Measuring Regulatory Complexity *

Jean-Edouard Colliard†
HEC Paris

Co-Pierre Georg‡
University of Cape Town
Deutsche Bundesbank

January 26, 2022

Abstract

Despite a heated debate on the perceived increasing 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 complexity measures, develop new ones, compute them on two examples—Basel I and the Dodd-Frank Act—and validate them using novel experiments that involve the computation of risk-weighted assets under various rules. Our framework offers a quantitative approach to the policy trade-off between the precision and the complexity of regulation. The toolkit we develop is freely available and allows researchers to measure the complexity of any normative text as well as test alternative measures of complexity.

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

JEL classification: G18, G28, G41.

*We are grateful to Bruno Biais, Claire Celerier, Igor Kozhanov, Iman van Lelyveld, Christoph Merkle, Sebastien Pouget, as well as to participants to the Paris Webinar on Financial Technology and Crypto, 2019 BCBS-CEPR Workshop on "Impact of regulation in a changing world: innovations and new risks", the 2019 NY Fed Fintech Conference, the 2016 "Journée de la Chaire ACPR", seminar audiences at Aix Marseille School of Economics, Autorité des Marchés Financiers, the Bank of England, HEC Paris, De Nederlandsche Bank, the Office of the Comptroller of the Currency, the University of Bonn, and the University of Exeter, the 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. The views expressed in this paper do not necessarily reflect the views of the Deutsche Bundesbank or the Eurosystem.

†E-Mail: colliard@hec.fr
‡E-Mail: co-pierre.georg@uct.ac.za

Electronic copy available at: https://ssrn.com/abstract=3523824

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 itself is aware of the issue, and considers simplicity as a desirable objective, to be traded off against 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), and 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 three 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 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.

[1]See Gai *et al.* (2019) provide 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`.

Electronic copy available at: https://ssrn.com/abstract=3523824

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 then turn to precise measures of complexity, for which we need a stylized representation of a regulation. We use the representation 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 very 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 by definition do 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

Electronic copy available at: https://ssrn.com/abstract=3523824

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. Participants to an experiment are given a regulation 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 forecast 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 improves the forecast relative to the mere length of the regulation. We propose four criteria a suitable measure of complexity should satisfy: the measure is negatively correlated with the number of correct answers, the measure is positively correlated with the time taken to answer, and either correlation still obtains after controlling for the length of the regulation. Only one of the five measures we consider satisfies these four criteria, suggesting that our experimental design is a powerful touchstone to test the validity of new measures. Importantly, 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 make this dictionary available online, so that interested researchers can compute our measures on other regulatory texts. 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.

Finally, we show how building on our approach could eventually lead to a quantitative

Electronic copy available at: https://ssrn.com/abstract=3523824

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 claims made by the industry so as not to reveal their lack of sophistication. Oehmke and Zawadowski (2019) develop a model which, applied to regulation, would assume that regulatory complexity is in itself desirable (e.g., it allows for more risk-sensitivity), but that regulators neglect that a more complex regulation consumes the limited attention of agents, and crowds out other activities. A striking prediction from the model is that observing that regulations are well understood by regulators and market participants is a sign that their complexity is suboptimally high. 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

[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)).

Electronic copy available at: https://ssrn.com/abstract=3523824

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. Applied to post-crisis regulatory reforms, the model suggests that the increased demand for regulation led to more complex and lower quality regulations, while a potentially better alignment between regulators and politicians led to more complex but higher quality regulations. Foarta and Morelli (2021) 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) and 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 complex assets.

Further from finance applications but also related to our study is a literature that tries to measure the complexity of solving mathematical problems for humans. In particular, the experimental approach we use in Section 3 is related to Murawski and Bossaerts (2016) and Franco *et al.* (2021). They ask participants to solve different versions of the knapsack problem, and study how the participants' 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

[5] Some empirical papers study the increase in the stringency or quantity of regulations. For instance, Kalmenovitz (2021) 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.

Electronic copy available at: https://ssrn.com/abstract=3523824

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. Recently, Oprea (2020) used an experimental approach to measure the cognitive costs of following different procedures (“implementation complexity"), and showed 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.

1 A unifying framework

1.1 Dimensions of complexity

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". In this section, we provide preliminary definitions to clarify different dimensions of complexity, notably those we are measuring in this paper.

We start by formalizing the analogy between regulations and algorithms. Knuth (1973) describes an algorithm as:[7] *“a finite set of rules that gives a sequence of operations for solving a specific type of problem.”* Surprisingly, a formal definition of an algorithm beyond the informal characterization provided above is not without difficulty, but for the purpose of our paper, this somewhat informal description of an algorithm is sufficient.

In the case of regulation, the “input" is a regulated entity and the output a regulatory ac-

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

[7] Knuth (1973) 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 programming languages. 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 paper.”*

Electronic copy available at: https://ssrn.com/abstract=3523824

tion. A regulated entity could be for example an individual financial institution, or the entire financial system. Examples of regulatory actions are imposing a fine on a bank, imposing higher capital requirements, or simply allowing the bank to continue operating. Formally, we define:

Definition 1. *A regulatory problem is a mapping $f : \mathscr{E} \to \Sigma$ from the set of regulated entities $\mathscr{E}$ to a set of regulatory actions Σ.*

An algorithm is a set of rules, such that by following them we can compute $f(x)$ given any input x. Similarly, a regulation is a sequence of instructions that implement an appropriate regulatory action to any regulated entity:

Definition 2. *A regulation $\tilde{f}$ is a sequence of elements taken in a vocabulary $\mathscr{V}$. This sequence of elements is interpreted through a language, and implements f.*

It is important to note that the same problem can be solved by different algorithms. Similarly, the same regulatory problem f can be solved by different regulations $\tilde{f}$.

Finally, once a particular algorithm to solve a problem has been chosen, the last step is to actually run the algorithm, which may take more or less time and computing power. Similarly, following the rules set in a given regulation may be more or less complicated for the regulatory authority and/or for the regulated entity. We call this last step “supervision":

Definition 3. *Supervision is the act of following $\tilde{f}$ to evaluate $f(e)$ for a given entity $e \in \mathscr{E}$.*

We can now define properties of measures of regulatory complexity corresponding to different dimensions of complexity. Assume we have a set $\tilde{\mathscr{F}} = \{\tilde{f}_1, \tilde{f}_2 ..., \tilde{f}_n\}$ of regulations solving the same regulatory problem f, and a set $\mathscr{E} = \{e_1, e_2, ..., e_m\}$ of regulated entities. Elements of these sets could be empirically observed (actual regulatory texts, actual banks) or hypothetical (variants on the text, hypothetical banks). Following the previous definitions, we can define a measure of regulatory complexity and give necessary conditions for different types of measures as follows:

Definition 4. *A measure of regulatory complexity μ is a mapping $\mu : \tilde{\mathscr{F}} \times \mathscr{E} \to \mathbb{R}$. If μ is a measure of problem complexity, then $\mu(\tilde{f}, e)$ is constant in $\tilde{f}$ and e. If μ is a measure of psy-*

Electronic copy available at: https://ssrn.com/abstract=3523824

chological complexity, then $\mu(\tilde{f}, e)$ *is constant in e but not necessarily in* $\tilde{f}$. *If* μ *is a measure of computational complexity, then* $\mu(\tilde{f}, e)$ *may depend both on e and* $\tilde{f}$.

These properties characterize an important distinction between three forms of regulatory complexity:

(i) 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 f, but is independent of which regulation $\tilde{f}$ implements f.

(ii) Regulatory complexity may also mean that the actual regulation used to solve the regulatory problem is complex, which may be due both to the complexity of the problem f and to the complexity of the particular $\tilde{f}$ that solves the problem. 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.

(iii) Finally, regulatory complexity may mean that applying a regulation to a particular entity is costly in terms of time and resources. The cost can be incurred by the supervisor (supervision costs) and by the regulated entities (compliance costs). 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.

