NEW DIMENSIONS OF REGULATORY COMPLEXITY AND THEIR ECONOMIC COST. AN ANALYSIS USING TEXT MINING

2021

BANCO DE ESPAÑA
Eurosistema

Documentos de Trabajo
N.º 2107

Juan de Lucio and Juan S. Mora-Sanguinetti

NEW DIMENSIONS OF REGULATORY COMPLEXITY AND THEIR ECONOMIC COST. AN ANALYSIS USING TEXT MINING

NEW DIMENSIONS OF REGULATORY COMPLEXITY AND THEIR ECONOMIC COST. AN ANALYSIS USING TEXT MINING [*]

Juan de Lucio [**]

UNIVERSIDAD DE ALCALÁ DE HENARES AND UNIVERSIDAD NEBRIJA

Juan S. Mora-Sanguinetti [***]

BANCO DE ESPAÑA

(*) The views expressed are those of the authors and should not be attributed to the Banco de España or the Eurosystem. We appreciate the comments of the participants in the Banco de España seminar.
(**) Juan de Lucio. Universidad de Alcalá de Henares and Universidad Nebrija Email: Juan.deLucio@UAH.es. Juan de Lucio acknowledges financial support from Comunidad de Madrid & UAH (ref: EPU-INV/2020/006).
(***) Juan S. Mora-Sanguinetti. Senior Economist (Banco de España - Eurosistema) and Lawyer (ICAM). Contact: Email: JuanS.Mora@bde.es; Banco de España - Eurosystem. Directorate General Economics, Statistics and Research. Alcalá 48, 28014 Madrid (España). Tel.: (0034) 913385197. FAX: (0034) 913385486.

Documentos de Trabajo. N.º 2107

2021

ISSN: 1579-8666 (on line)

Abstract

Complex or poorly drafted regulations are more difficult for economic agents to implement, eroding economic efficiency. The literature has so far concentrated on the analysis of regulatory complexity as a phenomenon related to the "quantity" of norms. Regulation can also be complex because of "qualitative" reasons such as its linguistic ambiguity or its relational structure (references between legal documents). This article innovates by analyzing these new dimensions of complexity: we develop new indicators for legibility and regulatory interconnectedness. To do so, we constructed a new database (RECOS - Regulatory Complexity in Spain) extracting information from 8,171 norms (61 million words) covering the regulation set of all the Spanish autonomous regions. We analyze the relationship between these new indicators and productivity (as a relevant economic variable) and judicial efficacy (as a relevant institutional-structural variable). While each of these areas should be analyzed in separate articles, this research shows that the new dimensions of regulation complexity matter and also have significant results.

Keywords: Regulatory Complexity, productivity, linguistic complexity, relational complexity, legal corpus.

JEL classification: O43, K2, R11, O47.

Resumen

Los agentes económicos tienen una mayor dificultad para implementar normativas complejas, produciéndose una disminución de la eficiencia económica. La bibliografía se ha concentrado hasta ahora en el análisis de la complejidad como un fenómeno relacionado con la «cantidad» de normas. Sin embargo, la regulación también puede ser difícil de implementar por razones «cualitativas»: por su ambigüedad lingüística o por tener una estructura relacional (referencias entre documentos legales) compleja. Este artículo innova analizando estas dimensiones y desarrolla nuevos indicadores de legibilidad e interconexión regulatoria. Para ello, construimos una nueva base de datos (RECOS - Complejidad Regulatoria en España) extrayendo información de 8.171 normas (61 millones de palabras) que abarcan la regulación de todas las comunidades autónomas españolas. En nuestro análisis también mostramos la relación de estos nuevos indicadores con variables económicas (productividad) e institucionales (la eficacia judicial). Si bien cada una de estas áreas deberá analizarse en artículos separados, esta investigación revela que las nuevas dimensiones de la complejidad de la regulación (la complejidad lingüística y la relacional) son relevantes y tienen impactos significativos.

Palabras clave: complejidad de la regulación, productividad, complejidad lingüística, complejidad relacional, corpus jurídico.

Códigos JEL: O43, K2, R11, O47.

1 Introduction

Regulation, as part of the institutional framework, is fundamental to economic development. If it is well designed, it can help to mitigate market failures, such as imperfect information or externalities, and can generally reduce transaction costs[1]. However, if regulation is poorly drafted, it could lead to increased transaction costs, negatively impacting economic efficiency. Indeed, Laffont and Tirole (1993) argue that market failures would be a necessary but not sufficient condition for resorting to regulation since its effects are conditioned by transactional, administrative-political or informational reasons (see also Mora-Sanguinetti and Pérez-Valls, 2020).

Regulation may imply both direct and indirect economic costs. The former refers to resources devoted to the compliance of regulations. Indirect costs are associated with changes in agents' behavior (firms, consumers or government). Both direct and indirect costs, may derive in a misallocation of resources. Indeed, the literature in Economics, and more specifically that which analyses the impacts of institutional quality (Acemoglu and Robinson, 2012), makes several references to the costs derived from administrative burdens (red tape) (Hampton, 2005) which are a main motivation for "better regulation" policies (Radaelli, 2007).

Several studies have also found specific examples of the costs involved in regulation: Djankov et al. (2006) showed that improving business regulations, from the worst quartile to the best, increase the annual growth rate by 2.3 percentage points. Coffey et al. (2020) identify sectors affected by the regulations and concluded that economic growth in the U.S. has been dampened by federal regulations by 0.8% per annum. Bailey and Thomas (2017) showed that the industries that are more intensively regulated experienced lower enterprise birth rate and slower employment growth. Chambers et al. (2019b) concluded that, in the case of the U.S., a 10% increase in the effective federal regulatory burden increase the poverty rate by 2.5%. There are also particular areas of regulation which have been intensively analyzed in the literature. One of them is the regulation of retail trade, which would have negative impacts on employment in the sector [Bertrand and Kramarz (2002) or Viviano (2008)].

In addition to the potential economic costs, there are also legal costs, which in turn can impact the economy indirectly. The OECD (Palumbo et al., 2013) found that a low

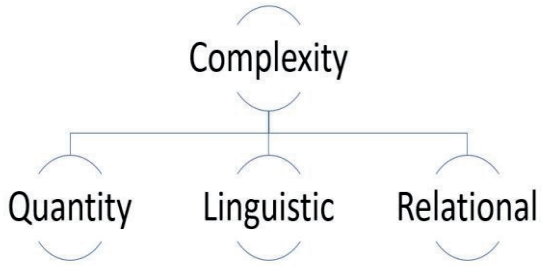
[1] The study of classical microeconomic theory, specifically the welfare theorems, leads to the same conclusion, as market failures violate the first theorem (see, among others, Mora-Sanguinetti and Salvador-Mora, 2016; Mora-Sanguinetti, 2019a).

quality of the regulation, measured with the World Bank's regulatory quality indicators, was related to more litigation in the countries analyzed. It should be noted that the increase in litigation has a relevant impact on the functioning of the judicial system, implying longer trial lengths. All this is important for the activity of lawyers and legal services professionals (Mora-Sanguinetti, 2019b).

One of the reasons why regulation may be inefficient is because it is “complex”. The concept of complexity refers to problems on the "form" and not of the specific topics covered by the regulation. A corpus of regulations can be complex in the first place because it is too broad (the “quantity” approach), i.e. there is an excessive volume of regulations or those regulations are coming from very different sources, making it difficult for economic agents to manage them and to verify their validity (Bardhan 2002; Di Vita, 2018). Secondly, norms may be "qualitatively complex" because they are ambiguously or poorly drafted (the "linguistic" approach), making them difficult for consumers and businesses to understand and comply with them. Additionally, complexity can derive from how rules are connected to each other (the “relational” approach). Norms drafted with more references to other norms demand more resources to be understood and enforced.

Figure 1 refers to these three sources of complexity.

