments described in Section 3 follow this approach. In contrast, Section 1.3 defines *ex-ante* measures of regulatory complexity, that are only based on the observation of a regulatory text and on ex ante principles.[8]

We give both "mistake-based" definitions of complexity as the propensity of making a mistake in the regulatory process, and "cost-based" definitions based on the cost (in terms of labor or mental effort) associated with each step of the regulatory process. Formally, we always consider a regulation F associated with a correct interpretation f. We assume a given entity x, and given types $\theta_D, \theta_I, \theta_S$ and effort levels e_D, e_I, and e_S.[9]

We first define *problem complexity*:

Definition 3. *The mistake-based problem complexity of regulation F is given by* $\Pr[D(f, \theta_D, e_D(f, \theta_D), \mathcal{L}, \mathcal{V}) \neq f]$. *The cost-based problem complexity of regulation F is given by $e_D(f, \theta_D)$.*

In words, a regulation has a higher mistake-based problem complexity when the regulator drafting such a regulation is more likely to make a mistake, so that the correct interpretation of the draft is actually different from what was intended. If we apply this definition to the

[8] The distinction of ex-ante and ex-post measures was prominent in the early literature on software complexity. Weyuker (1988), for example, develops a framework of axioms within which to evaluate software complexity measures without reference to specific code. See also Fenton (1994). For ex-post measures based on experiments with coders see, among others, Zhang and Baddoo (2007) who study the performance of different widely-used measures of software complexity, including the McCabe (1976) measures we also use.

[9] In applications it is straightforward to extend the formalism of this section to a population of regulated entities, regulators, readers, and supervisors, with types following a distribution.

context of coding, we would say that, for example, factoring a number has a higher problem complexity than sorting a vector, if we observed that a coder asked to program both tasks is more likely to succeed at the latter than at the former. A regulation has a higher cost-based problem complexity if we observe that the regulator spends more effort at drafting the regulation.

Similarly, we define *psychological complexity*:

Definition 4. *The mistake-based psychological complexity of regulation F is given by* $\Pr[I(F, \theta_I, e_I(F, \theta_I), \mathcal{L}, \mathcal{V}) \neq f]$*. The cost-based psychological coomplexity of regulation F is given by* $e_I(F, \theta_I)$*.*

In words, a regulatory text has a higher psychological complexity if it is more likely to be misinterpreted by a reader, or if the reader has to exert more effort interpreting it. Note that if two regulations F and F' have the same correct interpretation ($f = f'$) then they have the same problem complexity and hence any difference in psychological complexity only comes from the way the regulatory text is drafted. If instead one compares texts corresponding to different regulatory problems, then differences in problem complexity also generate differences in psychological complexity.

Finally, we define *computational complexity*:

Definition 5. *Denoting* $\tilde{f} = I(F, \theta_I, e_I(F, \theta_I), \mathcal{L}, \mathcal{V})$*, the mistake-based computational complexity of regulation F is given by* $\Pr[S(x, \tilde{f}, \theta_S, e_S(x, \tilde{f}, \theta_S)) \neq f(x)]$*. The cost-based computational complexity of regulation F is given by* $e_S(x, \tilde{f}, \theta_S)$*.*

In words, a regulatory text has a higher computational complexity if the action taken by a supervisor on a regulated entity is different from the one actually dictated by the text.

Note that texts with different levels of psychological complexity will also have different levels of computational complexity as a consequence: one reason why a supervisor may reach the wrong regulatory action is that he misunderstands the regulation in the first place.

We believe that these three dimensions of complexity capture some of the main ideas that people have in mind when talking about the complexity of regulation. Regulatory complexity may mean that the regulatory problem is complex, e.g., it deals with many different aspects of a bank's business, or foresees a large number of regulatory actions. We call this the *problem complexity* of regulation. Problem complexity depends on φ, but is independent of which regulatory text F solves it. Regulatory complexity may also mean that an actual text is complex, which may be due both to the complexity of the underlying problem φ and to the complexity of the particular text F. Following the computer science literature (e.g., Zuse (1990)), we call this dimension the *psychological complexity* of regulation, as it reflects the difficulty of understanding a particular solution to a problem. Finally, regulatory complexity may mean that applying a regulation to a particular entity is difficult. Imagine for instance a regulation that exempts small banks from most rules. It could then 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. Thus, this dimension depends on the entity to which the regulation is applied. Following again the computer science literature, we call this dimension the *computational complexity* of regulation.

Finally, our formalization makes it clear that any notion of regulatory complexity is necessarily relative to the regulated entities considered and the humans processing the regulation. Depending on how a regulatory text is written, it might for instance seem more complex to lawyers than to economists, or vice versa.

A few papers have proposed ex-post measures based on how much effort regulated entities spend on complying with regulations (computational complexity). For instance, Simkovic and Zhang (2020) propose a *Regulation Index* based on the proportion of regulation-related employees in different sectors, as measured in the Occupational Employment Statistics data from the U.S. Bureau of Labor Statistics. Kalmenovitz (2023) proposes four *RegIn* indexes of regulatory intensity, based on the number of forms required by Federal regulatory agencies in the U.S., 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 U.S. firms by *NetReg*, a measure based on the mention of regulatory topics in transcripts of earnings calls. Singla (2022) uses estimates of regulatory costs provided by U.S. regulatory agencies themselves at the level of each industry.

1.3 Ex-ante measures of complexity

Ex-post measures of regulatory complexity are appealing because they are empirically grounded. However, in policy applications in particular it is preferable to measure the complexity of a proposed text before it is actually implemented. It is then necessary to complement these ex-post measures with "ex-ante" measures, that can be directly computed based on any regulatory text F. In this section we derive a number of ex-ante measures. We will study in Section 3 to what extent these measures correlate with ex-post measures of regulatory complexity and can hence serve to proxy them.

We now show how several ex-ante measures of regulatory complexity can be derived by modeling a regulation like an algorithm in Halstead (1977). We consider regulation

F as a sequence of "n-grams" (expressions of length n that are elements in a language) $F = \{w_1, w_2...w_N\}$, from which we extract two sequences: a sequence of N_{OR} operators and a sequence of N_{OD} operands, with $N_{OR} + N_{OD} = N$. 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.

Using Halstead's definition, operands in an algorithm 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 $x = (x_1, x_2...x_n)$ which can be written as:

```
y = sqrt(x_1^2+x_2^2...+x_n^2)
```
(4)

Here, the operators are $=, sqrt, +, \hat{}$, and the operands $y, x_i, 2$. So 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 regulations and generic algorithms, we propose a slightly finer partition than Halstead's. Already in Halstead's work, the assignment operator (the $=$ sign in (4)) 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". Operators that are not regulatory operators fall into two categories: "logical operators" represent logical operations such as "if", "when", etc., while "mathematical operators" 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 now derive six measures of complexity within our extended framework.

First, the simplest measure of regulatory complexity is the total number of words N in a regulation, which we denote *length*. This measure is used for instance in Haldane and Madouros (2012).

Second, a popular measure in computer science is cyclomatic complexity (McCabe, 1976), which is the number of different paths an algorithm can follow. We denote it *cyclomatic*. This is measured in practice by the number N_L of different logical operators, as in, e.g., Li *et al.* (2015).

Third, the quantity of regulations, denoted *quantity*, can be measured by counting the total number of regulatory operators, N_R. This corresponds to the RegData measure of Al-Ubaydli and McLaughlin (2017), who count the number of words indicating a binding constraint in the U.S. Code of Federal Regulations.[10] A related example is Herring (2018), who measures complexity through the number of different capital ratios Global Systemically Important Banks need to comply with.

Fourth and fifth, Halstead (1977) suggests two additional measures, new to the literature on regulatory complexity. The three measures above depend on the actual text F and hence cannot capture problem complexity, which is independent of the text chosen to solve the underlying problem φ. How can one obtain a measure of problem complexity, that depends only on φ? Halstead's answer to this question is to look at the shortest possible program that can solve the problem, in the best possible programming language. Defining this algorithm

[10] See also McLaughlin *et al.* (2021) for a recent study using this measure.

is easy. For example, the shortest possible program to compute the vector norm is:

```
y = vecnorm(x_1,x_2...x_n)
```
(5)

where `vecnorm` is a function returning the vector norm. 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 (which in our example already exists in the programming language).

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 $\eta^*_{OD} = \eta_{OD}$. The volume of this minimal program, equal to $2 + \eta_{OD}$, is a measure of problem complexity called *potential volume* and denoted *potential*.

Finally, an interesting question to ask is whether an algorithm is close to the shortest possible algorithm. Adapting Halstead (1977), we define the *level* of an algorithm as $level = potential/length$. The measure *level* 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—a technical jargon opaque to outsiders. Conversely, a low value of *level* means that the regulation starts from elementary concepts and operations.

Sixth and last, by symmetry with *potential* we propose to also consider the number of unique operators η_{OR}, or *operator diversity*, denoted *diversity*, 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 or relies on different types of logical tests could be more difficult to understand.

For completeness, we briefly review other ex-ante measures that have been proposed in the literature but do not directly fit within our framework.

Kalmenovitz *et al.* (2022) propose a measure of regulatory fragmentation, *RegFragmentation*, which relies on counting the number of different regulatory agencies mentioned in the Federal Register and relevant for the same industry. This measure is best interpreted as a measure of computational complexity, the idea being that the overlap between different authorities makes compliance more costly. It can be included in our framework by counting separately the operands corresponding to regulatory authorities.

Amadxarif *et al.* (2019) use a number of measures from the linguistics literature, in particular *average word length*, the Maas' index of *lexical diversity* (Maas, 1972), and the Flesch-Kincaid grade level *readability metric* (Kincaid *et al.*, 1975). Katz and Bommarito (2014) and Li *et al.* (2015) also use *Shannon's entropy* as an alternative measure of lexical diversity. All these measures do not rely on a partition of words between operands and operators, and apply equally well to texts that have no normative or operational content. These measures aim at capturing the complexity of the style used by an author, which can be part of psychological complexity, rather than the complexity of the underlying ideas.

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* (how often a legal text is cited by other legal texts), *out-degree* (how often a legal text cites other legal texts), or different *network centralities* can then be interpreted as measures of psychological complexity.[11] 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.

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. This is different from other classifications we are aware of (e.g., in Amadxarif *et al.* (2019)), which are based on how the different measures are computed. In particular, our classification illustrates the special role of potential volume and quantity, the only measures of problem complexity.

[Insert Table 1 here.]

2 Basel I

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 use as an example the 1988 Basel I Accords (Basel Committee on Banking Supervision (1988)). We focus on Annex 2, "Risk weights by category of on-balance-sheet asset". As we will illustrate below, this is a natural starting point because this part of the regulation can easily be described as an algorithm. This allows us to compute our measures based both on an algorithmic representation of Basel I, and on the actual text. We then compare the results obtained in both cases and conclude that our measures can be applied directly on the text.

[11] Amadxarif *et al.* (2019), for example, discuss the use of PageRank centrality, which measures how often a node in a network is cited by nodes that themselves are cited often.

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. It maps different asset classes to different risk buckets, and different capital instruments to different 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%. As this succinct description makes clear, Basel I is easily described as an algorithm. We write a "pseudo-code" that implements the computation of risk-weighted assets described in the Annex 2 of the text, i.e., our code maps a bank balance sheet to total risk-weighted assets under Basel I. We give this program in Online Appendix OA.1. In this section, we briefly explain the structure of the program and give the associated complexity measures.

Annex 2 of the Basel I text is a list of balance sheet items associated with 5 different risk weights. For instance, in the 20% risk-weight category we have "Claims on banks incorporated in the OECD and loans guaranteed by OECD incorporated banks". In our code this is translated into:

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

We can easily identify the operands and operators in such a piece of code, and compute our measures of complexity. For instance here the operands are the different asset classes (e.g., `ASSET_CLASS`), characteristics (e.g., `ISSUER_COUNTRY`), values of these characteristics (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 conduct the same exercise for each of the 19 items covered by Basel I, and report the full results in Online Appendix OA.2.