An example: The length of bank capital regulation. In the example of capital regulation, a regulated entity is a bank, represented for instance by a list B of balance sheet items and values. The regulatory problem is to associate any possible bank balance sheet B to an action, the simplest ones being for instance "pass" or "fail". Regulation is then a series of operations on balance sheet items that ends with an outcome $\sigma \in \Sigma$. Haldane and Madouros (2012) for instance measure the complexity of banking regulation by the number of pages of the different Basel Accords. In our framework, the exact text of the Basel Accords is a particular regulation $\tilde{f}$ to solve an underlying regulatory problem. The length of the text is a

Electronic copy available at: https://ssrn.com/abstract=3523824

particular measure. Clearly, this measure depends on how the text is written, but not on which bank we apply the regulation to. It is thus a measure of psychological complexity, but not of problem complexity or computational complexity.

1.2 Extending the Halstead framework

In this section we show how several measures of regulatory complexity can be derived by modeling a regulation like an algorithm in Halstead (1977). We consider regulation $\tilde{f}$ as an sequence of "n-grams" (expressions of length n that are elements in a language) $\tilde{f} = \{w_1, w_2 ... w_N\}$, 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 $\tilde{f}$, 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)
```
(1)

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 (1)) 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, substraction, and so on. We denote $N_R, \eta_R, N_L, \eta_L, N_M, \eta_M$ the number of total regula-

Electronic copy available at: https://ssrn.com/abstract=3523824

tory operators, unique regulatory operators, total logical operators, unique logical operators, total mathematical operators, and unique mathematical operators, respectively. We have $N_R + N_L + N_M = N_{OR}$ and $\eta_R + \eta_L + \eta_M = \eta_{OR}$.

1.3 Measures of complexity

We now derive six measures of complexity within our extended framework. We also briefly survey other measures that have been proposed but do not readily fit in our 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.[8] 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 representation $\tilde{f}$ of regulation f. How can one obtain a measure of problem complexity, that depends only on f? 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 is trivial. For example, the shortest possible program to compute the vector norm is:

```
y = vecnorm(x_1,x_2...x_n)
```
(2)

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

Electronic copy available at: https://ssrn.com/abstract=3523824

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 $potential$.

Similarly, 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.

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.

For completeness, we briefly review other types of measures that have been proposed in the literature but do not directly fit within our framework and probably capture different dimension of complexity.

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 ap-

Electronic copy available at: https://ssrn.com/abstract=3523824

ply 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.[9] These network-based measures of complexity are quite different from our approach because they are based on references between different legal texts in a corpus.

Finally, some papers follow a very different approach by estimating how much effort regulated entities spend on understanding or complying with regulations. They thus propose measures of 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 (2021) 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.

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 1.1. 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.]

[9] 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.

Electronic copy available at: https://ssrn.com/abstract=3523824

2 Basel I

We now apply our measure empirically to an actual text, 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.

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. This is done by mapping 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 Appendix C. In this section, we briefly explain the structure of the program and give the associated 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`,

Electronic copy available at: https://ssrn.com/abstract=3523824

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. We can then compute our six measures of regulatory complexity for each item, as well as for the entire regulation. We report these measures in Table 2. In addition, Table 3 gives the pair-wise correlation coefficients between the different measures, across the 19 regulatory instructions. We report both the Pearson and Spearman correlation coefficients.

[Insert Tables 2 and 3 here.]

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.

2.2 Text analysis

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. Our objective is to test whether our approach can meaningfully be applied to a text directly, and hence more easily and on a wider scale than if one first has to translate a regulation into an algorithm.

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.[10] 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

[10]We report this modified text in Appendix C.

Electronic copy available at: https://ssrn.com/abstract=3523824

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”, etc.), hence we don’t take them into account. Table 4 gives the top 5 words in each category. We then reproduce Tables 2 and 3 using the measures based on our text analysis.

[Insert Tables 4, 5, and 6 here.]

We observe that the text-based measures tend to be less correlated with each other than the algorithm-based measures. The algorithm always uses the same logical structure to define the risk-weight in a particular rule. In contrast, the text version is sometimes ambiguous or leaves some elements implicit, it is then up to the reader to interpret the text. A good example is item (2a), which has $length = 43$ in the algorithmic version but $length = 22$ only in the text version. Interestingly, in both versions this item stands out as one of the most complex according to $cyclomatic$, $complexity$, and $potential$, but the compact way it is drafted gives it a low value of $length$ in the text version. Accordingly, this is also one of the rules with the highest $level$ in the text version.

Finally, we compute the correlations between the text-based measures and their algorithm-based counterparts. Table 7 gives the correlation coefficients.[11] The correlation coefficients we obtain are quite high, with the exception of $diversity$, which shows that the text-based analysis and the algorithm-based analysis are capturing similar patterns. Moreover, 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

[11] Formally the coefficients are not defined for $quantity$. However, 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.

Electronic copy available at: https://ssrn.com/abstract=3523824

potential are indeed capturing problem complexity, as they are less affected by a change in the language used.

[Insert Table 7 here.]

3 Experiments

3.1 Validation of the measures

While the different measures we have studied are intuitive, they 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 subjects to evaluate a regulatory action by computing regulatory quantities based on different regulations. We then measure to what extent the correctness of their output and how quickly subjects have computed it is correlated with different measures of regulatory complexity.

Generating a sample of test regulations. For our experiment 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 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. More precisely, for our random regulation to have the same structure as the Basel I regulation text (see Section 2.1), we de-

Electronic copy available at: https://ssrn.com/abstract=3523824

cide upfront on the number of IF-THEN-ELSE clauses we want to have. As with the actual Basel I regulation, we use 6 clauses in total. Within each clause, the algorithm then selects a number of random conditions (smaller or equal than some fixed positive bound, in our case 10). Each condition consists of operators and operands, e.g. `ASSET_CLASS == "cash"` that can be combined by AND and OR statements. We use only operands and operators that also exist in the Basel I regulation. Operands in our random regulation generator can take exactly the values they can take in the original Basel I text. For example, ASSET_CLASS can take the values {cash, claim, loan, premises, plant, equipment, real_estate, other_fixed_assets, other_investments, capital_instruments}. Different assets can have attributes, e.g., a claim can have (among other attributes) a ISSUER and a DENOMINATION. Fig. 9 shows a flowchart of the algorithm we use.

[Insert Fig. 9 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.[12] At the end of this process, we obtain 38 regulations that we use in our experiment. As shown in Table 8 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 9. 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 6).

[Insert Tables 8 and 9 here.]

Participants. 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

[12]The Appendix D shows an example of such a randomly generated regulation. In addition, the replication files of this paper give: (i) the program used to generate the random regulations; (ii) the raw regulations generated by the program; (iii) the final regulations we used in the experiment.

Electronic copy available at: https://ssrn.com/abstract=3523824

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, and 67.9% of their answers were correct.