Figure 1: The three dimensions of regulatory complexity.



Source: Own elaboration.

The first approach to complexity using the volume of norms has already been studied by some articles in the literature, relying on indicators which measure the number of legal texts. Using this type of indicators, Di Vita (2018) showed that regulatory complexity hinders regional growth in Italy. Kirchner (2012) showed that the growth in legislation, at a national level (Australia), measured by the number of pages is negatively related to growth in real income per capita in the short run. Mora-Sanguinetti and Pérez-Valls (2020) showed that regulation is negatively related to the number of firms in Spain

and may have effects on the size distribution of the Spanish firms, with an excessive SMEs share. This article continues that line of research by accounting for words and sentences.

The study of the last two approaches (complexity derived from linguistic or relational problems) has been scarce in the economic literature so far, being the main contribution of this article. We suggest using natural language processing (NLP) techniques to measure those dimensions. These tools are increasingly applied in the literature. They have been shown to be useful to measure economic uncertainty (see, Azqueta-Gavaldon et al. 2020, Ghirelli et al, 2019 and Baker et al. 2016). In the literature on financial markets, Tobback et al (2017) use NLP techniques to analyze the media's perception of the tone of the ECB´s monetary policy discourse, Ehrmann y Talmi (2020) focus on the semantic similarity of subsequent central bank statements to prove that volatility rises when changes occur after sequences of similar statements, Hansen et al. (2017) used them to detect issues in Federal Open Market Committee statements[2], Hassan et al. (2019) analyzed political risk using earnings conference calls, and Calomiris et al. (2020) derived a measure of firm-level regulatory costs from corporate earnings conference calls.

To the best of our knowledge, RegData is the only database that measures regulations quantitively using NLP techniques (see Al-Ubaydli and McLaughlin, 2017). The authors compiled industry-specific federal regulations for the US during the period 1997-2012 (see McLaughlin and Sherouse, 2019, for an update of the database). Using it, Davis (2017) related regulatory complexity with policy uncertainty, Chambers et al. (2019a) explored the link between regulation and prices, Coffey et al. (2020) regulation and economic growth and Bailey and Thomas (2017) industry regulation and enterprise birth and employment growth. Version 3.2 of the database was released in March 2020, cover the period 1970–2019 and include, for the first time, measures on complexity of regulatory texts with a similar methodology to the one developed in this paper, such as the Shannon Entropy legibility indicator.

Finally, our analysis is also related to Hurka and Haag (2019) which was the first article to use linguistic and structural-relational complexity measures based in NLP. They showed that the length of the decision-making process in the EU is influenced by different types of policy complexity.

[2] There is an increasing interest in analyzing the drafting and the complexity of the language employed in central bank statements. Hernández-Murillo and Shell (2014) and Coenen et al. (2017) have shown that statements have become more complex since the financial crisis. Jansen (2011) shows that more complex statements are associated with higher volatility. Haldane and McMahon (2018) focus on the simplification of central bank communications.

Our article develops and presents a new database (RECOS, Regulatory Complexity in Spain) covering the laws enacted by the regional governments over the whole Spanish democratic period (from 1978 until 2019). As far as we know this is the first effort to build a database using NLP techniques on regional regulations outside the USA. RECOS incorporates new legibility and relational indicators as well as the traditional volume indicators. As it is explained in sections 2 and 4, some of these measures of legal complexity are new in the literature. With our new indicators (on regulation readability and interdependency) in hand we explore the effects that regulatory complexity has on economic activity (productivity) and judicial efficacy (as representative of a relevant institutional dimension) using panel data analysis. The analysis is performed at the regional level.

The study of the Spanish case is of high interest for several reasons: first, it allows us to develop comparative indicators (by region) of regulatory complexity (and analyze its economic impact) without having to compare different languages. This last problem puts at risk other potential studies that would like to compare complexity in regions (or countries) with different languages. Spanish regions also provide an adequate comparison framework as all of them share the same basic institutional background: on the one side, the Spanish Constitution establishes a common framework of competences (without denying specificities)[3]. This implies that the set of subjects and problems solved by the regulations are more similar between Autonomous regions than between countries. On the other side, the international norms that affect Spain, similarly affect all the regions. Secondly, with regard to the analysis of economic impacts, specifically judicial data, the study of Spain is unique at the international level because it has richer judicial databases than many other jurisdictions. Thirdly, with regard to the construction of robustness checks, Spanish is a highly regulated language, with an official dictionary (developed by the Real Academia Española - RAE), which facilitates the development of objective measures of infrequent words. Finally, the case of Spain, beyond the studies mentioned, has not yet been widely analyzed even though it is one the largest economies of the Euro area.

The effects of the different complexity measurements are robust to different specifications and seem to be independent between each other. These results suggest that the linguistic and relational complexity dimensions introduced in this document are important perspectives of regulatory complexity, even more relevant than the traditional measures focused on the quantity of regulation, and affect different aspects of economic performance.

[3] See, for instance, articles 148 and 149 of the Spanish Constitution.

The remainder of the article proceeds as follows. Section 2 presents the database and the new indicators. Section 3 relates our indicators of regulatory complexity with measurements of productivity growth and judicial efficacy at the regional level using panel data analysis. Section 4 presents several robustness checks. The final section presents the main conclusions and future research avenues. The article is completed by an appendix which includes further detailed information.

2 Regulation in Spain

Our research analyzes all regional Laws and Decree-Laws published in the "Official State Bulletin" (*Boletín Oficial del Estado*, BOE) over the democratic period (from 1978 to 2019). In order to do so, we performed web data extraction (web scrapping) from *BOE.es*. We thus collected a set of regional regulations with the force of Law[4] summing up 8171 norms and 61 million words[5]. We analyze each regulation within the corpus using NLP techniques.

With the 1978 Constitution, Spain adopted a decentralized model of territorial and political organization. The regions (*Autonomous Communities*) have the power to adopt regulations with the force of law, and the Constitution allows them to have a very broad competence ceiling. In strict legal terms, the system is not called "federal" because, among other reasons, access to autonomy is not compulsory, even though all the regions have it [see, among others, López-Guerra (1994) or López-Guerra *et al*. (2018)]. The system as a whole has largely converged with the "classical" federal model and is, according to various international indicators, as decentralized as the federal ones (in the Regional Authority Index, RAI, Spain's score would be comparable to that of Canada or the USA and is higher than that of the UK; see Hooghe *et al*. 2008).

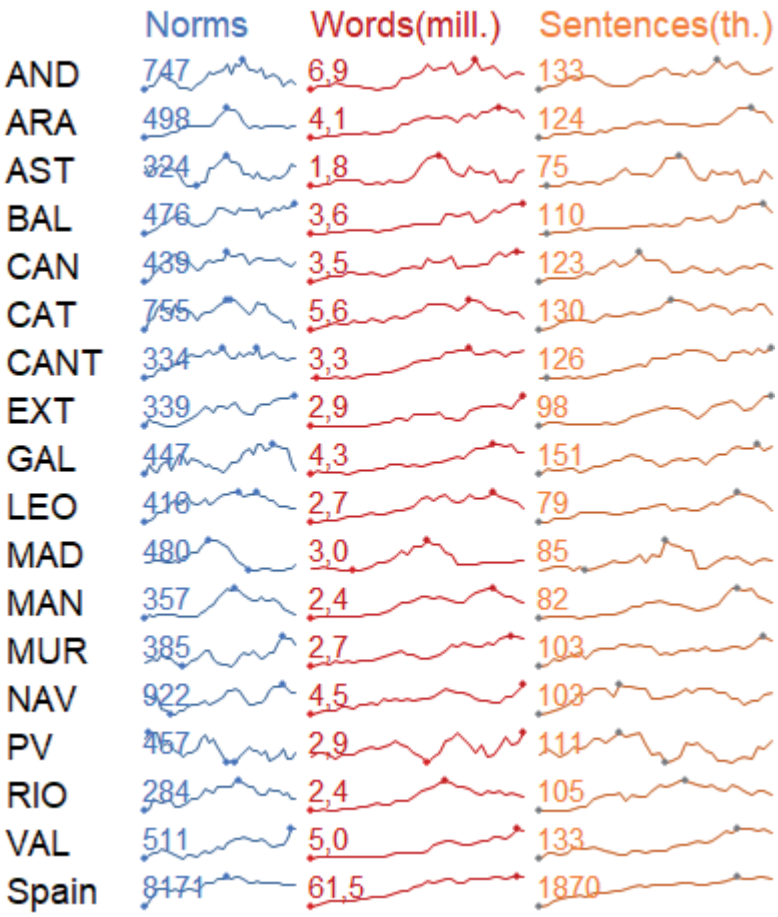
We compared the formal aspects of the regulations of the different regions. Although some regions have their own regional language (in addition to Spanish), we have obtained the official Spanish version of all of them. This mitigates comparison problems when using text analysis techniques. Additionally, as it was already discussed, all regions share a common "basic" institutional framework composed by the Spanish Constitution, the Organic Laws (among others) and the international legal framework.