2.2 Text analysis and comparison

We now repeat the same analysis of the Appendix 2 of Basel I, but relying this time on the actual text and not on our "translation" into code. A drawback of the text of Basel I's Appendix 2 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 to the original text.[12] We then classify as "operands" the words or word combinations that have the same function as operands in the program, more precisely economic entities (e.g., "bank" or "OECD"), concepts (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 are "shall" and "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 81 unique words out of the 86-word vocabulary used by 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 don't 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 compute the correlations between the text-based measures and their algorithm-

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

based counterparts. Table 2 gives the correlation coefficients.[13] The correlation coefficients we obtain are quite high, perhaps with the exception of *diversity*, which shows that the text-based analysis and the algorithm-based analysis are capturing similar patterns. Differences arise between the algorithm and the text because the text version is sometimes ambiguous or leaves some elements implicit. A good example is item (2a), which has $length = 43$ in the algorithmic version but $length = 22$ only in the text version. However, in both versions this item stands out as one of the most complex according to *cyclomatic*, *diversity*, and *potential*. More generally, the correlation is particularly high for measures of problem complexity (*quantity* and *potential*), which indeed should theoretically not depend on whether the regulation is expressed in English or in code.

Overall, we conclude from this comparison that measures of regulatory complexity relying on a text analysis can be a good proxy for the more theoretically founded measures based on the algorithmic version. This supports our adoption of the text-based approach for a full-scale regulatory text in Section 4.1. In addition, this analysis confirms that *quantity* and *potential* are indeed capturing problem complexity, as they are less affected by a change in the language used.

[Insert Table 2 here.]

3 Experiments

The purpose of ex-ante measures of complexity such as those studied in Sections 1.3 and 2 is that they can be computed at scale and used for empirical or policy applications. How-

[13]Formally the coefficients are not defined for *quantity*. 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.

ever, these measures are necessarily somewhat arbitrary and one may wonder whether they are good measures of regulatory complexity. The parallel with computer science suggests a methodology to test the relevance of the different measures: in computer science, 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 follow this idea and ask participants to an experiment to evaluate a regulatory action by computing regulatory quantities based on different regulations. For each regulation, we can compute the performance of the participants, which gives us ex-post measures of complexity. We then test whether different ex-ante measures of regulatory complexity can explain the the variation in ex-post measures.

3.1 Design

For our experiments we continue to rely on the Basel I regulation, this time as a testing ground. We generate a number of artificial "Basel-I like" instructions to compute risk-weighted assets based on a balance sheet, where the instructions vary in the number of asset classes to be considered, the different conditions attached to each asset class, and the number of different risk-weights, so that they will also have different ex-ante measures of regulatory complexity.

There is obviously a lot of flexibility and arbitrariness in writing artificial regulations. In order to tie our hands and avoid introducing potential biases by manually writing them, we generate a sample of randomized instructions for computing risk-weighted assets, all following the template of Basel I, but with random variations. 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. Figure 4 shows the possible attributes for each asset class and the values these attributes can take. Each randomly generated regulation consists of these building blocks that are connected using a random number of `AND` and `OR` statements which is no larger than the largest number of conditions in any `IF-THEN` clause in Basel I, which is six.

[**Insert Fig. 4 here.**]

As a last step, we manually check that the instructions make sense, e.g., they do not contain contradictory rules, and we make some minor manual changes to avoid ambiguities, grammar mistakes, etc.[14] At the end of this process, we obtain 38 regulations that we use in our experiment. As shown in Table 3 below, there is significant variation in all the complexity measures across the different regulations (in this section, all measures are computed based on the actual texts seen by the participants to the experiment). A limitation of our sample of randomly generated regulations is that several measures are quite correlated with each other, as seen in Table 4. Such a high correlation is to be expected: there is a natural correlation between the number of operands and operators, which we can also observe in the Basel I instructions (Table OA.5).

[14]The Appendix C shows an example of such a randomly generated regulation. In addition, the replication files of this paper give:[15] (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.

[Insert Tables 3 and 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. The students 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 offered (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.

Given the sanitary situation in early 2021, our experiment was conducted online. Each participant had to register on a website designed for conducting the experiment (https://regulatorycomplexity.org/).[16] After an introductory page, the participant registers and gives some background information. The participant is then shown a screen with expla-

[16]The interested reader can try the experiment anonymously by using the login "test_account" and password "test".

nations about the experiment and how to compute capital requirements. The next screen is a "test-round", which is the same for all participants (Figure 2). 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. If the answer is wrong, the participant is told so. In both cases, the participant is given an explanations 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. All the pages of the website are reproduced in Online Appendix OA.3.

[Insert Fig. 2 here.]

3.2 Results on mistake-based complexity

The answers form a balanced panel with 118 participants, indexed by i, answering a series of 9 questions each, indexed by t. The t-th question for each participant corresponds to regulation $R_{i,t} \in \{1...38\}$, which is randomly drawn from our 38 randomly generated regulation, with draws being independent across questions and students. We denote $correct_{i,t}$ a dummy variable equal to 1 if the participant i's answer to question t is correct. Following Definition 5, $correct_{i,t}$ is a measure of mistake-based complexity for participant i and regulation $R_{i,t}$. There is substantial variation across regulations. If one computes the proportion $correct_j$ of correct answers for each regulation j, the average of $correct_j$ is 68.61%, the standard deviation is 16.78%, the minimum 37.50%, first quartile 58.62%, median 64.68%, third quartile 82.14%, and two regulations have the maximum of 100.00%.

Denoting $\mu(R_{i,t})$ an ex ante complexity measure for regulation $R_{i,t}$, in order to show whether μ is a useful ex ante measure of complexity we study its power to explain the variation in *correct*. First, we evaluate the following probit model, at the participant-question level, using both participant and question fixed effects:

$$\Pr(correct_{i,t} = 1) = \Phi(\alpha + \beta\mu(R_{i,t}) + \gamma_i + \delta_t), \qquad (6)$$

where $\Phi(.)$ is the cdf of the standardized normal distribution. As our ex-ante measures are all based on the classification of words in the different regulations into operands and different types of operators, we first check that operands and operators have statistically significant different effects on $correct_{i,t}$. For each regulation j we compute $length_j$ the total length of the regulation, as well as $N_{OD,j}, N_{OR,j}, N_{R,j}, N_{L,j}, N_{M,j}$ the total numbers of operands, operators, regulatory operators, logical operators, and mathematical operators, respectively. We also compute the numbers $\eta_{OD,j}, \eta_{OR,j}, \eta_{R,j}, \eta_{L,j}, \eta_{M,j}$ of unique terms in each category. We then run the probit regression (6) on different subsets of these variables, as shown in

Table 5.

[Insert Table 5 here.]

Column (1) of Table 5 shows that *length* is negatively associated with correct answers, and together with participant and question fixed effects explains 24.3% of the variation of $correct_{i,t}$ across the 1,062 participant-question pairs (against 18.1% for a specification with fixed effects alone). In column (2) we split $length_j = N_{OD,j} + N_{OR,j}$ into its two components and run the probit separately on $N_{OD,j}$ and $N_{OR,j}$. Only the coefficient on $N_{OD,j}$ is statistically significant, suggesting that total operands and total operators play a different role. However, perhaps due to the small sample size, we can reject the hypothesis that the coefficients on $N_{OD,j}$ and $N_{OR,j}$ are equal to each other in this regression at the 10% level only (p-value of 9.6%). Column (3) shows that the total number of unique words is negatively associated with correct answers. Splitting again into unique operands and operators in Column (4), we obtain that unique operands are negatively associated with correct answers, whereas unique operators are positively associated. The difference between the two coefficients is statistically significant, and remains so after controlling for $length_j$ in Column (5) (p-value lower than 0.1% in both cases). Finally, in Column (6) we further decompose unique operators into unique mathematical operators $\eta_{M,j}$ and unique logical operators $\eta_{L,j}$. We do not include unique regulatory operators as $\eta_{R,j}$ is equal to 1 for every j. We find that the impact of unique mathematical operators is higher than that of unique logical operators. The coefficients on $\eta_{M,j}$ and $\eta_{L,j}$ are statistically significantly different from each other and from the coefficient on unique operands (all p-values are below 0.1%). We conclude from this series of regressions that separating the words used in the various regulations into operands and different categories of operators does indeed help explain the variation of mistake-based complexity across regulations.

We now turn to testing our ex-ante measures. First, we run the probit regression (6) on each of the six measures separately, across the 1,062 participant-question observations. In these regressions, we expect a reliable measure of mistake-based complexity to be negatively associated with correct answers, and the pseudo-R^2 of the regression measures how much of the variation of complexity across regulations is captured by the measure. The results are reported in Table 6. Second, we run each regression again, adding *length* as an independent variable. As *length* is a natural and standard measure of complexity, a new measure is useful only to the extent that it is able to explain the variation across regulations beyond what is already captured by *length* alone. Hence, we expect a good measure of mistake-based complexity to be significantly negatively associated with *correct* even after controlling for *length*, and to lead to a higher pseudo-R^2 than in the regression on *length* alone.

[Insert Table 6 and 7 here.]

We observe that *length*, *cyclomatic*, *quantity*, *potential*, and *diversity* have the expected negative correlation with *correct*, while *level* does not. In addition, once we control for length, we see that *quantity* stands out, as it is the only measure that has the expected negative correlation with *correct* and explains the variation across regulations beyond *length*. These results have natural interpretations. *cyclomatic* is a measure of psychological complexity, like *length*, which may be why it does not capture much beyond what is already captured by *length*. In contrast, *quantity* and *potential* are both measures of problem complexity and were hence expected to capture a dimension not already reflected in *length*, and we indeed find this is the case for *quantity*. Actually, *quantity* has a large marginal effect on *correct*: for a given length, participant, and question order, adding one rule (compared to an average of 4.79) decreases the probability of a correct answer from a baseline of 68.6% to 53.8%. Conversely, once controlling for *quantity* it seems *length* is unrelated to *correct*. *diversity* was introduced by symmetry with *potential*, but it does not rely on any theoretical foundation, and accordingly it performs poorly. *level* is given some theoretical foundations in Halstead (1977), but it is expected to have a different impact for more sophisticated and less sophisticated participants, hence it's correlation with *correct* is a priori ambiguous.

We conclude from this analysis that *quantity* seems to be the best ex-ante measure of mistake-based complexity, and the only one to provide information beyond *length*. For robustness, in Online Appendix OA.4 we repeat the same analysis using OLS regressions of the percentage of correct answers at the regulation level on the different measures, and we obtain very similar results.

3.3 Results on cost-based complexity

We repeat the analysis with cost-based complexity. For each participant i and question t answered by this participant, we know the time taken to answer, denoted $time_{i,t}$. According to Definition 5, we can consider the time taken for each regulation as a measure of cost-based complexity. However, there are three issues that complicate the measurement: (i) A participant may take an abnormally long or short time to answer because he or she misunderstood the regulation. Hence, it is not clear whether wrong answers given after a short time really reflect a low cost-based complexity; (ii) Some participants may have "given up" on some regulations that looked more daunting, and given a random answer after a short amount of time, making again such answers difficult to interpret; (iii) A few correct answers were given after a long time (the maximum being 958 seconds, or about 16 minutes). While it is possible that it actually took that long to the participants to answer, it is likely that they 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 we think are the least likely to be affected by them. Starting with 1,062 observations, we keep only those 728 that correspond to correct answers, which solves issues (i) and (ii). We then delete 6 observations for which $time_{i,t}$ is above 579 seconds (99 percentile of the initial distribution), which alleviates (iii). 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 question fixed effects:

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

[Insert Tables 8 to 10 here.]

We verify that separating operands and operators helps explaining the variation of $time_{i,t}$ across regulations. Table 8 shows that this is the case. Column (2) shows that total operands and operators have a different impact on *time* (p-value below 0.1%), Column (4) and (5) that unique operands and operators have a different impact, including when controlling for *length*, although the statistical significance is weaker (p-values of 8.7% and 8.5%, respectively).

However, and differently from the analysis of mistake-based complexity, we cannot reject the hypothesis that unique logical operators and unique mathematical operators have the same impact (p-value of 17.7%).