Experiment. 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/`). After an introductory page (Figure 1), the participant registers and gives some background information (Figure 2). The participant is then shown a screen with explanations about the experiment and how to compute capital requirements (Figure 3). The next screen is a “test-round", which is the same for all participants (Figure 4). 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 (Figures 5 and 6), 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

Electronic copy available at: https://ssrn.com/abstract=3523824

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 students.

[Insert Fig. 1 to 6 here.]

Results. We want to test the ability of our complexity measures to forecast the difficulty of computing risk-weighted assets depending on the regulation considered. A natural way to proceed is to create a measure of "difficulty" at the regulation level and see how it correlates with measures of complexity. Thus, for each randomly generated regulation $j \in 1,2,...38$ we create the variable $correct_j$, the percentage of correct answers for regulation j, and $time_j$, the average time taken to solve regulation j. In both cases we exclude from the analysis 7 students who took the test several times, and whose answers are potentially affected by a learning effect. To compute $time_j$, we exclude incorrect answers and answers with time below 7 seconds (1st percentile) or above 579 (99th percentile). We thus drop 342 of our 1062 observations.

We then run simple OLS regressions of $correct_j$ and $time_j$ on our measures of complexity, either on each measure individually, on each measure and length, on all measures at the same time except level (which is the ratio of two other measures).[13] We propose the following four criteria to judge the quality of a complexity measure X other than length:

(i) X is negatively correlated with the percentage of correct answers: When regressing $correct_j$ over X_j only, the coefficient on X_j is negative and significant at the 10% level.

(i') X explains the percentage of correct answers beyond length: When regressing $correct_j$ over $length_j$ and X_j, the coefficient on X_j is negative and significant at the 10% level.

(ii) X is positively correlated with the time taken to give a good answer: When regressing $time_j$ over X_j only, the coefficient on X_j is positive and significant at the 10% level.

(ii') X explains the time taken to give a good answer beyond length: When regressing $time_j$ over $length_j$ and X_j, the coefficient on X_j is positive and significant at the 10% level.

Ideally, a good measure of complexity should be negatively correlated with the probability of finding a correct answer and positively with the time taken to provide a good answer.

[13] In Appendix G, we also run panel regressions at the student-question level to further investigate the determinants of providing a correct answer and the time taken to provide it.

Electronic copy available at: https://ssrn.com/abstract=3523824

Moreover, given that *length* is easy to compute and widely used, it is natural to ask that a new measure captures something which is not already captured by *length*. Tables 10, 11, 12, and 13 show the results of the different regressions used to test our four criteria. In addition, Table 14 summarizes which criteria are satisfied for each measure.

[Insert Tables 10 to 14 here.]

Our four criteria allow us to significantly discriminate among the different measures. The measure *quantity* stands out, as it is the only one to pass the four criteria. The measure *potential* only fails criterion (i'). The measure *cyclomatic* is significantly correlated with *correct* and *time* with the expected signs, but not once we control for *length*, it thus fails criteria (i') and (ii'). *diversity* and *level* fail all but one criterion.

These results have natural interpretations. *cyclomatic* is a measure of psychological complexity, like *length*, which may be why it does not capture anything 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*. *diversity* was introduced by symmetry with *potential*, but it does not rely on any theoretical foundation, and accordingly it performs poorly. *level* is a special case, as it is closely related to the ratio of *potential* over *length*. It seems that the impact of *length* is stronger than the impact of *potential*, so that in univariate regressions *level* has the wrong sign. For criterion (ii') the sign of level is correct once we control for *length*, but in that case *level* essentially brings the same information as *potential*.

It is also instructive to look at the R^2 of the different regressions we conducted. *length* alone explains 27% of the variation of *correct* across the 38 regulations. If we add *quantity* the R^2 jumps to 42%, confirming that *quantity* brings information not already included in *length*. *potential* instead does not seem to bring additional information relative to *length*. If one looks at *time* instead, the roles of *potential* and *quantity* are flipped: *length* alone captures 58% of the variation of *time*, adding *potential* increases this number to 67%, but adding quantity brings it to 62% only. While *potential* and *quantity* are both measures of problem complexity, it seems they are capturing different subdimensions,

Electronic copy available at: https://ssrn.com/abstract=3523824

one seems more associated with making mistakes and the other one with taking a longer time to answer.

While we believe these results are interesting in their own right, our main conclusion is broader: our 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 one passed all four criteria, and a second one passed three of them. 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 measure of complexity and compare it to the five we considered. The only restriction is that the measure has to be text-based.

4 Applications

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

In order to apply the Halstead approach at scale on a variety of actual regulatory texts, one needs a dictionary of regulatory terms, with a classification of words into operators and operands. To start building such a dictionary and prove that our measures can be implemented on 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 hope to create 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

Electronic copy available at: https://ssrn.com/abstract=3523824

manually made a list of all such n-grams (for details see Appendix E). 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.[14]

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 require us to classify the words manually. We created an online dashboard to help with this task, shown in Fig. 7. The main function of this dashboard is to highlight words that have already been classified, so that we never have to classify the same word twice. After classifying words in the 16 Titles of the Dodd-Frank Act plus its introduction, we create a dictionary containing: 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 15 shows the top 10

[14] 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.

Electronic copy available at: https://ssrn.com/abstract=3523824

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 16.

[Insert Tables 15 and 16, and Fig. 7 here.]