[4] Regulations with the force of law are part of the regulations adopted by the regional administrations, which can also pass lower-level regulations. Taking Aranzadi as a general source [see Mora-Sanguinetti (2019a)], only in 2019 the regions adopted 7970 norms in total, being the level of administration that adopts more rules (the central administration adopted 1785 and the local-provincial administration 578). For further information, in 2020 the pattern continued: regional administrations adopted 9645 regulations, compared to 1929 for the central administration and 676 for the local-provincial administrations.
[5] For an average person reading 8 hours per day it would take more than a year just to read the corpus.

2.1 Measurement of the volume of regulations

Figure 2 shows the total number of norms (second column), the total number of words (third column) and the total number of sentences of those norms (fourth column) for each region. It also graphs the evolution of those measurements during the period of analysis. The markers (small rhombuses) of each series in Figure 2 highlight the minimum and the maximum level. Usually the maximum takes place in the most recent years; there is a generalized upward trend in the amount of regulations incorporated each year. We observe that some regions have adopted 3 times more norms than others. For instance, Navarra (NAV) has adopted more than nine hundred regulations meanwhile Rioja (RIO) has less than three hundred. Andalusia (AND) norms contain almost 7 million words, but Asturias (AST) ones only 1.8 million. For some regions the average number of words per sentence is 23 (Rioja) meanwhile in Andalusia is 52 words per sentence.

Figure 2: Basic volume indicators of the regional legal corpus



Source: Own elaboration.
Note: AND: Andalusia, ARA: Aragón, AST: Asturias, BAL: Balearic Islands; CAN Canary Islands, CAT: Catalonia. CANT: Cantabria, EXT: Extremadura, GAL: Galicia, LEO: Castile and Leon, MAD: Community of Madrid, MUR: Murcia; NAV: Navarra, PV: Basque Country, RIO: La Rioja, VAL: Valencia. "Spain" is the sum of all the regional norms (thus not including the central administration norms).

The measurements of the number of norms, sentences, and words are part of the "quantity approach" to complexity. However, these three indicators are independent: the length of the corpus is not necessarily related to the length of the norms within it. Indeed, the correlation between the number of norms and the number of words is 6% and with the number of sentences is 0%.[6] In this paper we will work with the measurement of the number of norms as this is the usual approach to the volume of regulation used in the literature.

2.2 The relational approach: Network analysis

An additional element that contributes to the complexity of regulation is the external interdependence between norms. Citations and references to other norms increase the cost of knowledge acquisition. That is, a legal text that obliges the reader to consult other legal texts to reach a complete understanding of its content will demand higher effort to be used and enforced. We have built an indicator which computes the average number of norms that are referenced in the norms that are adopted per each region in a specific year.

Figure 3 graphs the links between norms per each region during the analyzed period. Each network corresponds to a region. For the sake of simplicity, it only includes those texts with more than 50 links. Figure 3 illustrates that there are important differences between the networks of different regions. Some regions, like Aragon (ARA) and Valencia (VAL), have much more intricate structures. In contrast, Asturias (AST) or Navarra (NAV) do not have any norm with more than 50 links. We also have regions such as Extremadura (EXT), Castilla la Mancha (MAN) or La Rioja (RIO) with low connectivity within their legal corpus. A more interconnected graph points to a more complex structure of the legal corpus.

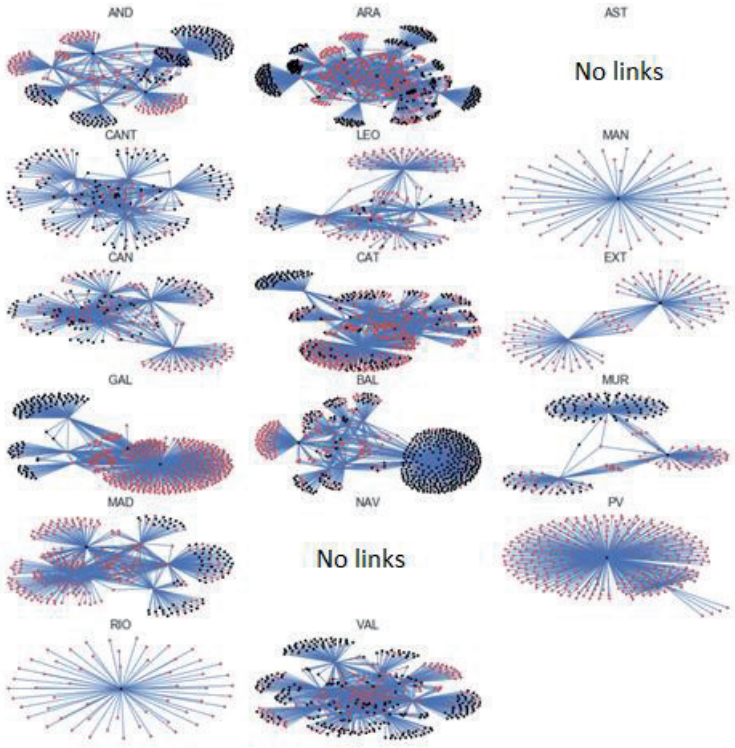
The network representation as graphed in Figure 3 provides a static proxy of the stock of regulations and their relational structure. In order to construct panel data estimations, we need to explore its time variance as well, i.e. the evolution of these network structures over time. Therefore, we also compute the average number of new links generated by the regulation adopted during each year according to equation (1).

[6] The correlation between words and sentences amounts to 90%. The correlation between words and paragraphs amounts to 79%. The correlation between sentences and paragraphs amounts to 87%.

$$\#Links_{reg,t} = \frac{\sum_{n\,\epsilon\,(reg,t)} Links_n}{\sum_{n\,\epsilon\,(reg,t)} n} \quad (1)$$

Where $\#Links_{reg,t}$ corresponds to the average number of links that incorporates the norms adopted by a region in the year, t. The numerator is the sum of the links incorporated in the norms, n, adopted during a year. Denominator is the number of norms adopted by the region, reg, in year, t.

Figure 3: Regional network of norms (only for norms with more than 50 inward-outward links)



Source: Own elaboration. Red: regional regulations. Black: other regulations (usually national).
Note: AND: Andalusia, ARA: Aragón, AST: Asturias, BAL: Balearic Islands; CAN Canary Islands, CAT: Catalonia. CANT: Cantabria, EXT: Extremadura, GAL: Galicia, LEO: Castile and Leon, MAD: Community of Madrid, MUR: Murcia; NAV: Navarra, PV: Basque Country, RIO: La Rioja, VAL: Valencia.