Table 9 shows that all the ex-ante measures of complexity except level have the expected positive correlation with *time*. When controlling for *length* in Table 10, we see that *potential* and to a lesser extent *quantity* are significantly and positively correlated with *time*. In terms of magnitude, for a given *length*, participant, and question rank, adding one extra unique operand (compared to an average of 16.66) increases *time* from an average of 132 seconds to 145 seconds.

For robustness, Online Appendix OA.4 shows the results obtained when aggregating answers at the regulation level. Online Appendix OA.5 shows the results obtained with several other filtering choices, including keeping all observations, and winsorizing instead of trimming outliers. In all cases we obtain qualitatively similar results, with the exception that the impact of *quantity*, already weaker in the main specification, becomes not significant at the 10% level in some specifications.

3.4 Discussion

The conclusion of our experimental analysis is that two of the five new ex-ante measures we consider, namely *potential* and *quantity*, seem to be good proxies for ex-post measures of complexity, beyond *length*. Interestingly, these are the two measures that were expected to capture problem complexity, a different dimension from the one captured by *length*. These two measures also seem to capture different subdimensions of problem complexity:

quantity is mostly related to mistake-based complexity, with an extra rule increasing the probability of a mistake by 14.8 percentage points, while *potential* is mostly related to cost-based complexity, with an extra unique operand increasing the time to provide a correct answer by 13 seconds (controlling for *length* in both cases). This suggests that more generally making a regulation easier to understand does not necessarily make it less costly to process, and conversely.

While we believe these results are interesting in their own right, our main conclusion is broader: this methodology inspired by the validation of algorithmic complexity measures in computer science provides a powerful touchstone for testing novel measures of regulatory complexity. Indeed, out of five measures we tested, only two pass the test. Reassuringly for our methodology, these two measures are also the ones that were expected to perform the best ex ante. As we provide the texts of the regulations we used and the results of the experiments online, other researchers have a tool to test any other text-based measure of complexity and compare it to the five we considered.

4 Applications

In this section we discuss the two main possible applications of our approach, and develop tools for these applications. The first one is to apply our ex-ante measures of complexity on various regulatory texts, either in the context of an academic study on the impact of regulatory complexity on economic outcomes, or as part of a policy process to keep track of the complexity of new proposed regulations. Our ex-ante measures can be applied at scale on a variety of texts, provided that one has a sufficiently rich dictionary of operands and operators. We explain in Section 4.1 how we developed such a dictionary. The second application is to use our measures in the context of a normative model of regulation, which could eventually be used by policymakers to quantify the trade-off between the complexity of regulation and other objectives. As a proof of concept, Section 4.2 builds a simple model of the trade-off pointed out by the BCBS between the risk-sensitivity and the simplicity of capital requirements.

4.1 A dictionary for positive analysis: The Dodd-Frank Act

To build a dictionary of operands and operators and prove that our measures can be implemented at a larger scale, we compute our complexity measures for the different titles of the 2010 Dodd-Frank Act. There are two reasons for this choice. First, the Dodd-Frank Act is one of the key regulations introduced after the financial crisis. It has triggered a lot of debate, in particular regarding its perceived complexity. Second, the Dodd-Frank Act touches upon a wide range of issues in finance, so that by classifying the words of the Dodd-Frank Act we created a comprehensive dictionary that can be used for a broad range of other regulatory texts.

The scale and scope of the Dodd-Frank Act also creates three new challenges compared to the more limited example of Basel I.

First, a lot of operands in the Dodd-Frank Act are "n-grams", expressions made of n distinct words. For instance, "Consumer Financial Protection Bureau" should be considered as one operand, not four distinct words. To take this into account, we read the entire Act and manually made a list of all such n-grams (for details see Online Appendix OA.6). We classified each n-gram into a category, and then removed them from further counts. That is, we made sure that "Consumer Financial Protection Bureau" is counted only once as an operand, not once as an operand and then again as four distinct words.

Second, some words in the text can sometimes be used as an operand and sometimes as an operator. The most prominent example is the word "is". In principle, "is" could be a regulatory operator (as in, e.g., "the risk-weight is 20%"). However, it could have a merely grammatical function to indicate the passive voice (e.g., "at the time at which each report is submitted", Sec. 112 (b)). We classify such ambiguous words in the category "other", and hence don't count them in our different measures.[17]

Third, the Act 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)) [...]" How should one deal with such a case? A possible solution would be to include the text referenced in the example as being implicitly part of the Act. 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. After classifying words in the 16 Titles of the Dodd-Frank Act plus its introduction, we created a dictionary containing:

[17] There is necessarily some judgement 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.

429 operators (230 logical operators, 161 regulatory operators, 38 mathematical operators), 5,872 operands, as well as 2,799 "other" words (2,450 legal references, 222 function words, and 127 ambiguous words). Table 11 shows the top 10 words in each category as well as the number of occurrences. Similarly to what we did in Section 2.2, we then compute different measures for the different titles of the Dodd-Frank Act, and the entire act separately. The results are reported in Table 12.

The objective of building this dictionary is that it can be used on other regulatory texts. To test whether the dictionary is rich enough, we conduct the following exercise. For each title i between 1 and 16 of the Dodd-Frank Act, 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 not able to classify based on the alternative dictionary. In addition, we also count the proportion of these unclassified words that are actually operands, operators of different types, and other words. As shown in Table 13, on average across all titles we are able to retrieve 86% 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 more than 96% of operators of all categories, so that measures relying on operators (cyclomatic and diversity) seem the easiest to compute on other texts without having to expand the dictionary.

[Insert Tables 11, 12, and 13 here.]

We made the dictionary of all the classified words in the Dodd-Frank Act available online. In addition, the code for the dashboard we used is available, and can be used to manually enrich our dictionary with words from other regulatory texts. Moreover, for regulatory texts with too many unclassified words our dictionary can be used to train a supervised machine learning algorithm to classify words into operands and operators.

4.2 Normative analysis: "balancing risk-sensitivity and simplicity"

To use our approach for normative purposes, we extend the framework of Section 1 by assuming that the regulator designing the regulation has a model of the economy that associates an entity x and a regulatory action y to some measure of social welfare $U(x, y)$. For a given regulation φ associating each possible entity to a regulatory action, the welfare achieved by the regulation is:

$$\mathcal{W}_0(\varphi, x) = U(x, \varphi(x)). \tag{8}$$

In a standard microeconomic model of regulation, we would solve for the $\varphi^*(x)$ that maximizes $\mathcal{W}_0(\varphi, x)$, and this would define the optimal regulation. We extend this standard case by taking into account that while ideally regulation φ associates x to $\varphi(x)$, the possibility of mistakes at the various steps studied in Section 1 implies that the actual regulatory action may be $\hat{f}(\varphi, x) \neq \varphi(x)$. More precisely, we assume that, for a given x, with probability $\hat{p}(\varphi, x)$ the regulation is correctly implemented and $\hat{f}(\varphi, x) = \varphi(x)$, whereas with the complementary probability $\hat{f}(\varphi, x)$ contains a random mistake following some distribution. In addition, we denote $\hat{t}(\varphi, x)$ the total effort (e.g., hours of work) spent on supervising entity x, and denote w the cost per unit of effort. We can then define a social welfare function that takes into account the costs of complexity:

$$\mathcal{W}(\varphi, x) = \hat{p}(\varphi, x)U(x, \varphi(x)) + (1 - \hat{p}(\varphi, x))\mathbb{E}[U(x, \hat{f}(\varphi, x))|\hat{f}(\varphi, x) \neq \varphi(x)] - w\hat{t}(\varphi, x). \quad (9)$$

Note that the choice of φ affects the probability $\hat{p}(\varphi, x)$ of a mistake, the distribution of $\hat{f}(\varphi, x)$ in case of a mistake, and the total effort $\hat{t}(\varphi, x)$. In particular, according to our definitions in Section 1, a more complex regulation φ means that the probability of a mistake and the effort costs are higher. Hence, there is potentially a trade-off between having a regulation "close" to φ^* and generating regulatory mistakes and costs.

We illustrate this approach with a simple model of risk-sensitive capital requirements. The intense policy debate on the complexity of capital requirements led the Basel Committee to publish a discussion paper on the trade-offs between "risk sensitivity, simplicity and comparability" (Basel Committee on Banking Supervision, 2013). Nine years later, 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 such a normative purpose and be used to think about the optimal level of complexity.

Assume a bank invests in assets that have a certain "asset class" $x \hookrightarrow \mathcal{U}[0, 1]$, and denote y the minimum level of capital the bank must have. In Online Appendix OA.7, we derive for illustration a simple function $U(x, y)$ in a model of bank risk-shifting, such that capital regulation can improve welfare.[18] We consider the following family of bank capital

[18]In future research going beyond this proof of concept, one should use a quantitative model of the economy rich enough to accommodate different regulations. 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.

regulations:

`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 y is the amount of equity the bank is required to have for an asset belonging to class y, the $\bar{x}_i$ are thresholds chosen by the regulator, the E_i^* are capital levels chosen by the regulator, and I is the number of risk buckets considered by the regulator. We denote $\mathcal{W}_0(I) = \mathbb{E}[U(x, y)]$ the expected economic welfare if the regulation above is perfectly implemented, with the E_i^* chosen optimally. The Appendix shows that this welfare increases in I. This captures in a stylized way the benefits of risk-sensitivity, the first leg of the trade-off described in Basel Committee on Banking Supervision (2013).

The second leg, simplicity, the opposite of complexity, can be captured by our complexity measures. In the regulation above, 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.

For illustration, we then use the estimates of Section 3 to translate the measures into the probability $\hat{p}(I)$ with which the regulation is applied without mistake and the total effort $\hat{t}(I)$. More specifically, we use the estimates from specification (3) in Table 6 and (4) in

Table 9 to compute, for every I:[19]

$$\hat{p}(I) = \Phi(2.877 - 0.507 quantity(I)) \quad (10)$$

$$\hat{t}(I) = 4.055 + 7.965 potential(I). \quad (11)$$

We can now quantitatively measure how increasing the number of distinct asset classes affects $\mathcal{W}_0(I)$ (welfare absent costs of complexity), $\hat{p}(I)$, $\hat{t}(I)$, and welfare $\mathcal{W}(I)$ (including the costs of complexity). Figure 3 displays the results for the particular welfare function derived in Online Appendix OA.7.[20] In particular, while $\mathcal{W}_0(I)$ is strictly increasing in I, we obtain that $\mathcal{W}(I)$ is bell-shaped and for more than $I = 3$ risk-buckets the costs of complexity outweigh the benefits. Hence, the policymaker can compute the optimal trade-off between "risk-sensitivity" and "simplicity".

[Insert Fig. 3 here.]

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 policymakers to complete two additional tasks: (i) develop quantitative models of regulation, rich enough to estimate the welfare impact of different regulatory

[19]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.

[20]In addition to the parameters discussed above, we assume that conditionally on a mistake being made $\hat{f}(\varphi, x)$ is uniformly distributed over $[0, 1]$. In addition, we assume $\lambda = 0.025, \delta = 0.01, p = 0.05, w = 0$. The effort costs then play no role in the graphs, but obviously a higher w would make $\mathcal{W}(I)$ decrease more quickly in I.

alternatives; (ii) run richer and more robust experiments to have a more precise view of the costs of psychological complexity for different audiences.[21]

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 six measures of regulatory complexity that can be applied to large scale regulatory texts, conduct a validation test that can be applied to any text-based measure, and study the trade-off between the costs and benefits of more complex regulations in a normative model.

The present work is only a first step in applying this new approach to the study of regulatory complexity, and is meant as a "proof of concept". We believe our first results are encouraging and highlight several promising avenues for future research.

First, our dictionary will allow other interested researchers to compute complexity measures for other regulatory texts and compare them to those we produced for Basel I and the Dodd-Frank Act. One can for instance compare the complexity of different regulatory topics, different updates of the same regulation, different national implementations, etc. A rich database of the complexity of different regulations could eventually be used in empirical studies aiming at testing some of the mechanisms that have been proposed in the theoretical literature.