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 17, 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, potential, and diversity) seem the easiest to compute on other texts without having to expand the dictionary.

[Insert Table 17 here.]

We made the dictionary of all the classified words in the Dodd-Rank Act available online. In addition, the code for the dashboard we used is available, and can be used to enrich our dictionary with words from other regulatory texts. We hope that through this collaborative tool other studies of regulatory complexity will be conducted, so that for instance the complexity of different types of regulation or regulations in different countries can be compared.

4.2 Towards a normative analysis: “balancing risk-Sensitivity and simplicity"

The length and perceived complexity of the Basel II and Basel III Accords have led to an intense debate in policy circles about the complexity of banking regulation. This led in particular the Basel Committee to publish a discussion paper on the trade-offs between “risk

Electronic copy available at: https://ssrn.com/abstract=3523824

sensitivity, simplicity and comparability" (Basel Committee on Banking Supervision, 2013). 8 years later, the right trade-off remains elusive, in particular due to the lack of a normative framework to think about regulatory complexity. In this section, we sketch how our framework could (after suitable extensions) eventually serve such a normative purpose and be used to think about the optimal level of complexity.

As a starting point, one needs a model of the economy or the financial system that can accommodate different regulations with different levels of complexity. In addition, this model ought to be quantitative. This is precisely where the literature on bank capital requirements is heading.[15] As a simple illustration, assume a bank invests in assets that have a certain "asset class" $x \in [0,1]$. Consider then the following family of bank capital regulations:

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

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

`...`

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

`else` $E \equiv E_I^*$

where E is the amount of equity the bank is required to have for an asset belonging to class x, 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 distinct asset classes considered by the regulation.

In Appendix F, we propose for illustration a simple model of bank risk-shifting in which such a capital regulation improves welfare. Moreover, the optimal welfare the regulator can achieve for a given number of asset classes I, denoted $\mathscr{W}(I)$, 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, E, the $\bar{x}_i$, and the E_i^*. We have $\eta_R = N_R = 1$, $\eta_L = 3$ 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

[15] 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.

Electronic copy available at: https://ssrn.com/abstract=3523824

$length(I)$, $cyclomatic(I)$, $quantity(I)$, $potential(I)$, $diversity(I)$, and $level(I)$ using the formulas in Table 1 and see how they vary with the number of asset classes I.

We can then use experiments such as those in Section 3 to translate the measures into a predicted cost of complexity, measured either as the frequency $\hat{p}(I)$ with which the regulation is misunderstood or as the average time $\hat{t}(I)$ taken to apply the regulation. More specifically, we use the estimates from Tables 11 and 13 to compute, for every I:

$$\begin{aligned} \hat{p}(I) &= 0.951 + 0.001 length(I) + 0.000 cyclomatic(I) \\ &\quad -0.106 quantity(I) - 0.013 potential(I) + 0.101 diversity(I), \end{aligned} \tag{3}$$

$$\begin{aligned} \hat{t}(I) &= 9.587 + 1.493 length(I) - 3.750 cyclomatic(I) \\ &\quad +10.431 quantity(I) + 6.581 potential(I) - 13.109 diversity(I). \end{aligned} \tag{4}$$

We can now quantitatively measure how increasing the number of distinct asset classes affects both social welfare and measures of the cost of complexity. Figure 8 displays the results for the particular welfare function derived in Appendix F. The last step would be for the policymaker designing the regulation to formulate explicit preferences over social welfare and complexity. With such preferences, the policymaker would be able to compute the optimal number of distinct asset classes and hence arrive at the optimal trade-off between "risk-sensitivity" (which improves welfare in this example) and "simplicity".

[Insert Fig. 8 here.]

The method we outline here is only a proof of concept, but to our knowledge this is the first proposal offering policymakers a scientific and 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 three additional tasks: (i) develop quantitative models of regulation, rich enough to estimate the welfare impact of different regulatory alternatives; (ii) formulate an explicit trade-off between welfare and regulatory simplicity; (iii) run richer and more robust experiments to have a more precise view of the costs of psychological complexity for different audiences.[16]

[16] More specifically, the type of experiment we consider in Section 3 should ideally be reproduced with regula-

Electronic copy available at: https://ssrn.com/abstract=3523824

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, to derive six measures of regulatory complexity that can be applied to large scale regulatory texts, to conduct a validation test that can be applied to any text-based measure, and to 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, the dictionary that we created will allow other interested researchers to compute various complexity measures for other regulatory texts and compare them to those we produced for Basel I and the Dodd-Frank Act. Moreover, the dictionary can be enriched in a collaborative way. Such a process would make the measures more robust over time and allow to 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 what is the impact of regulatory complexity, and in particular testing some of the mechanisms that have been proposed in the theoretical literature.

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.

tions actually under discussion, and with participants closer to the actual audience of regulatory texts (bankers, lawyers, regulators, etc.).

Electronic copy available at: https://ssrn.com/abstract=3523824

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. 10, 30

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. 11, 12, 30

ARORA, S., BARAK, B., BRUNNERMEIER, M. and GE, R. (2009). *Computational Complexity and Information Asymmetry in Financial Products*. Working paper. 5

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**. 4

BASEL COMMITTEE ON BANKING SUPERVISION (1988). *International convergence of capital measurement and capital standards*. Discussion paper, BCBS. 13

BASEL COMMITTEE ON BANKING SUPERVISION (2013). *The regulatory framework: balancing risk sensitivity, simplicity and comparability*. Discussion paper. 1, 4, 24

BEGENAU, J. and LANDVOIGT, T. (2021). Financial regulation in a quantitative model of the modern banking system. *Review of Economic Studies*, **Forthcoming**. 24

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**. 12, 30

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

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. 5

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

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). 4