The last column of Figure 4 presents our relational indicator. We observed that the maximum number of links in all the regions is always in a more recent year than the minimum. These reflects that the average number of external links of legal text has increased over the democratic period.

Figure 5 in the Appendix presents the distribution of the indicator across regions. On average, new laws incorporate references to 11 different legal text. Some regions like

Aragón (ARA) incorporates more external references, about 18, meanwhile other regions like Navarra (NAV) only incorporates 3 references. In Figure 3 we observed that Navarra does not have regulations with over 50 references, meanwhile Aragón had a very interconnected structure. The proposed indicator is able to capture some network characteristics which can be observed in the graphs (size, density and connectivity) as well as the evolution over the period of analysis.

2.3 The linguistic approach to complexity: indicators of legibility

The linguistic approach to complexity analyzes the lexical structure of the texts (paragraphs, sentences and words). Regulations may be "complex" because they are ambiguously or poorly drafted, making them difficult for consumers and businesses to understand and to comply with them. The data presented in the middle column of Figure 4 follows the μ indicator proposed by Muñoz and Muñoz (2006). This indicator has been used recently in other works which analyze the Spanish language [Brelsford *et al*. (2018) and Campillo et al. (2020)].[7] It is computed following equation 2.

$$\mu_n = \left(\frac{Wo_n}{Wo_n - 1}\right)\left(\frac{\overline{\mathrm{Le}}_n}{\sigma^2_{Le_n}}\right) * 100 \qquad (2)$$

Where, n, refers to norms; Wo, corresponds to the number of words; $\overline{\mathrm{Le}}$, is the average number of letters per word and, σ^2_{Le}, refers to the variance in the number of letters per word[8]. In practice, μ usually takes values between 0 and 100, although it can reach higher values. Greater values of the μ indicator are associated with better readability. Lower values correspond with more complex regulations, lower legibility.

The final value of the legibility indicator is the mean value of the legibility indicator of the norms approved in the region during the year, $Legib_{reg,t} = \bar{\mu}_{n\in(reg,t)}$.

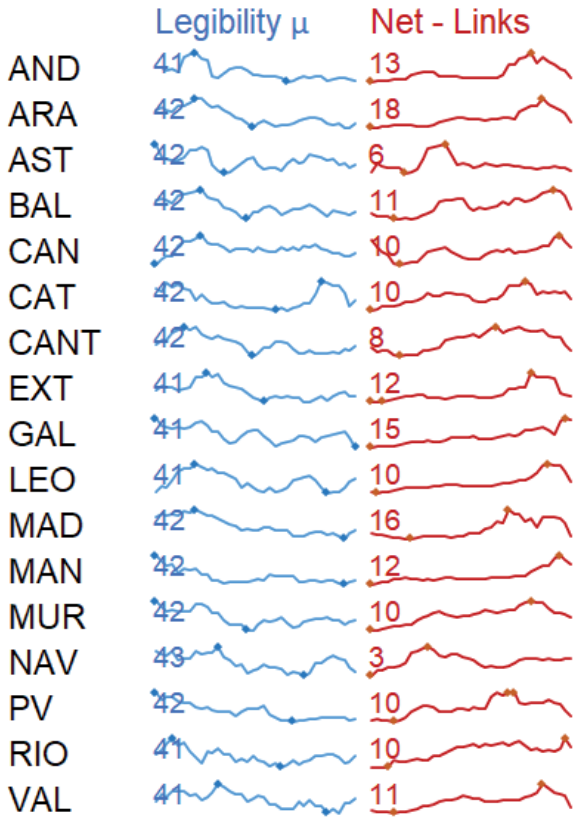
For most regions, we observe across regions a downward trend in legibility. The small graphs draw also the minimum and maximum values of the series. The ease with which a reader can decode norms seems to be lower during the last period of the series.

[7] Alternatives to our indicator are those derived from the work of Flesch (1948). Fernández (1959) and the Flesch-Szigriszt indicator (1993) adapted the indicator to the Spanish language but used an ad-hoc parameterization which is not without criticism.

[8] For texts with a number of words such as those considered in this document, the first parenthesis tends to one and therefore the formula is the inverse of Pearson's coefficient of variation, indicating the relationship between the standard deviation of a sample and its mean, divides by the standard deviation.

Please note that the maximum level of legibility used to be located at the beginning of the graph.

Figure 4: Basic regulatory complexity indicators at the regional level



Source: Own elaboration.
Note: AND: Andalusia, ARA: Aragón, AST: Asturias, BAL: Balearic Islands; CAN Canary Islands, CAT: Catalonia. CANT: Cantabria, EXT: Extremadura, GAL: Galicia, LEO: Castile and Leon, MAD: Community of Madrid, MUR: Murcia; NAV: Navarra, PV: Basque Country, RIO: La Rioja, VAL: Valencia. "Spain" is the sum of all the regional norms (thus not including the central administration norms).

Figure 5 in the Appendix presents the distribution across regions of the indicator. The usual values for the legibility indicator variables between 40 and 44, which correspond to difficult texts (the mean is 41,7, the median is 41,6 and the standard deviation is 1,4). All data but two are within the range 38-48 which, that according to Muñoz and Muñoz (2006) correspond with difficult texts (very difficult texts are below 31 and easy texts have μ indicator value over 70). Navarra (NAV) have legal norms that are more legible than Andalusia (AND), Extremadura (EXT) or Valencia (VAL). The variable considered for estimation is the average value of legibility indicator for the norms adopted during each year.

Lower legibility and higher interconnected laws, result in more complex regulations. Their effects are complementary but independent to the measurement of the quantity of regulation, as it will be shown in the next sections. Correlation matrix between

variables shows that the correlations between number of norms and legibility is only -2%, and -10% with the number of links. Number of links and legibility correlates -33%.

3 Examples of the impact of regulatory complexity on economic efficiency

From a theoretical point of view, more complex regulations are less effective in reducing transaction costs. In the worst case, complexity may entail costs for businesses and citizens and may lead to resource misallocation. In this section, we make use of the variables developed in section 2 to preliminarily explore their economic and judicial impacts. More specifically, we empirically analyze the relationship between regulatory complexity in Spain, focusing on the new indicators (linguistic and relational complexity) and two sets of structural variables at the regional level: (labor) productivity and judicial efficacy. It is part of the future research agenda to analyze in depth each of the different channel of these relationships.

3.1 Comprehensive macroeconomic impacts: labor productivity

We present the impacts of that complexity on a comprehensive macroeconomic measure: labor productivity, defined as value added (VA) per hour. We prefer this measure (VA per hour) to a more general one such as the GDP per capita for two reasons: it is more robust to underlying regional demographic and labor market trends (for instance, unemployment or participation rate) and it is closer to structural economic growth (Harvie et al. 2019). The source of regional macroeconomic data is De la Fuente (2019). Regional information on real VA per hour is available for the period 1977-2017. Those regions with more complex regulations are expected to experiment lower productivity. Within each region, periods with more complex regulations are expected to be negatively related to productivity.

This exploration is based on the fact that regulatory complexity is supposed to be related to the structural component of productivity. More specifically, the complexity of regulation has a negative impact on productivity through "total factor productivity" (TFP). TFP growth captures the effects of a number of mixed factors, including the impact of the quality of the institutional environment, such as the regulation of product and labor markets and the capacity of the economy to innovate on the productive use of labor,

capital and other inputs. The TFP growth is often defined as "technological progress" [see Scarpetta et al. (2002), Fuentes and Mora-Sanguinetti (2012), Mora-Sanguinetti (2021)].