[21]More specifically, the type of experiment we consider in Section 3 should ideally be reproduced with regulations actually under discussion, and with participants closer to the actual audience of regulatory texts (bankers, lawyers, regulators, etc.).

Second, the experiments we conducted and the validation criteria we propose allow interested researchers to test any alternative text-based measure and compare it to the six measures considered in this study. They could also serve as a useful benchmarking tool for policymakers drafting new regulations.

Finally, the measures of complexity we propose can be computed also on models of regulation, opening the possibility for policymakers to conduct the trade-off between the precision and the complexity of regulation under the guidance of a quantitative model.

References

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. 17, 48

Amadxarif, Z., Brookes, J., Garbarino, N., Patel, R. and Walczak, E. (2019). *The language of rules: textual complexity in banking reforms*. Staff Working Paper No.834, Bank of England. 19, 20, 48

Arora, S., Barak, B., Brunnermeier, M. and Ge, R. (2009). *Computational Complexity and Information Asymmetry in Financial Products*. Working paper. 6

Asriyan, V., Foarta, D. and Vanasco, V. (2021). The good, the bad and the complex: Product design with imperfect information. *American Economic Journal: Microeconomics*, **Forthcoming**. 6

Basel Committee on Banking Supervision (1988). *International convergence of capital measurement and capital standards*. Discussion paper, BCBS. 20

Basel Committee on Banking Supervision (2013). *The regulatory framework: balancing risk sensitivity, simplicity and comparability*. Discussion paper. 1, 5, 39, 40

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. 39

Bennett, B., Garvey, G., Milbourn, T. and Wang, Z. (2019). *The Role of Stock Price Informativeness in Compensation Complexity*. Working Paper Series 2019-7, Ohio State University, Charles A. Dice Center for Research in Financial Economics. 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, 48

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*. NBER Working Papers 26856, National Bureau of Economic Research, Inc. 15, 48

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. 3

Carlin, B. I. (2009). Strategic price complexity in retail financial markets. *Journal of Financial Economics*, **91** (3), 278 – 287. 6

—, Kogan, S. and Lowery, R. (2013). Trading complex assets. *The Journal of Finance*, **68** (5), 1937–1960. 6

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

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). 5

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. 6

Fenton, N. E. (1994). Software measurement: A necessary scientific basis. *IEEE Transactions on Software Engineering*, **20** (3), 199–206. 12

Foarta, D. and Morelli, M. (2022). *The Common Determinants of Legislative and Regulatory Complexity*. Working paper. 6

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. and Laibson, D. (2006). Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets. *The Quarterly Journal of Economics*, **121** (2), 505–540. 6

Gai, P., 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

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

Gutiérrez, G. and Philippon, T. (2019). *The Failure of Free Entry*. Working paper w26001, NBER. 6

Hakenes, H. and Schnabel, I. (2012). *Regulatory Capture By Sophistication*. Working paper. 5

Haldane, A. and Madouros, V. (2012). The dog and the frisbee. *Proceedings - Economic Policy Symposium - Jackson Hole*, pp. 109–159. 1, 17, 48

Halstead, M. H. (1977). *Elements of Software Science*. Elsevier. 2, 8, 15, 17, 18, 31, 48

Herring, R. J. (2018). The evolving complexity of capital regulation. *Journal of Financial Services Research*, **53** (2), 183–205. 17

IEEE (1990). IEEE Standard Glossary of Software Engineering Terminology. *IEEE Std 610.12-1990*, pp. 1–84. 8

Kalmenovitz, J. (2023). Regulatory Intensity and Firm-Specific Exposure. *The Review of Financial Studies*, hhad001. 6, 15, 48

—, Lowry, M. and Volkova, E. (2022). *Regulatory Fragmentation*. Working paper. 19, 48

Katz, D. and Bommarito, M. (2014). Measuring the complexity of the law: the united states code. *Artificial Intelligence and Law*, **22**. 19, 48

Kendall, C. and Oprea, R. (2021). *On the Complexity of Forming Mental Models*. Working paper. 7

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, 48

Knuth, D. E. (1973). *Fundamental Algorithms: The Art of Computer Programming*, vol. 1. Addison-Wesley, 2nd edn. 8

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). 2, 17, 19

Loughran, T. and McDonald, B. (2014). Measuring readability in financial disclosures. *The Journal of Finance*, **69** (4), 1643–1671. 5

Maas, H.-D. (1972). Über den Zusammenhang zwischen Wortschatzumfang und Länge eines Textes. *Zeitschrift für Literaturwissenschaft und Linguistik*, **2** (8). 19, 48

McCabe, T. J. (1976). A complexity measure. *IEEE Transactions on Software Engineering*, **SE-2** (4), 308–320. 12, 17, 48

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. 17

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. 6

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

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

Simkovic, M. and Zhang, M. B. (2020). *Regulation and Technology-Driven Entry: Measurement and Micro-Evidence*. Working paper. 15, 48

Singla, S. (2022). *Regulatory Costs and Market Power*. Working paper. 15, 48

Weyuker, E. J. (1988). Evaluating software complexity measures. *IEEE Transactions on Software Engineering*, **14** (9), 1357–1365. 12

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. 5

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. 12

Zuse, H. (1990). *Software Complexity: Measures and Methods*. New York, NY, United States: de Gruyter, 1st edn. 5, 8, 14

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

A Tables

Table 1: Summary of Measures of Regulatory Complexity.

Name	Source	Formula	Complexity Dimension	Approach
Length	e.g., Haldane and Madouros (2012)	N	Psychological	Ex-ante
Cyclomatic complexity	McCabe (1976)	N_L	Psychological	Ex-ante
Quantity of regulations	Al-Ubaydli and McLaughlin (2017)	N_R	Problem	Ex-ante
Potential volume	This paper and Halstead (1977)	$2 + \eta_{OD}$	Problem	Ex-ante
Operator diversity	This paper	η_{OR}	Psychological	Ex-ante
Level	This paper and Halstead (1977)	$\frac{2+\eta_{OD}}{N}$	Psychological	Ex-ante
RegFragmentation	Kalmenovitz *et al.* (2022)	-	Computational	Ex-ante
Average word length	e.g., Amadxarif *et al.* (2019)	-	Psychological	Ex-ante
Lexical diversity	Maas (1972)	-	Psychological	Ex-ante
Readability metric	(Kincaid *et al.*, 1975)	-	Psychological	Ex-ante
Shannon's entropy	e.g., Katz and Bommarito (2014)	-	Psychological	Ex-ante
PageRank	Amadxarif *et al.* (2019)	-	Psychological	Ex-ante
Network Centralities	Boulet *et al.* (2011)	-	Psychological	Ex-ante
Regulation Index	Simkovic and Zhang (2020)	-	Computational	Ex-post
RegIn	Kalmenovitz (2023)	-	Computational	Ex-post
NetReg	Calomiris *et al.* (2020)	-	Computational	Ex-post
Regulatory costs	Singla (2022)	-	Computational	Ex-post

Table 2: Correlation coefficients between the measures based on the algorithmic representation of Basel I and the measures based on the text of Basel I.

	Pearson	Spearman
length	0.76	0.84
cyclomatic	0.41	0.64
quantity	1	1
potential	0.82	0.8
diversity	0.4	0.48
level	0.39	0.43

Table 3: Summary statistics on complexity measures - sample of 38 randomly generated regulations.

	mean	sd	min	max
length	31.82	12.46	10.00	57.00
cyclomatic	5.32	3.58	1.00	13.00
quantity	4.79	1.19	2.00	6.00
potential	16.66	5.45	7.00	28.00
diversity	4.24	0.94	3.00	7.00
level	0.55	0.09	0.39	0.70

Table 4: Pairwise correlations between complexity measures, sample of 38 randomly generated regulations.

Panel A: Pearson Correlation Coefficients

	length	*cyclomatic*	*quantity*	*potential*	*diversity*	*level*
length	1	0.89	0.87	0.92	0.8	-0.7
cyclomatic	0.89	1	0.68	0.72	0.63	-0.79
quantity	0.87	0.68	1	0.82	0.7	-0.69
potential	0.92	0.72	0.82	1	0.83	-0.4
diversity	0.8	0.63	0.7	0.83	1	-0.43
level	-0.7	-0.79	-0.69	-0.4	-0.43	1

Panel B: Spearman Rank Correlation Coefficients

	length	*cyclomatic*	*quantity*	*potential*	*diversity*	*level*
length	1	0.89	0.85	0.91	0.83	-0.69
cyclomatic	0.89	1	0.75	0.7	0.67	-0.86
quantity	0.85	0.75	1	0.8	0.69	-0.65
potential	0.91	0.7	0.8	1	0.86	-0.39
diversity	0.83	0.67	0.69	0.86	1	-0.45
level	-0.69	-0.86	-0.65	-0.39	-0.45	1

Table 5: Correlation of mistake-based complexity with operands and operators. This table reports the coefficients, t-statistics (in brackets), and Pseudo-R^2, of probit regressions of $correct_{i,t}$ over different counts of total and unique operands and operators, with participant and question fixed effects.

	(1)	(2)	(3)	(4)	(5)	(6)
$length = N_{OD} + N_{OR}$	-0.037***				-0.044***	-0.042***
	(-7.95)				(-3.83)	(-3.59)
N_{OD}		-0.062***				
		(-3.94)				
N_{OR}		0.006				
		(0.22)				
$\eta_{OD} + \eta_{OR}$			-0.065***			
			(-7.27)			
η_{OD}				-0.144***	-0.070***	-0.093***
				(-7.44)	(-2.60)	(-3.32)
η_{OR}				0.451***	0.559***	
				(4.04)	(4.77)	
η_M						1.103***
						(6.30)
η_L						0.396***
						(3.22)
Pseudo-R^2	0.243	0.246	0.232	0.253	0.266	0.283

Table 6: Correlation of mistake-based complexity with ex-ante measures of complexity. This table reports the coefficients, t-statistics (in brackets), and Pseudo-R^2, of probit regressions of $correct_{i,t}$ over the six ex-ante measures of complexity separately, with participant and question fixed effects.

	(1)	(2)	(3)	(4)	(5)	(6)
length	-0.037***					
	(-7.95)					
cyclomatic		-0.095***				
		(-6.23)				
quantity			-0.507***			
			(-9.34)			
potential				-0.078***		
				(-7.59)		
diversity					-0.243***	
					(-4.34)	
level						4.061***
						(6.52)
Pseudo-R^2	0.243	0.217	0.277	0.237	0.198	0.221

Table 7: Correlation of mistake-based complexity with ex-ante measures of complexity, controlling for *length*.This table reports the coefficients, t-statistics (in brackets), and Pseudo-R^2, of probit regressions of $correct_{i,t}$ over *length* and each of the five other measures of complexity separately, with participant and question fixed effects.

	(1)	(2)	(3)	(4)	(5)	(6)
length	-0.037***	-0.055***	0.005	-0.030***	-0.065***	-0.031***
	(-7.95)	(-5.69)	(0.62)	(-2.75)	(-7.85)	(-5.16)
cyclomatic		0.068**				
		(2.13)				
quantity			-0.554***			
			(-5.92)			
potential				-0.017		
				(-0.70)		
diversity					0.438***	
					(4.16)	
level						1.275
						(1.55)
Pseudo-R^2	0.243	0.248	0.277	0.244	0.260	0.246

Table 8: Correlation of cost-based complexity with operands and operators. This table reports the coefficients, t-statistics (in brackets), and adjusted R^2, of OLS regressions of $time_{i,t}$ over different counts of total and unique operands and operators, with participant and question fixed effects. The sample is restricted to correct answers with $time \leq 579$.

	(1)	(2)	(3)	(4)	(5)	(6)
$length = N_{OD} + N_{OR}$	3.388***				0.654	0.597
	(14.39)				(1.01)	(0.92)
N_{OD}		7.102***				
		(8.04)				
N_{OR}		-3.193**				
		(-2.09)				
$\eta_{OD} + \eta_{OR}$			6.886***			
			(15.27)			
η_{OD}				8.452***	7.272***	7.601***
				(8.30)	(4.69)	(4.85)
η_{OR}				-3.309	-4.550	
				(-0.55)	(-0.75)	
η_M						-13.953
						(-1.51)
η_L						-1.296
						(-0.20)
R^2_a	0.445	0.461	0.462	0.464	0.464	0.465

Table 9: Correlation of cost-based complexity with ex-ante measures of complexity. This table reports the coefficients, t-statistics (in brackets), and adjusted R^2, of univariate regressions of *time* over the six ex-ante measures of complexity separately, with participant and question fixed effects. The sample is restricted to correct answers with $time \leq 579$.