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. 5

FOARTA, D. and MORELLI, M. (2021). *Complexity and the Reform Process*. Working paper. 5

Electronic copy available at: https://ssrn.com/abstract=3523824

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. 5

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. 5

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. 5

HAKENES, H. and SCHNABEL, I. (2012). *Regulatory Capture By Sophistication*. Working paper. 4

HALDANE, A. and MADOUROS, V. (2012). The dog and the frisbee. *Proceedings - Economic Policy Symposium - Jackson Hole*, pp. 109–159. 1, 8, 10, 30

HALSTEAD, M. H. (1977). *Elements of Software Science*. Elsevier. 2, 9, 10, 11, 30

HERRING, R. J. (2018). The evolving complexity of capital regulation. *Journal of Financial Services Research*, **53** (2), 183–205. 10

KALMENOVITZ, J. (2021). *Regulatory Intensity and Firm-Specific Exposure*. Working paper. 5, 12, 30

KATZ, D. and BOMMARITO, M. (2014). Measuring the complexity of the law: the united states code. *Artificial Intelligence and Law*, **22**. 11, 12, 30

KENDALL, C. and OPREA, R. (2021). *On the Complexity of Forming Mental Models*. Working paper. 6

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. 11, 30

KNUTH, D. E. (1973). *Fundamental Algorithms: The Art of Computer Programming*, vol. 1. Addison-Wesley, 2nd edn. 6

LI, W., AZAR, P., LAROCHELLE, D., HILL, P. and LO, A. W. (2015). Law is code: A software engineering approach to analyzing the united states code. *Journal of Business and Technology Law*, **10** (2). 1, 10, 11, 12

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

MAAS, H.-D. (1972). Über den Zusammenhang zwischen Wortschatzumfang und Länge eines Textes. *Zeitschrift für Literaturwissenschaft und Linguistik*, **2** (8). 11, 30

MCCABE, T. J. (1976). A complexity measure. *IEEE Transactions on Software Engineering*, **SE-2** (4), 308–320. 10, 30

Electronic copy available at: https://ssrn.com/abstract=3523824

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. 10

Murawski, C. and Bossaerts, P. (2016). How humans solve complex problems: The case of the knapsack problem. *Scientific Reports*, **6**. 5

Oehmke, M. and Zawadowski, A. (2019). *The Tragedy of Complexity*. Working paper. 4

Oprea, R. (2020). What makes a rule complex? *American Economic Review*, **110** (12), 3913–51. 6

Rubinstein, A. (1986). Finite automata play the repeated prisoner's dilemma. *Journal of Economic Theory*, **39** (1), 83 – 96. 6

Simkovic, M. and Zhang, M. B. (2020). *Regulation and Technology-Driven Entry: Measurement and Micro-Evidence*. Working paper. 12, 30

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. 4

Zuse, H. (1990). *Software Complexity: Measures and Methods*. New York, NY, United States: de Gruyter, 1st edn. 4, 8

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

Electronic copy available at: https://ssrn.com/abstract=3523824

A Tables

Table 1: Summary of Complexity Measures.

Name	Source	Formula	Complexity Dimension
Length	e.g., Haldane and Madouros (2012)	N	Psychological
Cyclomatic complexity	McCabe (1976)	N_L	Psychological
Quantity of regulations	Al-Ubaydli and McLaughlin (2017)	N_R	Problem
Potential volume	This paper and Halstead (1977)	$2+\eta_{OD}$	Problem
Operator diversity	This paper	η_{OR}	Psychological
Level	This paper and Halstead (1977)	$\frac{2+\eta_{OD}}{N}$	Psychological
Average word length	e.g., Amadxarif *et al.* (2019)	-	Psychological
Lexical diversity	Maas (1972)	-	Psychological
Readability metric	(Kincaid *et al.*, 1975)	-	Psychological
Shannon's entropy	e.g., Katz and Bommarito (2014)	-	Psychological
PageRank	Amadxarif *et al.* (2019)	-	Psychological
Network Centralities (Eigenvector, Betweenness, Closeness)	Boulet *et al.* (2011)	-	Psychological
Regulation Index	Simkovic and Zhang (2020)	-	Computational
RegIn	Kalmenovitz (2021)	-	Computational

Table 2: 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

Electronic copy available at: https://ssrn.com/abstract=3523824

Table 3: 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 4: 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		

Electronic copy available at: https://ssrn.com/abstract=3523824

Table 5: 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 6: 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

Electronic copy available at: https://ssrn.com/abstract=3523824

Table 7: Correlation coefficients between the measures based on the algorithm and the measures based on the text.

	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 8: 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 9: 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

Electronic copy available at: https://ssrn.com/abstract=3523824

Table 10: Criterion (i): Correlation with *correct*. This table reports the coefficients, t-statistics (in brackets), and R^2, of univariate regressions of *correct* over the six 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^2	0.268	0.158	0.414	0.254	0.063	0.177

Table 11: Criterion (i'): Correlation with *correct* beyond *length*. This table reports the coefficients, t-statistics (in brackets), and R^2, of regressions of *correct* over *length* and each of the five other measures of complexity separately (columns (1) to (5)). In addition, column (6) reports the results of a regression of *correct* over all measures except level.

	(1)	(2)	(3)	(4)	(5)	(6)
length	-0.011**	0.002	-0.005	-0.012***	-0.006**	0.001
	(-2.54)	(0.63)	(-0.93)	(-3.88)	(-2.17)	(0.07)
cyclomatic	0.015					-0.000
	(1.01)					(-0.00)
quantity		-0.111***				-0.106**
		(-3.04)				(-2.70)
potential			-0.006			-0.013
			(-0.48)			(-0.97)
diversity				0.083*		0.101**
				(2.03)		(2.51)
level					0.225	
					(0.58)	
R^2	0.289	0.421	0.273	0.345	0.275	0.518

Electronic copy available at: https://ssrn.com/abstract=3523824

Table 12: Criterion (ii): Correlation with $time$. This table reports the coefficients, t-statistics (in brackets), and R^2, of univariate regressions of $time$ over the six measures of complexity separately.

	(1)	(2)	(3)	(4)	(5)	(6)
length	3.256***					
	(7.09)					
cyclomatic		8.259***				
		(4.01)				
quantity			33.870***			
			(6.96)			
potential				7.954***		
				(8.46)		
diversity					34.565***	
					(4.65)	
level						-227.829**
						(-2.40)
R^2	0.582	0.309	0.574	0.665	0.375	0.138

Table 13: Criterion (ii'): Correlation with time beyond $length$. This table reports the coefficients, t-statistics (in brackets), and R^2, of regressions of time over $length$ and each of the five other measures of complexity separately (columns (1) to (5)). In addition, column (6) reports the results of a regression of time over all measures except level.

	(1)	(2)	(3)	(4)	(5)	(6)
length	5.615***	1.831**	0.287	3.269***	4.181***	1.493
	(6.01)	(2.05)	(0.26)	(4.17)	(6.85)	(0.56)
cyclomatic	-9.197***					-3.750
	(-2.83)					(-0.83)
quantity		17.229*				10.431
		(1.84)				(1.09)
potential			7.349***			6.581*
			(2.96)			(2.03)
diversity				-0.210		-13.109
				(-0.02)		(-1.33)
level					190.396**	
					(2.17)	
R^2	0.660	0.619	0.666	0.582	0.632	0.712

Electronic copy available at: https://ssrn.com/abstract=3523824

Table 14: Summary of the criteria for each measure. This table reports for each complexity measure other than length whether it passed each of the criteria (i), (i'), (ii), and (ii'). A ✓ symbol indicates that the measure satisfies the criterion, and a ✗ symbol that it does not.