We estimated equation (3) with panel data covering the 17 Spanish regions. Our dependent variable is productivity. As it was discussed before, our measures of complexity are: first, the number of norms as a proxy of the volume of regional regulation (we expect a negative relationship); second, average μ legibility as a proxy of linguistic complexity (we expect a positive relationship); and third, our relational variable, measured with the average number of links present in the new norms (we expect a negative relationship). We present the correlation matrix of these variables and other descriptive statistics in the Appendix, Figure 6. Our estimates include regional and time fixed effects, all variables enter the estimation in logs. As in the rest of estimates, errors have been clustered at the regional level.

$$Prod_{reg,t} = Cte + \beta_1 \underbrace{\#Norms_{reg,t-i}}_{Quantity} + \beta_2 \underbrace{Legib_{reg,t-i}}_{Linguistic} + \beta_2 \underbrace{\#Links_{reg,t-i}}_{Relational} + \varepsilon_{reg,t} \quad (3)$$

Dimensions of complexity:

Estimates are presented in Table 1. Columns (1) to (3) present the results when the complexity variables are included independently. The estimates take the form of a traditional convergence regression, which control for the initial productivity value. All measurements of complexity have the expected sign. The new complexity indicators have a (statistically) significant effect. Column (4) includes both the lexical complexity and the relational complexity dimensions. Both maintain their sign but the relational complexity dimension loses its statistically significance level when the quantity indicator is also incorporated. Column (5) includes all the complexity indicators simultaneously; the estimated parameters maintain the sign, magnitude and significant levels.

Table 1 is in line with the findings of Di Vita (2018) which approximates the complexity of regulation through the quantity approach (number or volume of norms) and shows that an increase in the number of norms reduces productivity. Our results also show that other dimensions of regulation affect productivity, consistently with Coffey et al. (2020) and Djankov et al. (2006). First, legibility is positively related to productivity. Finally, the analysis of our "relational" variable also shows that a greater number of links to other norms in the new regulations reduces productivity. These results are new in the literature and confirm that when analyzing regulatory complexity it is useful to examine further factors in addition to the volume of regulations. These other factors do not seem

to be related to the volume of regulation (number of norms) which has been traditionally used as proxy for complexity.

Table 1: Labor productivity per hour. Panel data Fixed effects

	(1)	(2)	(3)	(4)	(5)
# Norms $_t$	0.00232				0.00181
	(0.00158)				(0.00150)
Legibility $_t$		0.0769***		0.0642**	0.0649**
		(0.0261)		(0.0226)	(0.0225)
# Links $_t$			-0.00348*	-0.00257	-0.00221
			(0.00196)	(0.00186)	(0.00221)
Constant	0.551***	0.259	0.544***	0.309	0.315
	(0.146)	(0.196)	(0.140)	(0.177)	(0.183)
Labor productivity $_{t-1}$	0.947***	0.948***	0.949***	0.948***	0.947***
	(0.0148)	(0.0144)	(0.0144)	(0.0146)	(0.0152)
Fixed effects					
Time	Yes	Yes	Yes	Yes	Yes
Regional	Yes	Yes	Yes	Yes	Yes
Observations	583	583	583	583	583
R-squared	0.985	0.985	0.985	0.985	0.985

Robust (clustered) standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1
Source: Own elaboration

The effects presented in table 1 are contemporary. However, there are numerous judicial rulings (from the Constitutional Court)[9] in Spain that suggest that regulation is predictable (and therefore could have anticipated effects on economic efficiency). Mora-Sanguinetti and Pérez-Valls (2020) thus analyzed the advanced effects of the quantitative complexity of regulation (in addition to the contemporary ones) on business demographics. It is debatable whether agents can foresee linguistic or relational complexity beyond quantitative complexity.

3.2 Effects of complexity on legal costs (litigation and judicial efficacy)

Regulatory complexity may also imply legal costs. That is, those related to the effectiveness of the judicial system and the enforcement mechanisms available in the economy. As noted in the introduction, the OECD (Palumbo et al., 2013) found that a low

[9] STC 9/2019 of 17 January (BOE no. 39 of 14 February 2019), STC 167/2016 of 6 October (BOE no. 276 of 15 November 2016), STC 121/2016 of 23 June (BOE no. 181 of 28 July 2016) among many others (see Mora-Sanguinetti and Pérez-Valls, 2020).

quality of the regulation was related to more litigation in the countries analyzed. An increase in litigation is also related to lower judicial efficacy. In other words, the judicial system would show higher rates of congestion or longer trial lengths. Judicial inefficacy has important implications for economic efficiency in Spain, reducing credit (Mora-Sanguinetti et al., 2017), investment at the enterprise level (Dejuán and Mora-Sanguinetti, 2021) or the proportion of rented housing (Mora-Sanguinetti, 2012).

Our dependent variables in this case will be a measure of trial length in the civil jurisdiction following the CEPEJ (2016) approach (see equation 4). The trial length is a measure that approximates the congestion of the judicial system. In other words, its ineffectiveness in swiftly resolving cases brought before the courts. The raw data for its calculation comes from the General Council of the Judiciary.

$$Trial\ length_{reg,t}(CEPEJ) = \frac{Pending\ cases_{reg,t}}{Cases\ resolved_{reg,t}} * 365 \quad (4)$$

We analyze this effect by means of equation (5). As before, we estimated the equation with panel data covering the 17 Spanish regions. All variables enter the estimation in logs. Following the empirical strategy explained above, we introduce each variable of regulatory complexity independently. All estimates include fixed effects at the region level and time dummies. As control variables we include productivity and the number of lawyers [see Mora-Sanguinetti and Garoupa (2015) and Carmignani and Giacomelli (2010)]. Errors have been clustered at the regional level.

$$Jud.Cong_{reg,t} = Cte + \beta_1\, Complexity_{reg,t-1} + \sum_{k=1}^{K} \delta_k Control^{k}_{reg,t-1} + Jud.Cong_{reg,t-1} + \varepsilon_{reg,t} \quad (5)$$

Table 2 shows the results. As it was indicated, the analysis is carried out for the civil jurisdiction, which disciplines private contracts (between citizens and between companies). Specifically, we build the measurements of trial length for the executions. As a robustness check, in a set of estimates (1 to 3) we exclude family law conflicts and in estimates 4 to 6 we include the whole set of civil cases. The signs are as expected, and the impacts are significant for the relational complexity variables.

The results suggest that greater relational complexity (that is, the need to consult or understand a greater number of interconnected norms in order to make use of the law) makes the functioning of the courts (which basically have to apply the body of regulation to resolve a specific conflict) more difficult or slower. The work of the courts does not seem to be affected either by the number of rules *per se* or by the linguistic complexity of those rules, possibly suggesting that legal professionals are trained precisely to master these two dimensions.

Table 2: Judicial congestion (trial length). Panel data Fixed effects

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
	Civil (without family)			*Total civil*		
# Norms $_{t-1}$	-0.00619			-0.00833		
	(0.0161)			(0.0146)		
Legibility $_{t-1}$		-0.126			-0.0696	
		(0.246)			(0.229)	
# Links $_{t.1}$			0.0232*			0.0212**
			(0.0110)			(0.00996)
Judicial Cong. $_{t-1}$	0.690***	0.689***	0.684***	0.699***	0.699***	0.694***
	(0.0556)	(0.0550)	(0.0563)	(0.0564)	(0.0557)	(0.0569)
Constant	6.820	7.327	7.076	7.003	7.282	7.246
	(6.490)	(6.536)	(6.344)	(6.007)	(6.115)	(5.877)
Fixed effects						
Time	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Regional	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Controls (lawyers, productivity)	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Observations	*272*	*272*	*255*	255	255	255
R-squared	*0.807*	*0.807*	*0.810*	0.783	0.783	0.786

Robust (clustered) standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1
Source: Own elaboration

4 Robustness checks

We perform two alternative robustness checks: first we consider two alternative measures of linguistic complexity: entropy and percentage of non-frequent distinct word (type) in the adopted norms; secondly, we modified the original database to only include rules with a minimum number of words, thus eliminating very short rules.