	(1)	(2)	(3)	(4)	(5)	(6)
length	3.388***					
	(14.39)					
cyclomatic		9.338***				
		(10.18)				
quantity			32.996***			
			(13.80)			
potential				7.965***		
				(15.39)		
diversity					39.316***	
					(12.28)	
level						-265.538***
						(-6.90)
R_a^2	0.445	0.363	0.433	0.465	0.403	0.308

Table 10: Correlation of cost-based complexity with ex-ante measures of complexity, controlling for *length*. This table reports the coefficients, t-statistics (in brackets), and adjusted R^2, of regressions of *correct* over *length* and each of the five other measures of complexity separately, with participant and question fixed effects. The sample is restricted to correct answers with $time \leq 579$.

	(1)	(2)	(3)	(4)	(5)	(6)
length	3.388***	5.371***	2.234***	0.556	2.946***	4.183***
	(14.39)	(10.46)	(4.39)	(0.88)	(6.81)	(12.74)
cyclomatic		-8.084***				
		(-4.33)				
quantity			13.072**			
			(2.56)			
potential				6.807***		
				(4.79)		
diversity					6.899	
					(1.22)	
level						165.462***
						(3.44)
R_a^2	0.445	0.461	0.450	0.465	0.445	0.455

Table 11: Top 10 words in each category, entire Dodd-Frank Act.

Operands		Operators					
		Regulatory		Logical		Mathematical	
COMMISSION	1573	SHALL	3595	AND	9352	ADDING	267
PERSON	920	AMENDED	651	OR	8928	ADDITIONAL	125
BUREAU	788	REQUIRED	548	ANY	4007	TOTAL	101
CORPORATION	771	ESTABLISHED	282	AS	2646	MINIMUM	86
INFORMATION	731	ESTABLISH	247	OTHER	1546	EXCEED	70
DATE	692	REQUIRE	220	NOT	1128	OVER	69
STATE	607	PRESCRIBED	219	AFTER	906	ADDED	68
APPROPRIATE	569	DETERMINES	212	INCLUDING	761	INCREASE	48
REPORT	564	PRESCRIBE	202	EACH	687	MAXIMUM	41
AUTHORITY	552	DETERMINE	181	WITH RESPECT TO	678	MINIMIZE	28

Table 12: Complexity measures of the 16 titles of the Dodd-Frank Act.

Title	length	cyclomatic	quantity	potential	diversity	level
1	10581	2271	729	1389	190	0.13
2	16388	4479	852	1559	212	0.10
3	7269	2052	335	889	130	0.12
4	1938	466	117	444	94	0.23
5	3539	828	163	784	107	0.22
6	7662	1960	503	1040	157	0.14
7	32055	8195	2195	2127	231	0.07
8	3852	882	263	634	119	0.16
9	26319	5826	1614	2533	277	0.10
10	31872	7938	1916	2724	277	0.09
11	3277	764	220	674	113	0.21
12	780	155	49	248	42	0.32
13	575	141	31	152	32	0.26
14	16126	3389	866	2068	237	0.13
15	2013	376	106	549	76	0.27
16	68	22	3	32	14	0.47
Entire Act	164314	39744	9962	5874	429	0.04

Table 13: Fraction of words found in each title of the Dodd-Frank Act, using dictionaries built from the other titles only.

Title	All	Operands	Operators			Other
			Logical	Regulatory	Mathematical	
1	0.89	0.89	0.92	1.00	0.88	0.84
2	0.92	0.94	0.98	0.97	0.93	0.81
3	0.83	0.93	1.00	0.96	1.00	0.66
4	0.93	0.92	0.98	1.00	1.00	0.91
5	0.87	0.84	1.00	0.97	1.00	0.90
6	0.86	0.90	0.98	0.98	0.92	0.73
7	0.80	0.83	0.95	0.98	0.80	0.70
8	0.94	0.95	1.00	1.00	1.00	0.88
9	0.77	0.81	0.93	0.94	0.95	0.60
10	0.75	0.81	0.91	0.93	0.90	0.55
11	0.90	0.91	0.97	0.97	1.00	0.84
12	0.95	0.95	1.00	1.00	1.00	0.95
13	0.87	0.89	1.00	1.00	1.00	0.80
14	0.77	0.80	0.90	0.84	0.95	0.64
15	0.85	0.84	0.98	0.97	1.00	0.85
16	0.91	0.87	1.00	1.00	.	0.91
Average	0.86	0.88	0.97	0.97	0.96	0.79

B Figures

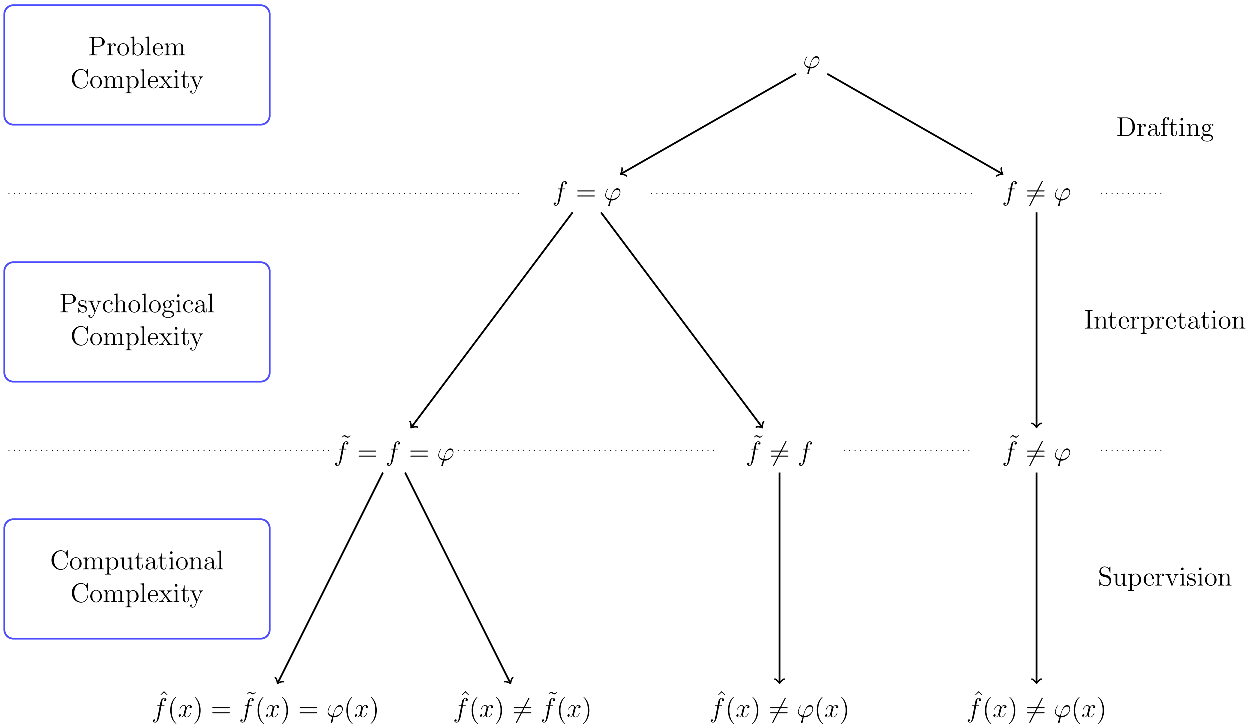


Figure 1: How problem- psychological- and computational complexity affect the likelihood of correctly solving the regulatory problem φ for a given entity e. If the regulation is drafted incorrectly, then $f \neq \varphi$. If a correct regulation is interpreted incorrectly, then $\tilde{f} \neq f$. Lastly, if a mistake is made at the supervision stage then $\hat{f}(x) \neq \tilde{f}(e)$.

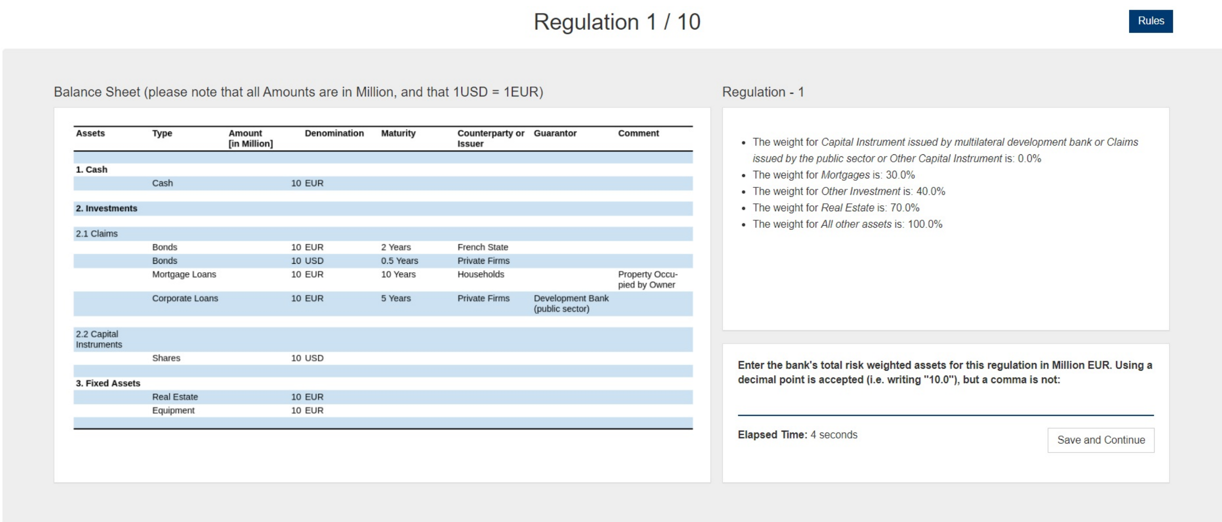


Figure 2: Online experiment - Test round.