	(i)	(i')	(ii)	(ii')
cyclomatic	✓	✗	✓	✗
quantity	✓	✓	✓	✓
potential	✓	✗	✓	✓
diversity	✗	✗	✓	✗
level	✗	✗	✗	✓

Table 15: 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

Electronic copy available at: https://ssrn.com/abstract=3523824

Table 16: 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 17: 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

Electronic copy available at: https://ssrn.com/abstract=3523824

B Figures

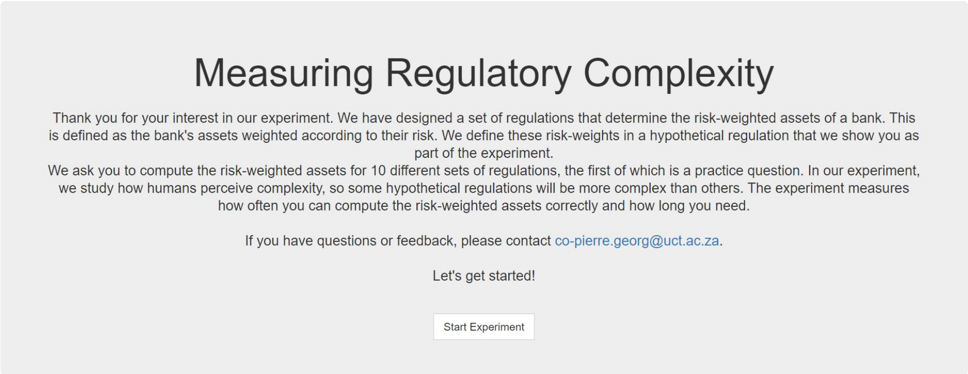

Figure 1: Online experiment - Welcome page.

Electronic copy available at: https://ssrn.com/abstract=3523824

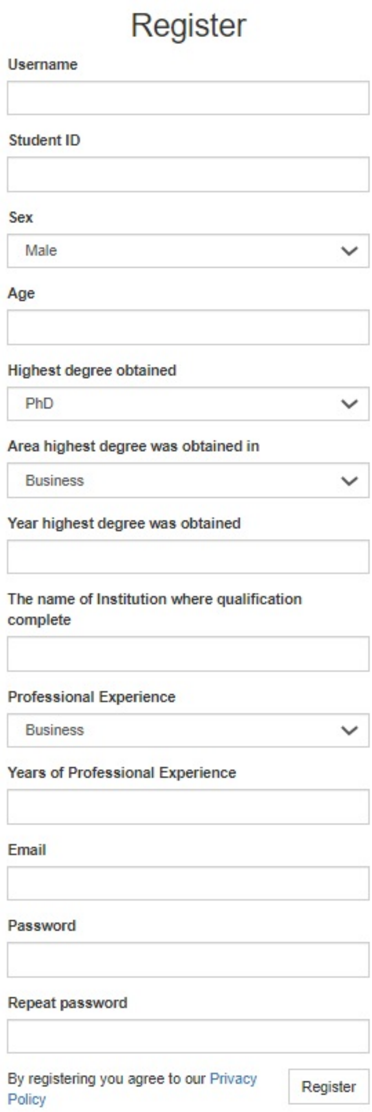

Figure 2: Online experiment - Registration page.

Electronic copy available at: https://ssrn.com/abstract=3523824

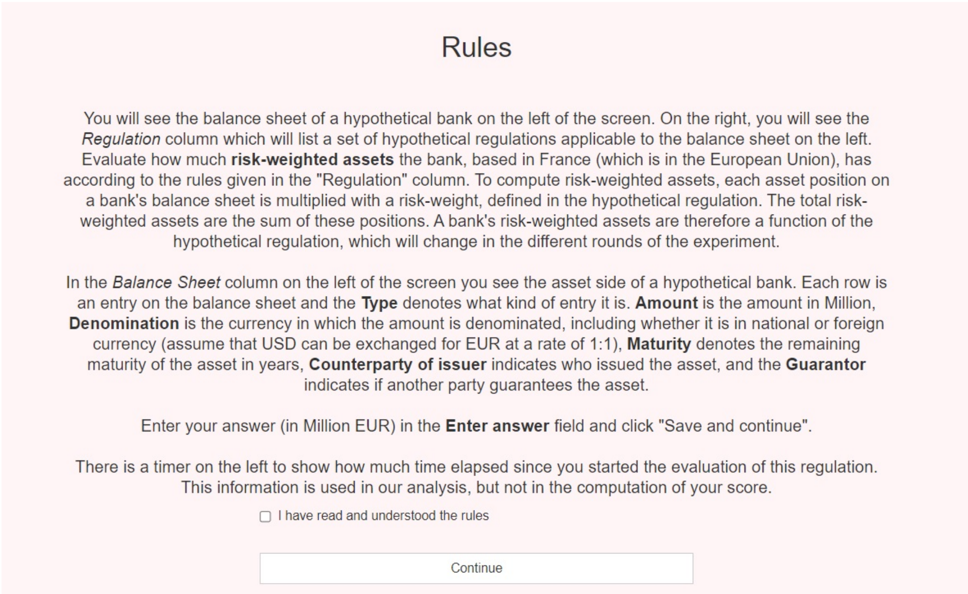

Figure 3: Online experiment - Instructions page.

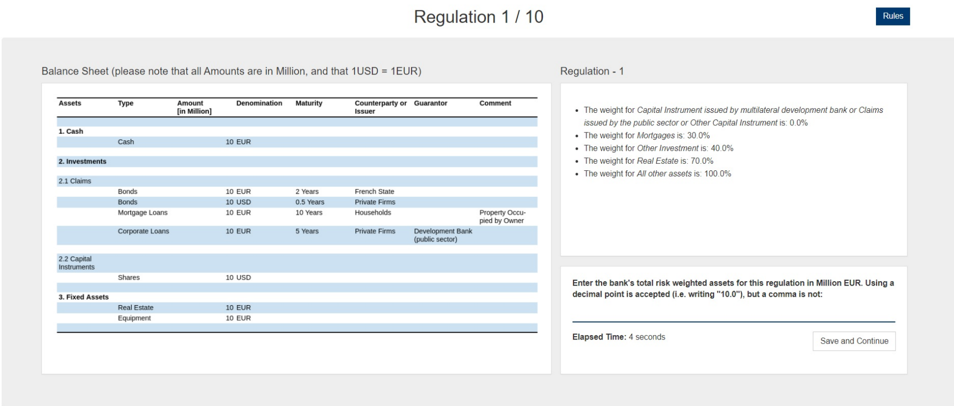

Figure 4: Online experiment - Test round.

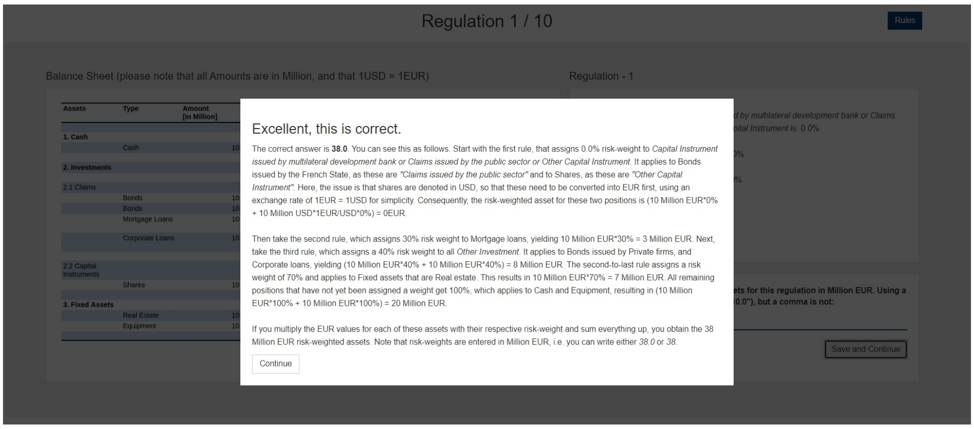

Figure 5: Online experiment - Feedback after correct answer in the test round.

Electronic copy available at: https://ssrn.com/abstract=3523824

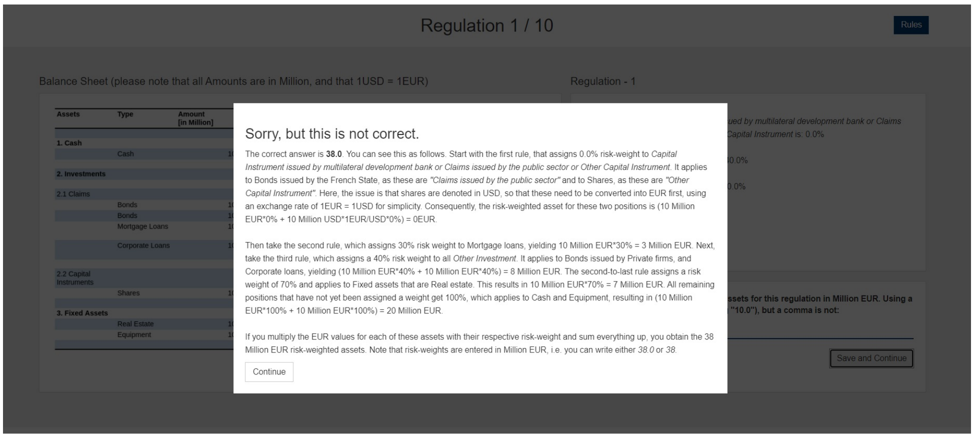

Figure 6: Online experiment - Feedback after wrong answer in the test round.

Electronic copy available at: https://ssrn.com/abstract=3523824

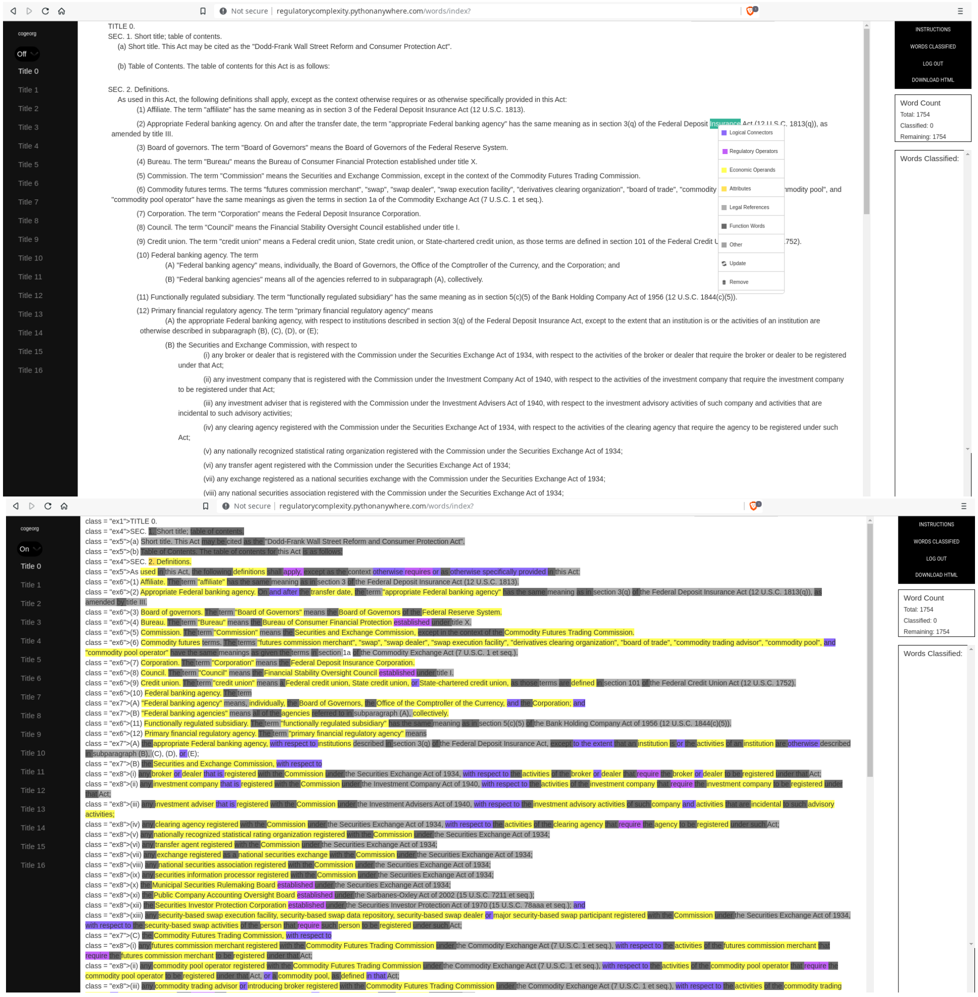

Figure 7: The Dashboard we developed to help us classify words in the Dodd-Frank Act as one of the following seven categories: Logical Connectors, Regulatory Operators, Economic Operands, Attributes, Legal References, Function Words, or Other. Top: The plain text of the Dodd-Frank Act. When highlighting 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.

Electronic copy available at: https://ssrn.com/abstract=3523824

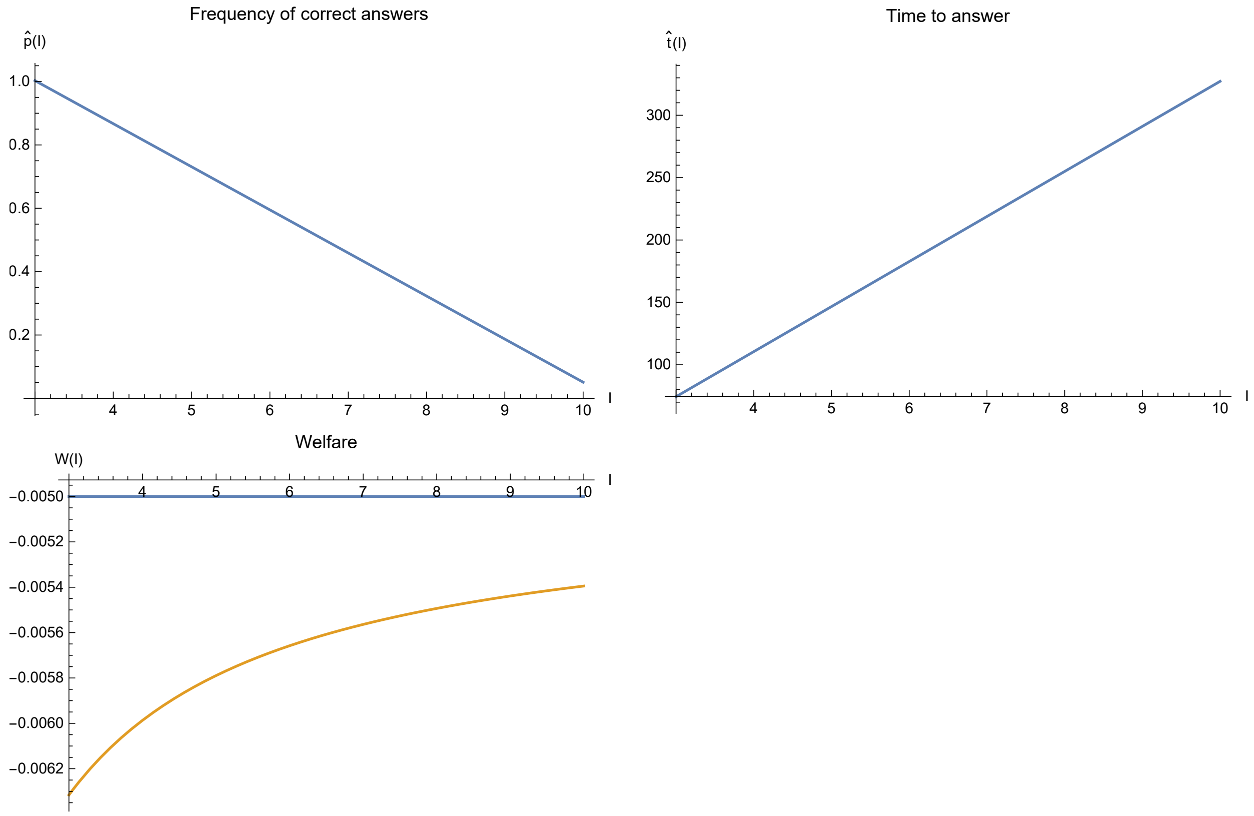

Figure 8: Frequency of correct answers, time to answer, and social welfare, as functions of the number of distinct asset classes.

Electronic copy available at: https://ssrn.com/abstract=3523824

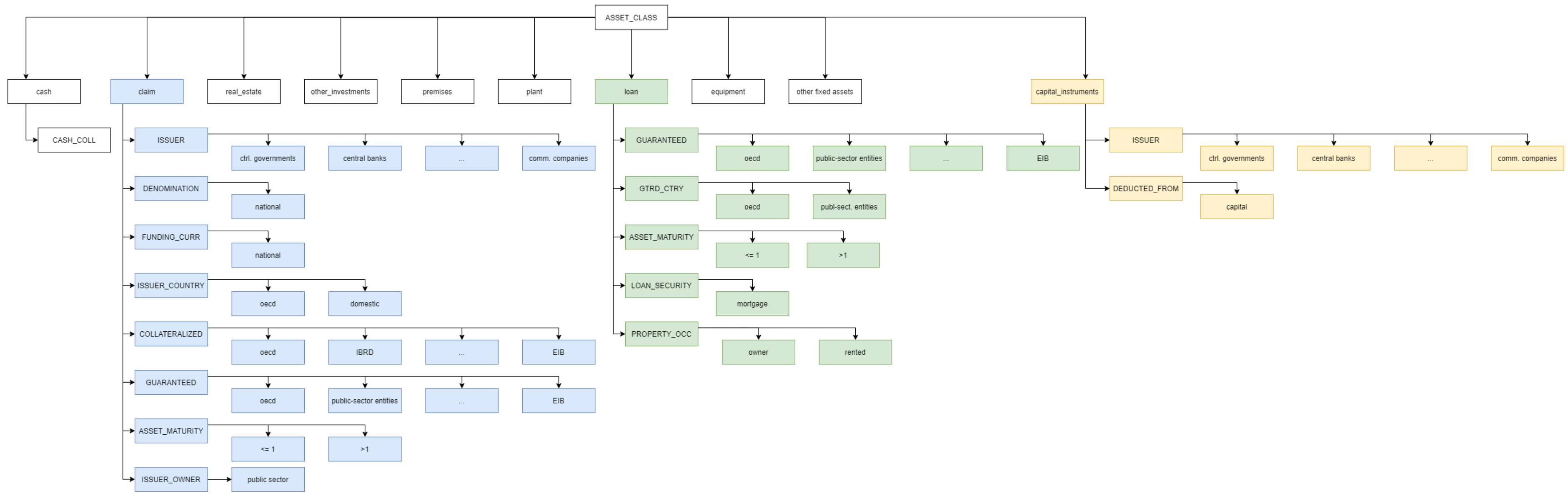

Figure 9: Decision tree with possible values for the "random regulation generator". Each box is an Economic Operand. Arrows indicate possible values of an Economic Operand in a bank's balance sheet. Colors are used for visual guidance only to make it easier to see which choices are part of the same branch in the tree.

Electronic copy available at: https://ssrn.com/abstract=3523824

C 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 5.

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)

Electronic copy available at: https://ssrn.com/abstract=3523824

IF (
    ASSET_CLASS == "claims" AND (
        (ISSUER == "IBRD" OR ISSUER == "IADB" OR ISSUER == "AsDB" OR ISSUER == "AfDB" OR
            ISSUER == "EIB") OR
        (GUARANTEED == "IBRD" OR GUARANTEED == "IADB" OR GUARANTEED == "AsDB" OR
            GUARANTEED == "AfDB" OR GUARANTEED == "EIB") OR
        (COLLATERALIZED == "IBRD" OR COLLATERALIZED == "IADB" OR COLLATERALIZED == "AsDB" OR
            COLLATERALIZED == "AfDB" OR COLLATERALIZED == "EIB")
) THEN:
    risk_weight = 0.2;
	Claims on multilateral development banks (IBRD, IADB, AsDB, AfDB, EIB) and claims guaranteed by, or collateralized by securities issued by such banks shall have a 20% risk weight

IF (
    ASSET_CLASS == "claims" AND (
        (ISSUER == "bank" AND ISSUER_COUNTRY == "oecd")
) OR
    ASSET_CLASS == "loans" AND
        (GUARANTEED == "bank" AND GUARANTEED_COUNTRY == "oecd")
) THEN:
    risk_weight = 0.2;
	Claims on banks incorporated in the OECD and loans guaranteed by OECD incorporated banks shall have a 20% risk weight

IF (
    ASSET_CLASS == "claims" AND (
        (ISSUER == "bank" AND ISSUER_COUNTRY != "oecd" AND ASSET_MATURITY <= 1)
) OR
    ASSET_CLASS == "loans" AND
        (GUARANTEED == "bank" AND GUARANTEED_COUNTRY != "oecd" AND ASSET_MATURITY <= 1)
) THEN:
    risk_weight = 0.2;
	Claims on banks incorporated in countries outside the OECD with a residual maturity of up to one year and loans with a residual maturity of up to one year guaranteed by banks incorporated in countries outside the OECD shall have a 20% risk weight

IF (
    ASSET_CLASS == "claims" AND (
        (ISSUER == "public_sector_entities" AND ISSUER != "central_governments" AND
            ISSUER_COUNTRY == "oecd" AND ISSUER_COUNTRY != "domestic")
) OR
    ASSET_CLASS == "loans" AND
        (GUARANTEED == "public_sector_entities" AND GUARANTEED != "central_governments" AND
            GUARANTEED_COUNTRY == "oecd" AND GUARANTEED_COUNTRY != "domestic")
) THEN:
    risk_weight = 0.2;
	Claims on non-domestic OECD public-sector entities, excluding central government, and loans guaranteed by such entities shall have a 20% risk weight

Electronic copy available at: https://ssrn.com/abstract=3523824

<pre>IF (
    ASSET_CLASS == "cash" AND (
        CASH_COLLECTION == "in_process"
)
) THEN:
    risk_weight = 0.2;</pre>
	Cash items in process of collection shall have a 20% risk weight

<pre>IF (
    ASSET_CLASS == "loans" AND
        (LOAN_SECURITY == "mortgage" AND
(PROPERTY_OCCUPIED == "owner" OR
            PROPERTY_OCCUPIED ==
"rented"))
) THEN:
    risk_weight = 0.5;</pre>
	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

<pre>IF (
    ASSET_CLASS == "claims" AND (
        ISSUER == "private_sector"
) THEN:
  risk_weight = 1.0;</pre>
	Claims on the private sector shall have a 100% risk weight

<pre>IF (
    ASSET_CLASS == "claims" AND (
        (ISSUER == "banks" AND
ISSUER_COUNTRY != "oecd" AND
ASSET_MATURITY > 1)
) THEN:
  risk_weight = 1.0;</pre>
	Claims on banks incorporated outside the OECD with a residual maturity of over one year shall have a 100% risk weight

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

<pre>IF (
    ASSET_CLASS == "claims" AND (
        (ISSUER ==
"commercial_companies" AND ISSUER_OWNER
== "public_sector")
) THEN:
  risk_weight = 1.0;</pre>
	Claims on commercial companies owned by the public sector shall have a 100% risk weight

<pre>IF (
    (ASSET_CLASS == "premises" OR
ASSET_CLASS == "plant" OR ASSET_CLASS ==
"equipment" OR
        ASSET_CLASS ==
"other_fixed_assets") OR
) THEN:
  risk_weight = 1.0;</pre>
	Premises, plant and equipment and other fixed assets shall have a 100% risk weight

<pre>IF (
    (ASSET_CLASS == "real_estate" OR
ASSET_CLASS == "other_investments") OR
) THEN:
  risk_weight = 1.0;</pre>
	Real estate and other investments (including non-consolidated investment participations in other companies) shall have a 100% risk weight

<pre>IF (
    ASSET_CLASS == "capital_instruments"
AND (
        (ISSUER == "banks" AND
DEDUCTED_FROM != "capital")
) THEN:
  risk_weight = 1.0;</pre>
	Capital instruments issued by other banks (unless deducted from capital) shall have a 100% risk weight

<pre>ELSE THEN:
  risk_weight = 1.0;</pre>
	All other assets shall have a 100% risk weight

Electronic copy available at: https://ssrn.com/abstract=3523824

D 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%

E 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. 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, Func-

Electronic copy available at: https://ssrn.com/abstract=3523824

tion 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.

F 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:

$$\lambda(1-p) > \delta. \tag{5}$$