4.1. Additional measures of linguistic complexity

In this section we introduce two additional measures of linguistic complexity: entropy and share of not frequent words. These variables replace the μ linguistic complexity index.

4.1.1. Entropy

An additional measure of linguistic complexity is the "entropy" indicator proposed by Katz and Bommarito (2014) and Shannon (1951).[10] To build the entropy indicator we sum up the probability of occurrence of each distinct word, p_w, within the set of total words, W_n, of a norm, n, multiplied by their logarithm in base 2 of this probability (see equation 5). Entropy characterize the uncertainty or variance in a system[11], in this case the normative system. A minimum level of entropy is necessary to have information within the text, but it is considered that a greater value of entropy characterizes more difficult legal texts which include a greater variation of concepts.

$$Entropy_{\ n} = -\sum_{w \in W_n} p_w \, log_2(p_w) \qquad (5)$$

The final value for the entropy indicator, specific for each region and year, is the simple mean of the entropy indicator for each norm adopted in the region during the year, $Entropy_{reg,t} = \overline{Entropy_{n \in (reg,t)}}$.

4.1.2. Share of non-frequent words

We also construct a new index of similarity between the vocabulary used in the legal texts and that most frequently used in the Spanish language according to the Royal Academy of the Spanish Language[12]. We obtain the percentage of words used in the legal texts which do not appear among the 10.000 more frequently used words in the Spanish language[13]. In equation 6, the numerator counts the different words in a norm, Wd_n, that do not appear among the 10.000 most frequent words in Spanish, RAE. The denominator counts the total number of different words in that text. On average, between 55% and 60% of the words in the regional legal corpus do not belong to the list of 10.000 used words. The greater the share of infrequent words the lower the legibility. This is thus an inversed index of the simplicity of the vocabulary used in the regulatory framework.

[10] Another work following this line of research (i.e. quantitative analysis of the law) is Friedrich et al. (2020), who computed entropy indicators based on the written text of opinions published by the U.S. Supreme Court and the German Bundesgerichtshof. As indicated, one advantage of our analysis is that we are not forced to compare different languages since all our sources are officially in Spanish.

[11] For texts with a number of words such as those considered in this document, the first parenthesis tends to one and, therefore, the formula is the inverse of Pearson's coefficient of variation, which indicates the relationship between the standard deviation of a sample and its mean, all divided by the standard variation.

[12] Information obtained from Current Spanish Reference Corpus, CREA, Real Academia Española - RAE.es

[13] We also used the percentage of (unique) frequent words over the total number of distinct words used in the text. Results do not vary. Results are available on request.

$$\text{Freq}_n = 100 \cdot \frac{\sum_{Wd_n \notin RAE} Wd_n}{\sum Wd_n} \quad (6)$$

The regional range of variation and the correlations between variables are presented in Figure 5 and Figure 6 in the Appendix. The entropy indicator is negatively correlated with the legibility indicator μ discussed above (-0.65). The share of *infrequent* words in the Spanish language has a low negative correlation with the entropy indicator (-0.47) and a positive correlation with legibility (0.82).

We again analyzed the impacts of regulatory complexity on productivity and legal costs using these new approaches to complexity. Table 3 presents the estimations: Columns 1 and 2 analyze the impacts on productivity. Columns 3 and 4 discuss the implications for legal costs. As has been done so far, the different dimensions of complexity are incorporated independently. In the case of productivity growth, the entropy indicator is significant and with the expected negative sign. In the case of judicial congestion (trial length) the share of infrequent words is significant with the expected positive value.

Table 3: Robustness alternative indicators of legibility. Panel data Fixed effects

	(1)	(2)		(3)	(4)
	Dep var: Productivity			Dep var: judicial congestion (civil, without family)	
Entropy $_t$	*-0.0508**		*Entropy* $_{t-1}$	*0.405**	
	(0.0263)			*(0.198)*	
Share unique $_t$		*-0.0319*	*Share unique* $_{t-1}$		*0.292***
		(0.0245)			*(0.125)*
Productivity $_{t-1}$	*0.949****	*0.949****	*Jud Cong* $_{t-1}$	*0.688****	*0.683****
	(0.0149)	*(0.0147)*		*(0.0565)*	*(0.0566)*
Constant	*0.871****	*0.663****	*Constant*	*4.185*	*6.138*
	(0.214)	*(0.138)*		*(6.445)*	*(6.490)*
Fixed effects			*Fixed effects*		
Time	*Yes*	*Yes*	*Time*	*Yes*	*Yes*
Regional	*Yes*	*Yes*	*Regional*	*Yes*	*Yes*
			Lawyers + Prod t-1	*Yes*	*Yes*
Observations	*583*	*583*	*Observations*	255	255
R-squared	*0.985*	*0.985*	*R-squared*	0.810	0.810

Robust (clustered) standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1
Source: own elaboration

4.2. Different data sets

Finally, as an additional robustness check, we have worked with a modified database: we have only taken into account the norms of a certain length (excluding therefore the shorter ones). In the first place, we included only the regulations with more than 200 words and, secondly, the regulations with more than 1000 words.

The reason for making these changes is that some regulations with low economic or legal content have the force of law, such as those that just modify a ministerial or department structure. Those norms are usually short. When we select longer regulations, we are removing 197 and 2029 norms respectively form the original data set. Table 4 presents the results in terms of the relationship between complexity and productivity (or legal costs) using equation (3) for productivity growth (see columns 1 and 2). Columns 3 and 4 present estimated parameters following equation (5) for judicial congestion (trial length).

Our variables of regulatory complexity always have the expected sign. Legibility seems to have a significant impact on productivity, while the relational complexity measure seems to have impacts on the legal costs (judicial congestion). Taken altogether, these results show that it may be necessary to select only the regulations with the greatest economic and legal content in order to truly understand the impacts of complexity on productivity and legal costs.

Table 4: Robustness Dep. var.: Growth VA per hour.

	Productivity growth (1) Norms>200 words	*Productivity growth (2) Norms>1000 words*	*Judicial cong. (3) Norms>200 words*	*Judicial cong. (4) Norms>1000 words*
# Norms	*0.00226*	*0.00139*	*-0.00811*	*0.00432*
	(0.00159)	*(0.00157)*	*(0.0165)*	*(0.0196)*
Legibility	*0.0795***	*0.0710****	*-0.176*	*0.0692*
	(0.0313)	*(0.0210)*	*(0.255)*	*(0.469)*
# Links	*-0.00348**	*-0.00336*	*0.0232**	*0.0254***
	(0.00196)	*(0.00199)*	*(0.0110)*	*(0.0110)*
Fixed effects				
Time	*Yes*	*Yes*	*Yes*	*Yes*
Regional	*Yes*	*Yes*	*Yes*	*Yes*
Observations	*583*	*583*	*255*	*255*

Robust (clustered) standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1
Source: Authors' calculations

5 Concluding remarks

Regulation, as part of the institutional framework, matters for economic efficiency. While the objective of regulation is to reduce transaction costs and mitigate other market failures, its effects may be less beneficial or even counterproductive if the legal corpus is poorly designed. One of the reasons that can make regulation inefficient is its "complexity".

"Complexity" has been measured by some recent works in terms of "volumes" or quantities of regulations (for instance, the number of norms or pages). In this article we argue that complexity has additional dimensions such as legibility, that is the ease of reading the norms, and the relational structure of the norm, the average number of external regulations to which a certain legislative text refers. Natural language processing (NLP) techniques allow to develop objective indicators on those dimensions and to analyze their effects.

This article makes a first contribution to the literature consisting of the construction of a new database (RECOS, Regulatory Complexity in Spain) which includes information on the laws (and other norms with the force of Law) adopted by the Autonomous Regions since the beginning of the democratic period in Spain. We also build a set of indicators covering the new dimensions of complexity: legibility and network structure. Secondly, this research provides a first exploration of the effects of the new dimensions of complexity on various structural variables; labor productivity growth and judicial efficacy.