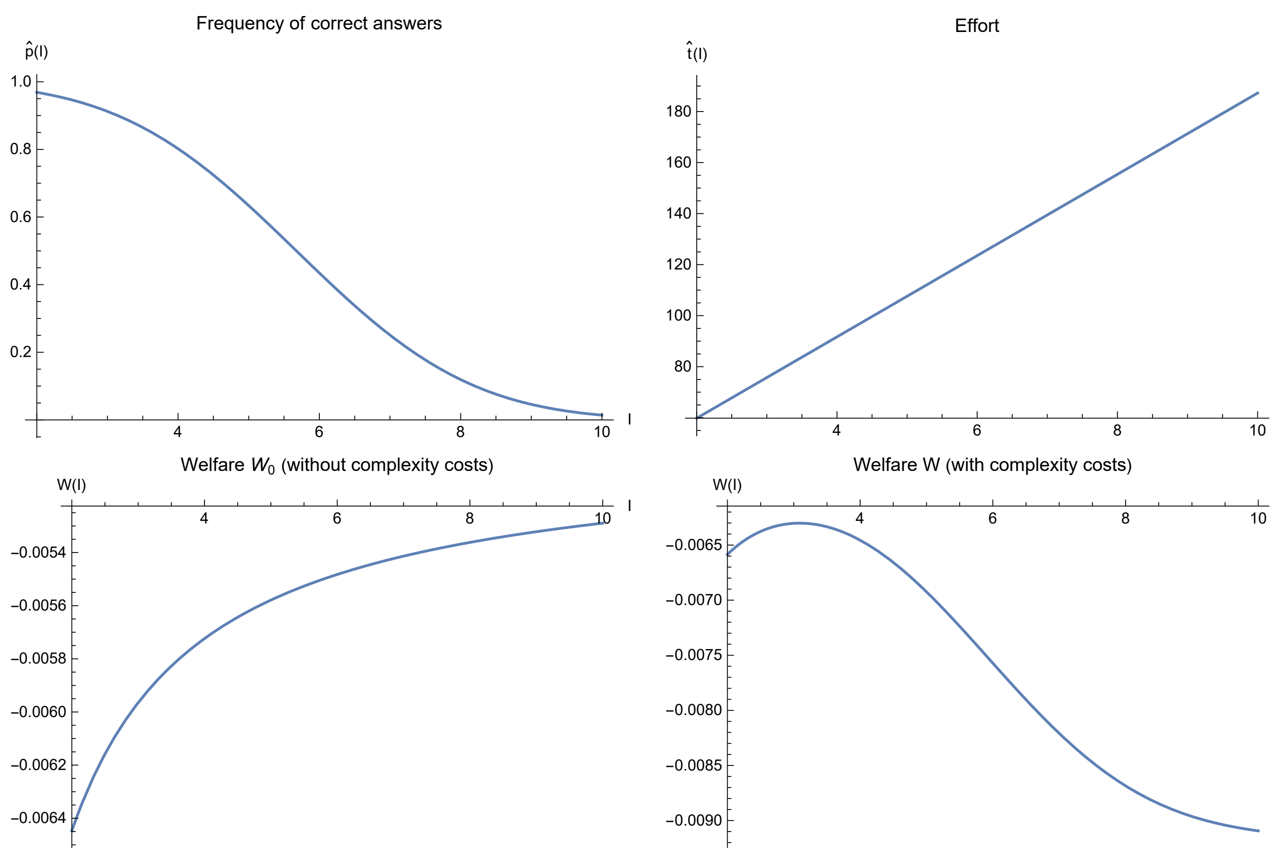


Figure 3: Frequency of correct implementation, effort spent, social welfare without complexity costs, and social welfare with complexity costs, as functions of the number I of risk buckets.

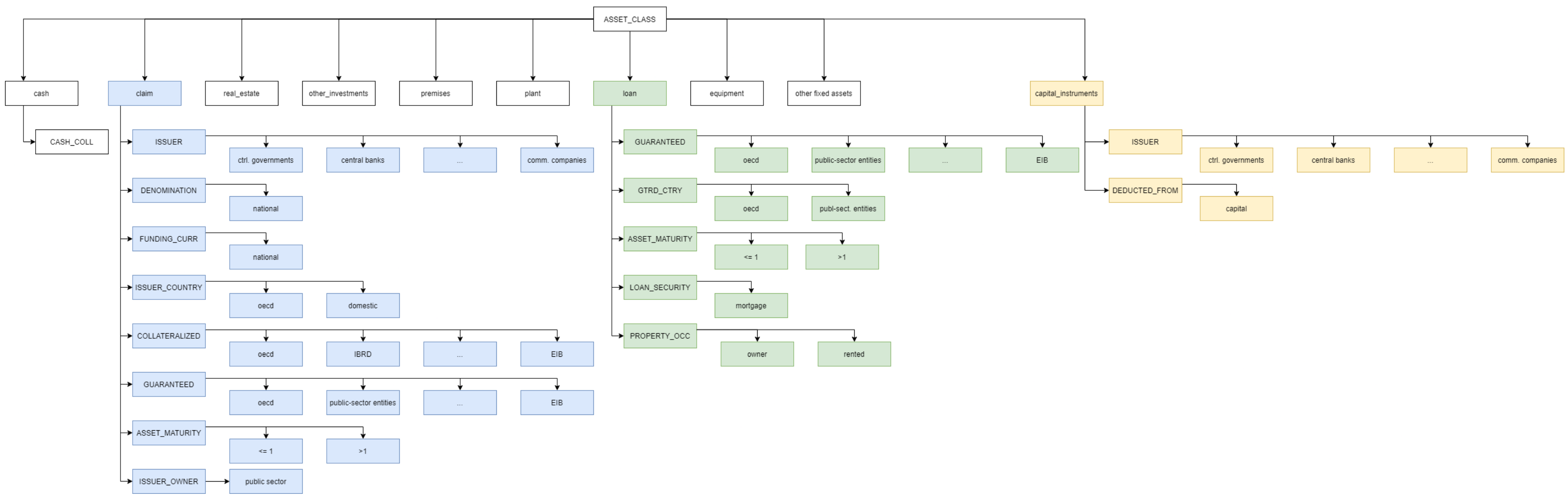


Figure 4: Decision tree with possible values for the "random regulation generator". Each box is an Economic Operand. Arrows indicate possible values of an Economic Operand in a bank's balance sheet. Colors are used for visual guidance only to make it easier to see which choices are part of the same branch in the tree.

C 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%

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-weighted assets

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)

<table>
<tr><td>

```
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;
```

</td><td>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</td></tr>
<tr><td>

```
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;
```

</td><td>Claims on banks incorporated in the OECD and loans guaranteed by OECD incorporated banks shall have a 20% risk weight</td></tr>
<tr><td>

```
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;
```

</td><td>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</td></tr>
<tr><td>

```
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;
```

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

`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;`
`IF (`
`    (ASSET_CLASS == "real_estate" OR ASSET_CLASS == "other_investments") OR`
`) THEN:`
`  risk_weight = 1.0;`
	Premises, plant and equipment and other fixed assets shall have a 100% risk weight
	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. We observe that the text-based measures tend to be less correlated with each other than the algorithm-based measures.

Table OA.1: Complexity measures of the 19 items of Basel I (algorithmic version).

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, across the 19 items of Basel I (algorithmic version). *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 in each category in Basel I (text version).

Operands		Operators: Regulatory		Logical		Mathematical	
risk weight	19	have	19	and	12	up to	2
claims	15	shall	19	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 of the 19 items of Basel I (text version).

Regulation	*length*	*cyclomatic*	*quantity*	*potential*	*diversity*	*level*
1a	5	0	2	5	2	1
1b	16	2	2	12	3	0.75
1c	12	2	2	9	4	0.75
1d	16	1	2	13	3	0.81
2a	22	3	2	18	5	0.82
3a	21	2	2	17	4	0.81
3b	14	1	2	9	3	0.64
3c	26	3	2	13	5	0.5
3d	18	3	2	14	5	0.78
3e	8	0	2	8	2	1
4a	15	2	2	13	3	0.87
5a	7	0	2	7	2	1
5b	13	1	2	11	4	0.85
5c	17	3	2	12	6	0.71
5d	10	0	2	10	2	1
5e	12	3	2	9	4	0.75
5f	15	5	2	10	6	0.67
5g	12	2	2	9	4	0.75
5h	7	1	2	5	4	0.71
Total	266	34	38	69	14	0.26

Table OA.5: Pairwise correlations between complexity measures, across the 19 items of Basel I (text version). *quantity* is not included, as it is constant across items.

Panel A: Pearson Correlation Coefficients

	length	*cyclomatic*	*potential*	*diversity*	*level*
length	1	0.65	0.88	0.63	-0.62
cyclomatic	0.65	1	0.48	0.89	-0.69
potential	0.88	0.48	1	0.44	-0.24
diversity	0.63	0.89	0.44	1	-0.72
level	-0.62	-0.69	-0.24	-0.72	1

Panel B: Spearman Rank Correlation Coefficients

	length	*cyclomatic*	*potential*	*diversity*	*level*
length	1	0.71	0.93	0.64	-0.42
cyclomatic	0.71	1	0.56	0.89	-0.64
potential	0.93	0.56	1	0.47	-0.1
diversity	0.64	0.89	0.47	1	-0.67
level	-0.42	-0.64	-0.1	-0.67	1

OA.3 All pages of the experiment's website

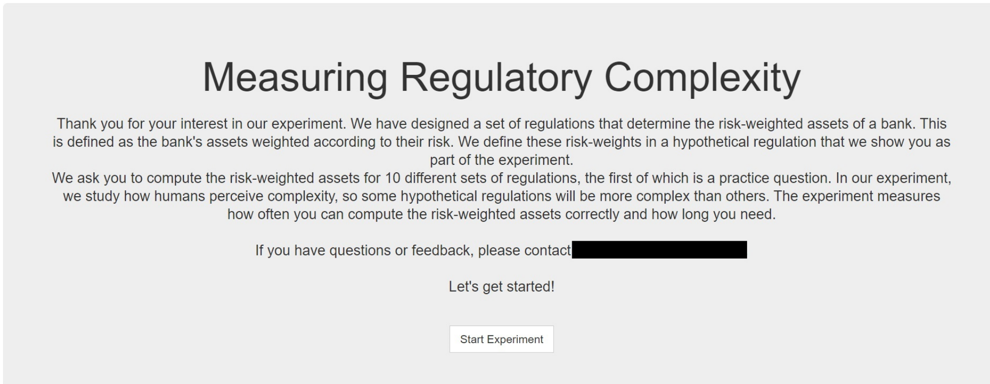
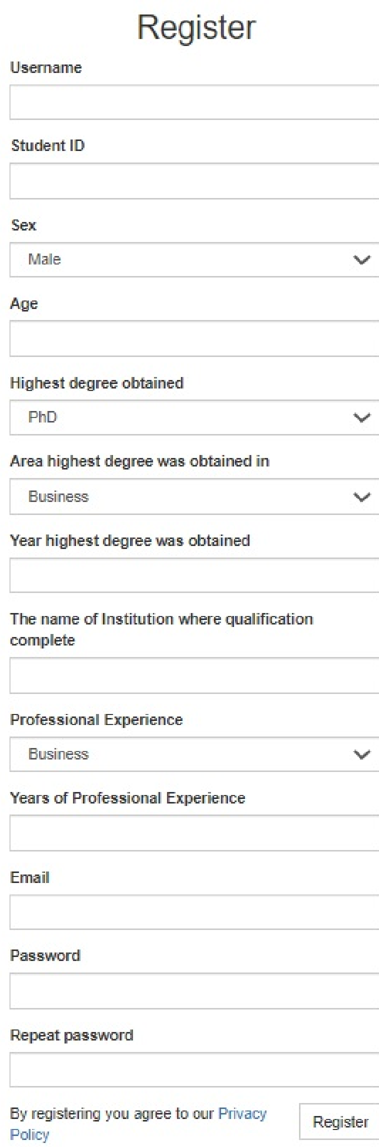


Figure OA.1: Online experiment - Welcome page.



Register

Username

Student ID

Sex

Male

Age

Highest degree obtained

PhD

Area highest degree was obtained in

Business

Year highest degree was obtained

The name of Institution where qualification complete

Professional Experience

Business

Years of Professional Experience

Email

Password

Repeat password

By registering you agree to our Privacy Policy

Register

Figure OA.2: Online experiment - Registration page.

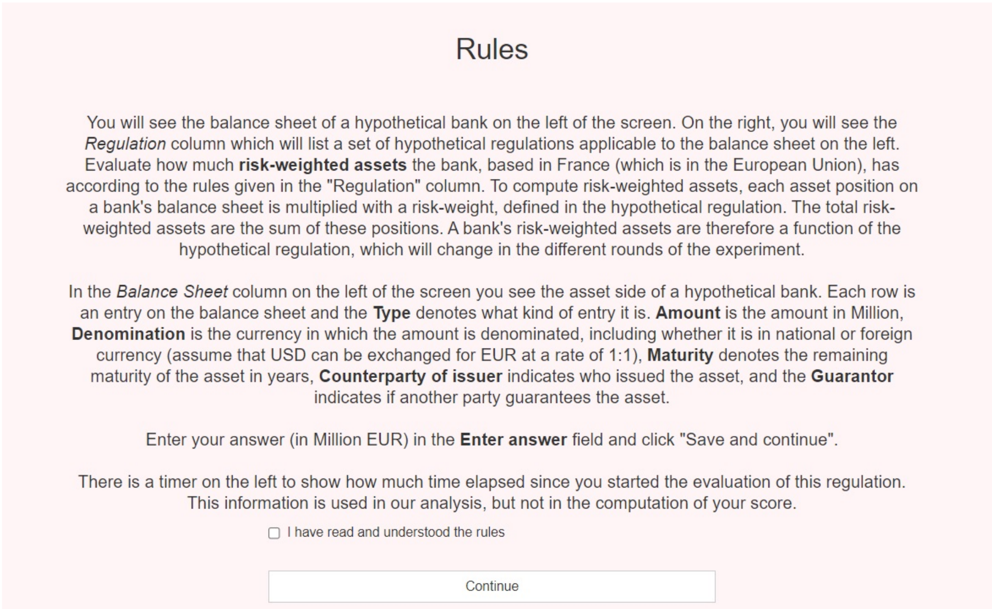


Figure OA.3: Online experiment - Instructions page.