For a given level of equity E and an asset type x, total welfare writes as:

$$pr(x) + (1-p)[1-x-\lambda \min(x-E,0)] - \delta E. \tag{6}$$

We want to derive an objective function for the regulator. As $pr(x)+(1-p)(1-x)$ is exoge-

Electronic copy available at: https://ssrn.com/abstract=3523824

nously given, we can consider the following objective function:

$$\mathscr{W}(E,x) = -\lambda(1-p)\min(x-E,0) - \delta E. \tag{7}$$

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

$$\int_0^1 \mathscr{W}(x,x)dx = \int_0^1 -\delta x dx = -\frac{\delta}{2}. \tag{8}$$

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 $E \geq E_i$. For a given interval $[a,b]$ the optimal capital requirement $E^*_{a,b}$ is given by:

$$E^*_{a,b} = b - \delta\frac{b-a}{\lambda(1-p)}. \tag{9}$$

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

$$\begin{aligned} W_{a,b}(E) &= \int_a^b [-\lambda(1-p)\min(x-E,0) - \delta E]dx && (10)\\ &= -\lambda(1-p)\int_E^b (x-E)dx - \delta E(b-a) && (11)\\ &= -\lambda(1-p)\frac{(b-E)^2}{2} - \delta E(b-a). && (12) \end{aligned}$$

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

Note that we indeed have $a \leq E^*_{a,b} \leq b$. 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 $1-p$).

Electronic copy available at: https://ssrn.com/abstract=3523824

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

$$W_{a,b}(E^*_{a,b}) = \delta(b-a)\left[\frac{\delta(b-a)}{2\lambda(1-p)} - b\right]. \tag{13}$$

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:

$$\begin{aligned} W_{0,\bar{x}}(E^*_{0,\bar{x}}) + W_{\bar{x},1}(E^*_{\bar{x},1}) &= \delta\bar{x}\left[\frac{\delta\bar{x}}{2\lambda(1-p)} - \bar{x}\right] + \delta(1-\bar{x})\left[\frac{\delta(1-\bar{x})}{2\lambda(1-p)} - 1\right] & (14)\\ &= \delta\bar{x}(1-\bar{x})\frac{\lambda(1-p)-\delta}{\lambda(1-p)} - \frac{\delta}{2\lambda(1-p)}[\delta - 2\lambda(1-p)]. & (15) \end{aligned}$$

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:

$$\begin{aligned} W_{i/I,(i+1)/I}(E^*_{i/I,(i+1)/I}) &= \frac{\delta}{I}\left[\frac{\delta}{2I\lambda(1-p)} - \frac{i+1}{I}\right] & (16)\\ &= \frac{\delta}{I^2}\left[\frac{\delta - 2\lambda(1-p)}{2\lambda(1-p)} - i\right]. & (17) \end{aligned}$$

We use this lest expression to compute total welfare:

$$\mathscr{W}(I) = \sum_{i=0}^{I-1} W_{i/I,(i+1)/I}(E^*_{i/I,(i+1)/I}) = -\frac{\delta}{2} - \frac{\delta}{2I\lambda(1-p)}[\lambda(1-p)-\delta]. \tag{18}$$

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 (18) to plot $\mathscr{W}(I)$ on Fig. 8, with $\lambda = 0.05$, $\delta = 0.01$, and $p = 0.05$. These parameters are meant for illustration only.

Electronic copy available at: https://ssrn.com/abstract=3523824

G Experiments - Panel analysis

In this section we go beyond the cross-sectional analysis at the regulation level of Section 3 and further analyze the determinants of the participants' answers by running panel regressions. The participants' answers form a balanced panel with 118 students, indexed by i, answering a series of 9 questions each, indexed by t. The t-th question for each student is drawn randomly from our 38 randomly generated regulation, and draws are independent across questions and students. Denoting $X_{i,t}$ a complexity measure for the $t-$th regulation given to student i, and $Y_{i,t}$ a measure of the performance of student i on question t, we study the power of X to explain the variation in Y. First, we evaluate the following probit model:

$$\ln\left(\frac{\Pr(correct_{i,t}=1)}{\Pr(correct_{i,t}=0)}\right) = \alpha + \beta X_{i,t}, \tag{19}$$

where $correct_{i,t} = 1$ if student i gave the correct answer to question t, and 0 otherwise. We estimate the model with student and question fixed effects.[17]

Table 18: Explaining the probability to find the correct answer. OLS regression with student and question fixed effects.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Length	-0.037***						-0.002
	(-7.95)						(-0.07)
Cyclomatic		-0.095***					-0.012
		(-6.23)					(-0.23)
Quantity			-0.507***				-0.563***
			(-9.34)				(-5.25)
Potential				-0.078***			-0.061*
				(-7.59)			(-1.72)
Diversity					-0.243***		0.555***
					(-4.34)		(4.61)
Level						4.061***	
						(6.52)	
Log-Likelihood	-416.15	-430.59	-397.89	-419.72	-441.26	-428.57	-385.92
Pseudo-R^2	0.243	0.217	0.277	0.237	0.198	0.221	0.298

The results are displayed in Table 18. We obtain that all measures of complexity except *level* are associated with a lower probability of answering correctly. In terms of explana-

[17] This is in principle unnecessary as regulations are randomly assigned to students, but we include these effects for robustness due to the relatively small sample size. Note that with student fixed effects there is no variation to explain for the 3 students who did not answer a single question correctly and the 14 students who answered all questions correctly.

Electronic copy available at: https://ssrn.com/abstract=3523824

tory power, measured either by the log-likelihood of the model or the pseudo-R^2, the best measure is $quantity$, followed by $length$ and $potential$. Column (7) uses all measures as regressors, except length (which is mechanically related to $length$ and $potential$). This regression must be interpreted with caution as the regressors are highly correlated with each other. However, we observe that the Pseudo-R^2 in this regression is 111% higher than in the regression using Column (1) only, which strongly suggests that the additional measures capture dimensions of complexity which are not already subsumed in $length$.

Next, we run a similar analysis but on the time taken to provide a correct answer. As in Section 3, we exclude answers with time below 7 seconds (1st percentile) or above 579 (99th percentile). We also keep only the correct answers, and run the following regression with OLS, both without and with student and question fixed effects, and where $time_{i,t}$ is the time (in seconds) taken by student i to answer question t.

$$time_{i,t} = \alpha + \beta X_{i,t}. \tag{20}$$

The results are displayed in Table 19. We obtain that all measures except $level$ are positively associated with the time to answer, again with the exception of $level$. The measure with the strongest explanatory power is $potential$, closely followed by $length$ and $quantity$. The coefficients have a straightforward interpretation: a regulation with an extra word takes an extra 3.4s to process on average, whereas a regulation with an extra unique operand takes 8.1s more, and a regulation with an extra regulatory instruction 34.6s. We again find that adding extra measures to $length$ increases the explanatory power, though not to the same extent as in the previous table, and that $potential$ seems to be a less noisy measure than $length$.

Electronic copy available at: https://ssrn.com/abstract=3523824

Table 19: Explaining the time taken to provide a correct answer. OLS regression with student and question fixed effects.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Length	3.394***						1.219
	(14.39)						(0.75)
Cyclomatic		9.338***					-2.598
		(10.17)					(-0.95)
Quantity			33.159***				7.841
			(13.83)				(1.37)
Potential				7.992***			5.772***
				(15.41)			(2.95)
Diversity					39.387***		-4.339
					(12.28)		(-0.71)
Level						-265.670***	
						(-6.89)	
R^2	0.54	0.47	0.53	0.55	0.50	0.42	0.56

Electronic copy available at: https://ssrn.com/abstract=3523824