The results of our estimations show that the new dimensions of regulatory complexity are negatively related to productivity and judicial efficacy. That is, judicial efficacy and labor productivity growth seem to decrease when the norms are less legible and when the legal texts are more are more difficult to use because they require access to a greater number of regulations (due to their network structure). To the best of our knowledge, this is the first study showing the potential impacts of these new dimensions of regulatory "complexity" on different measures of economic efficiency.

Our research rationalizes the efforts of public administrations to achieve "better regulation". Indeed, some countries provide guidelines (or even create public bodies) that try to improve the quality of norms. For instance, the Government of Spain has initiatives in place to improve regulatory quality and to assess the need for new regulation (see

Royal Decree 931/2017 [14] and Mora-Sanguinetti, 2019a). Specifically, the quality of regional regulation should never be neglected. At the moment, nearly 70% of all Spanish regulations come from the regional level (see Mora-Sanguinetti and Pérez-Valls, 2020). As it was already highlighted, the regional analysis of the different dimensions of complexity of our research reveals substantial differences in regulatory complexity among the Spanish regions. This makes it possible to identify areas for improvement in the quality of regulation.

This article leaves open the normative question of what the optimal level of regulatory complexity is, taking into account the optimal level of economic growth. Futures avenues of research should also distinguish between pure economic legislation (labor markets, trade and retail, sectors regulation, etc.) and legislation on topics which *prima facie* may have only indirect effects on economic performance (prisons, health, etc.). Addressing the analysis of regulation by theme will also allow to clarify, among other things, if the increase in the regulatory power (increase in competences) of the regions drives the increase in the number of regulations. In this first analysis we decided to follow a more general approach, covering all norms. This strategy is robust to the possible criticism that we have produced a selection bias. This paper opens the road to use NLP to construct additional measures of complexity such us in-depth sentence structure (conditional, subordinate, etc.), vocabulary accuracy (use of vague terms) or analysis of the internal structure of the regulation.

[14] Royal Decree 931/2017, of 27 of October.

References

Acemoglu, D., and J. A. Robinson (2012). *Why nations fail: The origins of power, prosperity, and poverty*, Crown Books.

Al-Ubaydli, O., and P. A. McLaughlin (2017). "Regdata: A numerical database on industry-specific regulations for all United States industries and federal regulations, 1997–2012", *Regulation and Governance*, 11(1), pp. 109-123.

Azqueta-Gavaldón, A., D. Hirschbühl, L. Onorante and L. Saiz (2020). *Economic Policy Uncertainty in the Euro Area: An Unsupervised Machine Learning Approach*, ECB Working Papers, 2359.

Bailey, J. B., and D. W. Thomas (2017). "Regulating away competition: The effect of regulation on entrepreneurship and employment", *Journal of Regulatory Economics*, 52(3), pp. 237-254.

Baker, S. R., N. Bloom and S. J. Davis (2016). "Measuring economic policy uncertainty", *The Quarterly Journal of Economics,* 131(4), pp. 1593-1636.

Bardhan, P. (2002). "Decentralization of Governance and Development", *Journal of Economic Perspectives,* 16(4), pp. 185-205.

Bertrand, M., and F. Kramarz (2002). "Does entry regulation hinder job creation? Evidence from the French retail industry", *The Quarterly Journal of Economics*, 117(4), pp. 1369-1413.

Brelsford, K. M., E. Ruiz and L. Beskow (2018). "Developing informed consent materials for non-English-speaking participants: An analysis of four professional firm translations from English to Spanish", *Clinical Trials*, 15(6), pp. 557-566.

Calomiris, C. W., H. Mamaysky and R. Yang (2020). *Measuring the Cost of Regulation: A Text-Based Approach,* NBER Working Papers, 26856.

Campillo, J., A. Morato, I. Maqueda and S. Sánchez-Cuadrado (2020). *Readability of Spanish e-government information*, 15th Iberian Conference on Information Systems and Technologies (CISTI), Seville (Spain), pp. 1-4. DOI: 10.23919/CISTI49556.2020.9141000.

Carmignani, A., and S. Giacomelli (2010). *Too many lawyers? Litigation in Italian civil courts*, Temi di discussione, 745, Banca d´Italia.

CEPEJ [European Commission for the Efficiency of Justice] (2016). *European Judicial Systems Efficiency and Quality of Justice*, CEPEJ Studies, No. 23. http://www.coe. int/t/dghl/cooperation/cepej/evaluation/default_en.asp.

Chambers, D., C. A. Collins and A. Krause (2019a). "How do federal regulations affect consumer prices? An analysis of the regressive effects of regulation", *Public Choice,* 180(1-2), pp. 57-90.

Chambers, D., P. A. McLaughlin and L. Stanley (2019b). "Regulation and poverty: an empirical examination of the relationship between the incidence of federal regulation and the occurrence of poverty across the US states", *Public Choice*, 180(1-2), pp. 131-144.

Coenen, G., M. Ehrmann, G. Gaballo, P. Hoffmann, A. Nakov, S. Nardelli, E. Persson and G. Strasser (2017). *Communication of monetary policy in unconventional times*, ECB Working Papers, 2080, June.

Coffey, B., P. A. McLaughlin, and P. Peretto (2020). "The cumulative cost of regulations", *Review of Economic Dynamics*, 38, pp. 1-21.

Davis, S. J. (2017). *Regulatory Complexity and Policy Uncertainty: Headwinds of Our Own Making*, Becker Friedman Institute for Research in Economics Working Papers, No. 2723980.

De la Fuente, A. (2019). "Series largas de algunos agregados económicos y demográficos regionales", *Estudios sobre la economía española* (2019/19), Fedea.

Dejuán, D., and J. S. Mora-Sanguinetti (2021). "Which legal procedure affects business investment most, and which companies are most sensitive? Evidence from microdata", *Economic Modelling,* 94, pp. 201-220.

Di Vita, G. (2018). "Institutional quality and the growth rates of the Italian regions: The costs of regulatory complexity", *Papers in Regional Science*, 97(4), pp. 1057-1081.

Djankov, S., C. McLiesh and R. M. Ramalho (2006). "Regulation and growth", *Economics Letters*, 92(3), pp. 395-401.

Ehrmann, M., and J. Talmi (2020). "Starting from a blank page? Semantic similarity in central bank communication and market volatility", *Journal of Monetary Economics*, 111, pp. 48-62.

Fernández, J. (1959). "Medidas sencillas de lecturabilidad", *Consigna*, 214, pp. 29-32.

Flesch, R. (1948). "A new readability yardstick", *Journal of Applied Psychology*, 32, pp. 221-33.

Friedrich, R., M. Luzzatto and E. Ash (2020). "Entropy in Legal Language", in *Proceedings of the 2020 Natural Legal Language Processing (NLLP) Workshop*, 24 August 2020, San Diego, US. ACM, New York, NY, USA, 6 pp.

Ghirelli, C., M. Gil, J. J. Pérez and A. Urtasun (2019). "Measuring economic and economic policy uncertainty and their macroeconomic effects: the case of Spain", *Empirical Economics*, pp. 1-24. https://doi.org/10.1007/s00181-019-01772-8.

Haldane, A., and M. McMahon (2018). "Central bank communications and the general public", *AEA Papers and Proceedings,* 108, pp. 578-583.

Hampton, P. (2005). *Reducing administrative burdens*, London [Hampton Report].

Hansen, S., and M. McMahon (2016). "Shocking language: Understanding the macroeconomic effects of central bank communication", *Journal of International Economics*, 99, S114-S133.

Hansen, S., M. McMahon and A. Prat (2017). "Transparency and deliberation within the FOMC: a computational linguistics approach", *The Quarterly Journal of Economics,* 133 (2), pp. 801-870.