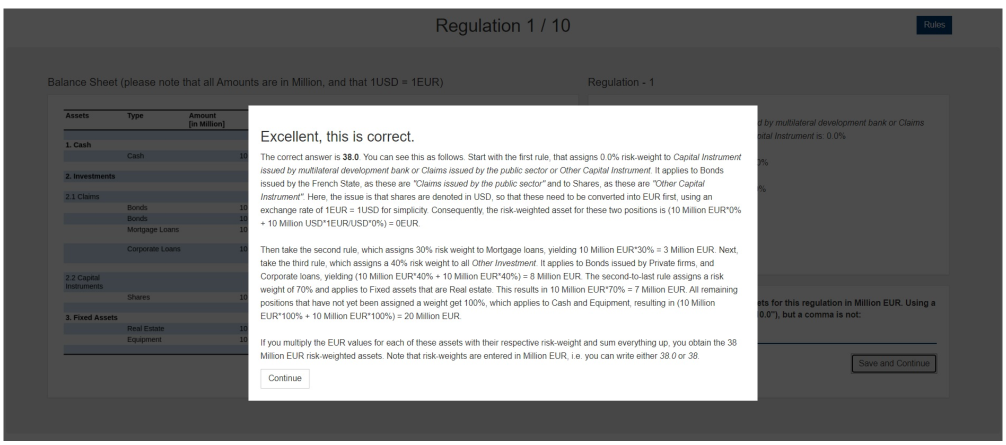


Figure OA.4: Online experiment - Feedback after correct answer in the test round.

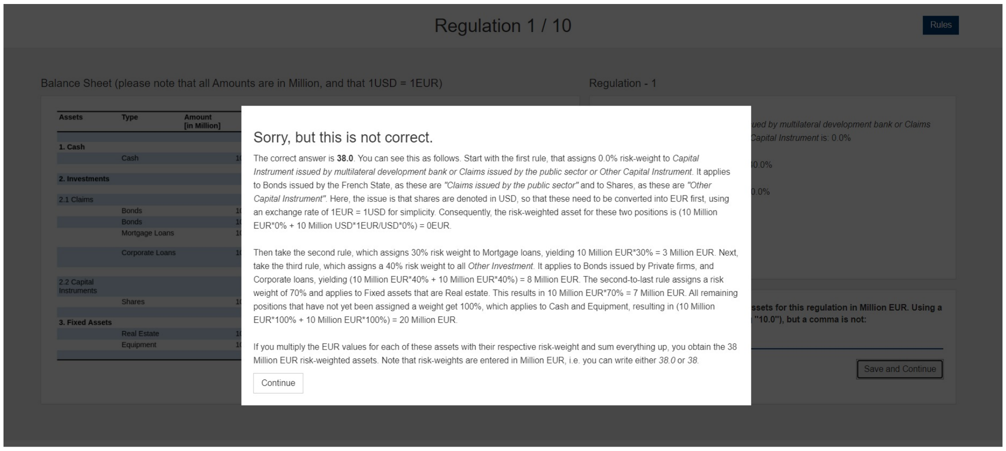


Figure OA.5: Online experiment - Feedback after wrong answer in the test round.

OA.4 Experiments - Regulation-level analysis

In this section we replicate the results of Section 3 after aggregating the answers of all participants at the regulation level. For each regulation $j \in \{1, 2..., 38\}$ we compute the average proportion $correct_j$ of correct answers and the average time taken $time_j$ (excluding incorrect answers and times above 579 seconds, as in the main analysis of Section 3). This gives us a database with 38 observations, one for each regulation. We then run OLS regressions of $correct_j$ and $time_j$ on the same measures of complexity as in Section 3. Tables OA.6 to OA.8 below correspond to Tables 5 to 7, and Tables OA.9 to OA.11 correspond to 8 to 10. The results are qualitatively the same as in our preferred specification at the participant-question level.

Table OA.6: Correlation of mistake-based complexity with operands and operators. This table reports the coefficients, t-statistics (in brackets), and adjusted R^2, of OLS regressions of $correct_j$ over different counts of total and unique operands and operators.

	(1)	(2)	(3)	(4)	(5)	(6)
$length = N_{OD} + N_{OR}$	-0.007***				-0.007	-0.006
	(-3.63)				(-1.41)	(-1.35)
N_{OD}		-0.013*				
		(-1.75)				
N_{OR}		0.003				
		(0.23)				
$\eta_{OD} + \eta_{OR}$			-0.013***			
			(-3.26)			
η_{OD}				-0.029***	-0.017	-0.020*
				(-3.88)	(-1.42)	(-1.82)
η_{OR}				0.097**	0.107**	
				(2.21)	(2.45)	
η_M						0.217***
						(3.37)
η_L						0.067
						(1.48)
R_a^2	0.247	0.240	0.206	0.308	0.327	0.398

Table OA.7: Correlation of mistake-based complexity with ex-ante measures of complexity. This table reports the coefficients, t-statistics (in brackets), and adjusted R^2, of OLS regressions of *correct* over the six ex-ante measures of complexity separately.

	(1)	(2)	(3)	(4)	(5)	(6)
length	-0.007***					
	(-3.63)					
cyclomatic		-0.019**				
		(-2.60)				
quantity			-0.091***			
			(-5.04)			
potential				-0.016***		
				(-3.51)		
diversity					-0.045	
					(-1.56)	
level						0.813***
						(2.78)
R_a^2	0.247	0.134	0.398	0.234	0.037	0.154

Table OA.8: Correlation of mistake-based complexity with ex-ante measures of complexity, controlling for *length*. This table reports the coefficients, t-statistics (in brackets), and adjusted R^2, of OLS regressions of *correct* over *length* and each of the five other measures of complexity separately.

	(1)	(2)	(3)	(4)	(5)	(6)
length	-0.007***	-0.011**	0.002	-0.005	-0.012***	-0.006**
	(-3.63)	(-2.54)	(0.63)	(-0.93)	(-3.88)	(-2.17)
cyclomatic		0.015				
		(1.01)				
quantity			-0.111***			
			(-3.04)			
potential				-0.006		
				(-0.48)		
diversity					0.083*	
					(2.03)	
level						0.225
						(0.58)
R^2_a	0.247	0.248	0.388	0.231	0.307	0.233

Table OA.9: Correlation of cost-based complexity with operands and operators. This table reports the coefficients, t-statistics (in brackets), and adjusted R^2, of OLS regressions of *time* over different counts of total and unique operands and operators. The sample is restricted to correct answers with $time \leq 579$.

	(1)	(2)	(3)	(4)	(5)	(6)
$length = N_{OD} + N_{OR}$	3.256***				0.570	0.576
	(7.11)				(0.52)	(0.52)
N_{OD}		7.171***				
		(4.49)				
N_{OR}		-3.523				
		(-1.30)				
$\eta_{OD} + \eta_{OR}$			6.809***			
			(8.07)			
η_{OD}				9.666***	8.594***	8.555***
				(5.77)	(3.23)	(3.13)
η_{OR}				-12.026	-12.925	
				(-1.24)	(-1.30)	
η_M						-11.821
						(-0.75)
η_L						-13.331
						(-1.21)
R_a^2	0.573	0.629	0.634	0.661	0.654	0.643

Table OA.10: Correlation of cost-based complexity with ex-ante measures of complexity. This table reports the coefficients, t-statistics (in brackets), and adjusted R^2, of univariate regressions of *time* over the six ex-ante measures of complexity separately. The sample is restricted to correct answers with *time* ≤ 579.

	(1)	(2)	(3)	(4)	(5)	(6)
length	3.256***					
	(7.11)					
cyclomatic		8.310***				
		(4.06)				
quantity			33.663***			
			(6.88)			
potential				7.939***		
				(8.45)		
diversity					34.465***	
					(4.64)	
level						-228.199**
						(-2.41)
R_a^2	0.573	0.295	0.556	0.656	0.357	0.115

Table OA.11: Correlation of cost-based complexity with ex-ante measures of complexity, controlling for *length*. This table reports the coefficients, t-statistics (in brackets), and adjusted R^2, of regressions of *correct* over *length* and each of the five other measures of complexity separately. The sample is restricted to correct answers with *time* $\leq$ 579.

	(1)	(2)	(3)	(4)	(5)	(6)
length	3.256***	5.549***	1.899**	0.324	3.286***	4.177***
	(7.11)	(5.93)	(2.12)	(0.30)	(4.20)	(6.87)
cyclomatic		-8.940***				
		(-2.74)				
quantity			16.403*			
			(1.75)			
potential				7.254***		
				(2.93)		
diversity					-0.487	
					(-0.05)	
level						189.632**
						(2.17)
R^2_a	0.573	0.638	0.596	0.647	0.560	0.613

OA.5 Experiments - Alternative filters

In this section we check that the results reported in Tables 9 and 10 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 (Tables OA.12 and OA.13) ; (ii) we keep the outliers but exclude incorrect answers (Tables OA.14 and OA.15) ; (iii) we keep incorrect answers and exclude outliers with time above 579 seconds (Tables OA.16 and OA.17) ; (iv) we keep all observations (Tables OA.18 and OA.19).

The results of 9 and 10 on *potential* are robust across all specifications. However, the positive coefficient on *quantity* when controlling for *length* is no longer significant when the outliers on *time* are included (specifications (ii) and (iv)) or winsorized (specification (iii)).

More precisely, the coefficient on *quantity* drops from 13.072 with a t-stat of 2.56 in Table 10 to 7.373 with a t-stat of 1.28 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 robust to how outliers are treated.

Table OA.12: Correlation of cost-based complexity with ex-ante measures of complexity. This table reports the coefficients, t-statistics (in brackets), and adjusted R^2, of univariate regressions of *time* over the six ex-ante measures of complexity separately, with participant and question fixed effects. The sample is restricted to correct answers. Observations with $time_{i,t}$ above 579 are winsorized.

	(1)	(2)	(3)	(4)	(5)	(6)
length	3.713***					
	(14.81)					
cyclomatic		10.860***				
		(11.31)				
quantity			35.135***			
			(13.55)			
potential				8.507***		
				(15.19)		
diversity					41.307***	
					(11.94)	
level						-315.129***
						(-7.70)
R_a^2	0.447	0.378	0.422	0.454	0.390	0.313

Table OA.13: Correlation of cost-based complexity with ex-ante measures of complexity, controlling for *length*. This table reports the coefficients, t-statistics (in brackets), and adjusted R^2, of regressions of *correct* over *length* and each of the five other measures of complexity separately, with participant and question fixed effects. The sample is restricted to correct answers. Observations with $time_{i,t}$ above 579 are winsorized.

	(1)	(2)	(3)	(4)	(5)	(6)
length	3.713***	5.091***	2.952***	1.437**	3.585***	4.397***
	(14.81)	(9.16)	(5.47)	(2.11)	(7.89)	(12.44)
cyclomatic		-5.569***				
		(-2.78)				
quantity			8.692			
			(1.59)			
potential				5.497***		
				(3.59)		
diversity					2.018	
					(0.34)	
level						141.549***
						(2.73)
R^2_a	0.447	0.453	0.448	0.457	0.446	0.453

Table OA.14: Correlation of cost-based complexity with ex-ante measures of complexity. This table reports the coefficients, t-statistics (in brackets), and adjusted R^2, of univariate regressions of *time* over the six ex-ante measures of complexity separately, with participant and question fixed effects. The sample is restricted to correct answers. Observations with $time_{i,t} > 579$ are included.

	(1)	(2)	(3)	(4)	(5)	(6)
length	3.806***					
	(14.33)					
cyclomatic		11.229***				
		(11.10)				
quantity			35.685***			
			(12.98)			
potential				8.618***		
				(14.46)		
diversity					41.938***	
					(11.45)	
level						-331.373***
						(-7.71)
R^2_a	0.434	0.370	0.407	0.437	0.377	0.310

Table OA.15: Correlation of cost-based complexity with ex-ante measures of complexity, controlling for *length*. This table reports the coefficients, t-statistics (in brackets), and adjusted R^2, of regressions of *correct* over *length* and each of the five other measures of complexity separately, with participant and question fixed effects. The sample is restricted to correct answers. Observations with $time_{i,t} > 579$ are included.

	(1)	(2)	(3)	(4)	(5)	(6)
length	3.806***	5.100***	3.160***	1.791**	3.759***	4.427***
	(14.33)	(8.65)	(5.53)	(2.47)	(7.81)	(11.80)
cyclomatic		-5.229**				
		(-2.46)				
quantity			7.373			
			(1.28)			
potential				4.867***		
				(2.99)		
diversity					0.739	
					(0.12)	
level						128.356**
						(2.34)
R_a^2	0.434	0.439	0.435	0.441	0.433	0.438

Table OA.16: Correlation of cost-based complexity with ex-ante measures of complexity. This table reports the coefficients, t-statistics (in brackets), and adjusted R^2, of univariate regressions of *time* over the six ex-ante measures of complexity separately, with participant and question fixed effects. The sample is restricted to answers with $time \leq 579$. Incorrect answers are included.

	(1)	(2)	(3)	(4)	(5)	(6)
length	3.236***					
	(16.76)					
cyclomatic		8.915***				
		(12.48)				
quantity			31.785***			
			(15.79)			
potential				7.326***		
				(16.84)		
diversity					36.647***	
					(13.45)	
level						-276.547***
						(-9.20)
R_a^2	0.488	0.428	0.474	0.489	0.441	0.388

Table OA.17: Correlation of cost-based complexity with ex-ante measures of complexity, controlling for *length*. This table reports the coefficients, t-statistics (in brackets), and adjusted R^2, of regressions of *correct* over *length* and each of the five other measures of complexity separately, with participant and question fixed effects. The sample is restricted to answers with $time \leq 579$. Incorrect answers are included.

	(1)	(2)	(3)	(4)	(5)	(6)
length	3.236***	4.631***	2.238***	1.611***	3.131***	3.592***
	(16.76)	(11.08)	(5.85)	(3.23)	(9.15)	(13.57)
cyclomatic		-5.499***				
		(-3.76)				
quantity			11.868***			
			(3.01)			
potential				3.972***		
				(3.53)		
diversity					1.707	
					(0.37)	
level						74.121**
						(1.97)
R_a^2	0.488	0.495	0.492	0.494	0.487	0.489

Table OA.18: Correlation of cost-based complexity with ex-ante measures of complexity. This table reports the coefficients, t-statistics (in brackets), and adjusted R^2, of univariate regressions of *time* over the six ex-ante measures of complexity separately, with participant and question fixed effects. The sample includes all answers.

	(1)	(2)	(3)	(4)	(5)	(6)
length	3.652***					
	(16.45)					
cyclomatic		10.625***				
		(13.18)				
quantity			34.465***			
			(14.64)			
potential				8.050***		
				(15.87)		
diversity					39.086***	
					(12.32)	
level						-330.806***
						(-9.64)
R_a^2	0.468	0.421	0.442	0.459	0.409	0.376

Table OA.19: Correlation of cost-based complexity with ex-ante measures of complexity, controlling for *length*. This table reports the coefficients, t-statistics (in brackets), and adjusted R^2, of regressions of *correct* over *length* and each of the five other measures of complexity separately, with participant and question fixed effects. The sample is restricted to answers with $time \leq 579$. The sample includes all answers.

	(1)	(2)	(3)	(4)	(5)	(6)
length	3.652***	4.507***	3.049***	2.520***	3.929***	3.916***
	(16.45)	(9.26)	(6.95)	(4.38)	(10.14)	(12.78)
cyclomatic		-3.345**				
		(-1.97)				
quantity			7.232			
			(1.59)			
potential				2.779**		
				(2.13)		
diversity					-4.589	
					(-0.87)	
level						54.825
						(1.25)
R_a^2	0.468	0.469	0.468	0.470	0.467	0.468

OA.6 A dictionary for studying the complexity of regulatory texts

As discussed in Section 4.1, we have created a dictionary consisting on n-grams that appear in the text version of the Dodd-Frank Act. We have followed the following steps to create our dictionary:

1. We started by manually classifying n-grams using the dashboard discussed in Section 4.1, and reproduced in Fig. OA.6 below. This results in 6,115 unique entries and a marked-up version of the Dodd-Frank Act 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 as an n-gram.

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 Dodd-Frank Act 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 Dodd-Frank Act.

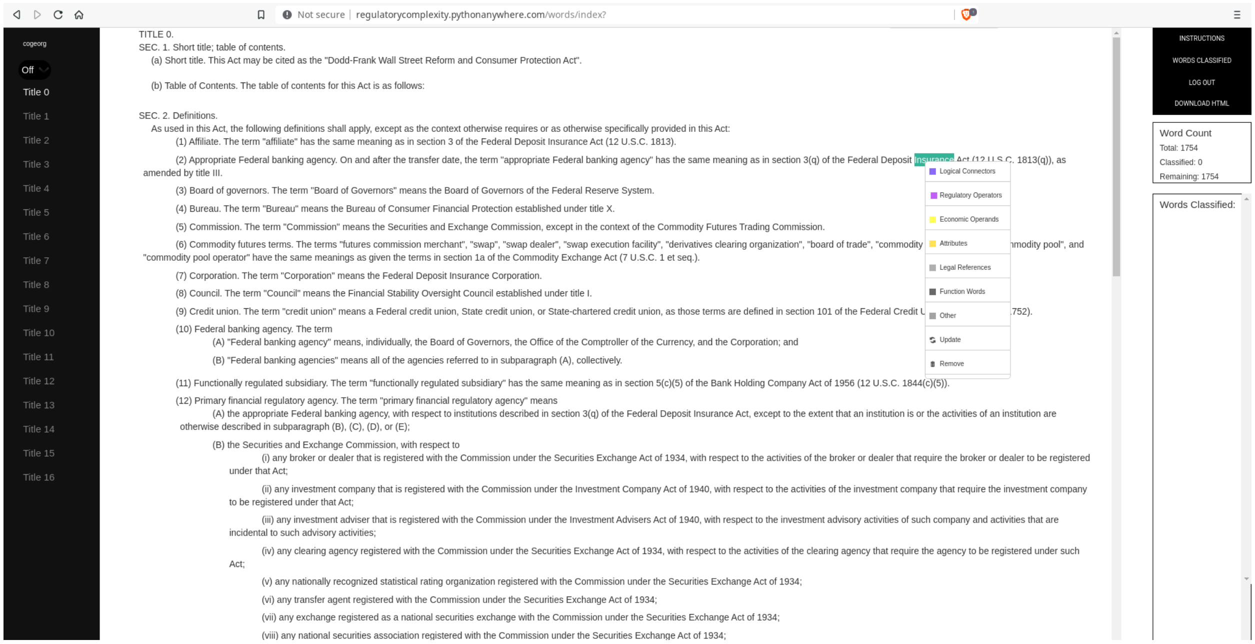


Figure OA.6: The Dashboard we developed to help us classify words in the Dodd-Frank Act as one of the following seven categories: Logical Connectors, Regulatory Operators, Economic Operands, Attributes, Legal References, Function Words, or Other. Top: The plain text of the Dodd-Frank Act. When highlighting 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 Dodd-Frank Act when all words and phrases have been classified.

OA.7 A model of risk-sensitivity

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 is a continuum $x \in [0, 1]$ of asset types. The bank starts with an asset of type x, drawn from the uniform distribution over $[0, 1]$. 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 $1 - x$, i.e., the bank makes a loss of x on its investment. If $E < x$ the bank defaults, and the government has to repay $D - (1 - x) = x - E$ to the depositors.

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 equity y and an asset type x, total welfare writes as:

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

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

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

As long as $y < x$, we have $\partial \mathcal{W}/\partial y = \lambda(1-p) - \delta$, which by assumption is positive. It is then clear that the optimal regulation would be to have $y^*(x) = x$ for any x, so that the bank never defaults. Total expected welfare would then be:

$$\int_0^1 \mathcal{W}(x,x)dx = \int_0^1 -\delta x dx = -\frac{\delta}{2}. \tag{OA.4}$$

Such a regulation requires to associate a continuum of different asset types to different levels of capital, which may be very complex, and hence costly.

We assume instead that the regulator defines different buckets, that is, intervals $[a_i, b_i]$ such that if $x \in [a_i, b_i]$ then $y \geq E_i$. For a given interval $[a,b]$ the optimal capital requirement $y^*_{a,b}$ is given by:

$$y^*_{a,b} = b - \delta \frac{b-a}{\lambda(1-p)}. \tag{OA.5}$$

Proof: For a given $y \in [a,b]$, total welfare is given by:

$$\begin{aligned} \mathcal{W}_{a,b}(y) &= \int_a^b [-\lambda(1-p)\min(x-y,0) - \delta y]dx && \text{(OA.6)} \\ &= -\lambda(1-p)\int_y^b (x-y)dx - \delta y(b-a) && \text{(OA.7)} \\ &= -\lambda(1-p)\frac{(b-y)^2}{2} - \delta y(b-a). && \text{(OA.8)} \end{aligned}$$

Maximizing this quantity with respect to y gives the desired result. ■

Note that we indeed have $a \leq y^*_{a,b} \leq b$. This means that banks with assets x close to a will be over-capitalized (they have more capital than what is necessary to sustain the losses 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:

$$\mathcal{W}_{a,b}(y^*_{a,b}) = \delta(b-a)\left[\frac{\delta(b-a)}{2\lambda(1-p)} - b\right]. \tag{OA.9}$$

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:

$$\begin{aligned} W_{0,\bar{x}}(y^*_{0,\bar{x}}) + W_{\bar{x},1}(y^*_{\bar{x},1}) &= \delta\bar{x}\left[\frac{\delta\bar{x}}{2\lambda(1-p)} - \bar{x}\right] + \delta(1-\bar{x})\left[\frac{\delta(1-\bar{x})}{2\lambda(1-p)} - 1\right] && \text{(OA.10)} \\ &= \delta\bar{x}(1-\bar{x})\frac{\lambda(1-p)-\delta}{\lambda(1-p)} - \frac{\delta}{2\lambda(1-p)}[\delta - 2\lambda(1-p)]. && \text{(OA.11)} \end{aligned}$$

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:

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

We use this lest expression to compute total welfare:

$$\mathcal{W}_0(I) = \sum_{i=0}^{I-1} W_{i/I,(i+1)/I}(y^*_{i/I,(i+1)/I}) = -\frac{\delta}{2} - \frac{\delta}{2I\lambda(1-p)}[\lambda(1-p) - \delta]. \tag{OA.14}$$

Total welfare is thus increasing in I, and converges to the continuous case $-\delta/2$ as $I \to +\infty$. Without any cost of complexity, it would be optimal to define as many risk buckets as possible. We use expression (OA.14) to plot $\mathcal{W}_0(I)$ on Fig. 3, with $\lambda = 0.05$, $\delta = 0.01$, and $p = 0.05$. These parameters are meant for illustration only.

OA.8 Legislative History of the Dodd-Frank Act

Summary:

- Introduced in the House of Representatives as *"The Wall Street Reform and Consumer Protection Act of 2009"* (H.R. 4173) by Barney Frank (D-MA) on December 2, 2009
- (Committee consideration by Financial Services)
- Passed the House on December 11, 2009 (223-202)
- Passed the Senate with amendment on May 20, 2010 (59-39)
- Reported by the joint conference committee on June 29, 2010; agreed to by the House on June 30, 2010 (237-192) and by the Senate on July 15, 2010 (60-39)
- Signed into law by President Barack Obama on July 21, 2010