Harvie, D., G. Slater, B. Philp and D. Wheatley (2009). "Economic well-being and British regions: the problem with GDP per capita", *Review of Social Economy*, 67(4), pp. 483-505.

Hassan, T., S. Hollander, L. van Lent and A. Tahoun (2019). "Firm level political risk: Measurement and effects", *The Quarterly Journal of Economics,* 134(4), pp. 2135-2202.

Hernández-Murillo, R., and H. Shell (2014). "Rising complexity of the FOMC statement", *Economic Synopses,* 23(2), available at https://research.stlouisfed.org/publications/economic- synopses/2014/11/05/the- rising- complexity- of- the- fomc-statement/.

Hooghe, L., A. H. Schakel and G. Marks (2008). "Appendix B: Country and Regional Scores", *Regional & Federal Studies*, 18(2), pp. 259-274.

Hurka, S., and M. Haag (2019). "Policy complexity and legislative duration in the European Union", *European Union Politics,* 21(1), pp. 87-108.

Jansen, D. J. (2011). "Does the Clarity of Central Bank Communication Affect Volatility in Financial Markets? Evidence from Humphrey-Hawkins Testimonies", *Contemporary Economic Policy*, 29(4), pp. 494-509.

Katz, D. M., and M. J. Bommarito (2014). "Measuring the complexity of the law: the United States code", *Artificial intelligence and law*, 22(4), pp. 337-374.

Kirchner, S. (2012). "Federal legislative activism in Australia: a new approach to testing wagner's law", *Public Choice*, 153(3-4), pp. 375-392.

Laffont, J.-J., and J. Tirole (1993). *A Theory of Incentives in Procurement and Regulation*, The MIT Press, Cambridge, Massachusetts, USA.

López Guerra, L. (1994). *Introducción al Derecho Constitucional*, Tirant Lo Blanch, Valencia.

López Guerra, L., E. Espín, J. García Morillo, P. Pérez Tremps and M. Satrústegui (2018). *Derecho Constitucional Volumen II*, 11ª Edición, Tirant Lo Blanch, Valencia.

McLaughlin, P. A., and O. Sherouse (2019). "RegData 2.2: a panel dataset on US federal regulations", *Public Choice,* 180(1-2), pp. 43-55.

Mora-Sanguinetti, J. S. (2012). "Is judicial inefficacy increasing the weight of the house property market in Spain? Evidence at the local level", *SERIEs, Journal of the Spanish Economic Association,* 3(3), pp. 339-365.

Mora-Sanguinetti, J. S. (2019a). "La 'complejidad' de la regulación española ¿Cómo medirla? ¿Qué impacto económico tiene?", *Revista ICE*, 907, pp. 147-162.

Mora-Sanguinetti, J. S. (2019b). *Producción normativa, mercado de la abogacía y litigación en España*, Dossier, Diciembre/2019, Thomson Reuters.

Mora-Sanguinetti, J. S. (2021). "Justicia y economía: la eficiencia del sistema judicial en España y sus impactos económicos", *Papeles de Economía Española* (forthcoming).

Mora-Sanguinetti, J. S., and A. Fuentes (2012). *An Analysis of Productivity Performance in Spain Before and During the Crisis: Exploring the Role of Institutions*, OECD Economics Department Working Papers, No. 973.

Mora-Sanguinetti, J. S., and R. Pérez-Valls (2020). "How does regulatory complexity affect business demography? Evidence from Spain", *European Journal of Law & Economics*. DOI: 10.1007/s10657-020-09650-w.

Mora-Sanguinetti, J. S., and S. Salvador-Mora (2016). "El impacto económico de la actividad normativa. Los riesgos de la 'gran convergencia'", *Revista de Economía Industrial*, 398, pp. 117-128.

Mora-Sanguinetti, J. S., and N. Garoupa (2015). "Do lawyers induce litigation? Evidence from Spain, 2001-2010", *International Review of Law and Economics*, 44, pp. 29-41.

Mora-Sanguinetti, J. S., M. Martínez-Matute and M. García-Posada (2017). "Credit, crisis and contract enforcement: evidence from the Spanish loan market", *European Journal of Law & Economics,* 44(2), pp. 361-383.

Muñoz, M., and J. Muñoz (2006). *Legibilidad Mμ*, Viña del Mar, Chile.

Palumbo, G., G. Giupponi, L. Nunziata and J. S. Mora-Sanguinetti (2013). *The Economics of Civil Justice: New Cross-Country Data and Empirics*, OECD Economics Department Working Papers, No. 1060.

Radaelli, C. M. (2007). "Whither better regulation for the Lisbon agenda?", *Journal of European public policy*, 14(2), pp. 190-207.

Scarpetta, S., P. Hemmings, T. Tressel and J. Woo (2002). *The Role of Policy and Institutions for Productivity and Firm Dynamics: Evidence from Micro and Industry Data*, OECD Economics Department Working Papers, No. 329.

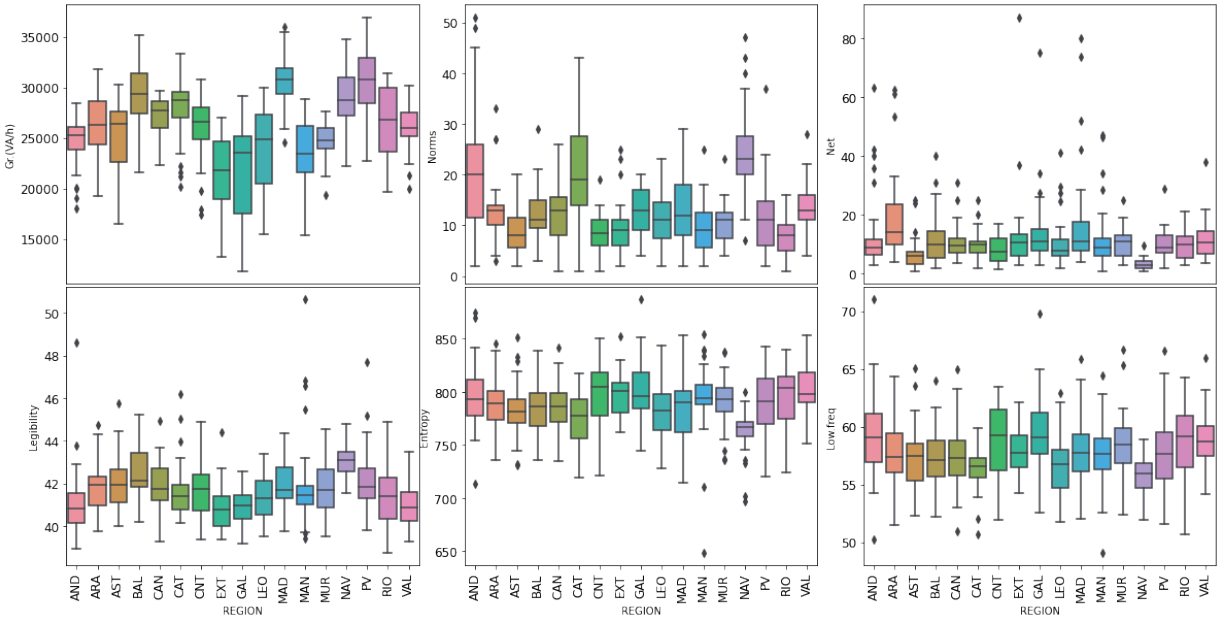
Shannon, C. E. (1951). "Prediction and entropy of printed English", *Bell System Technical Journal,* 30(1), pp. 50-64.

Tobback, E., S. Nardelli, and D. Martens (2017). *Between hawks and doves: measuring central bank communication,* ECB Working Papers, No. 2085.

Viviano, E. (2008). "Entry regulations and labour market outcomes: Evidence from the Italian retail trade sector", *Labour Economics,* 15(6), pp. 1200-1222.

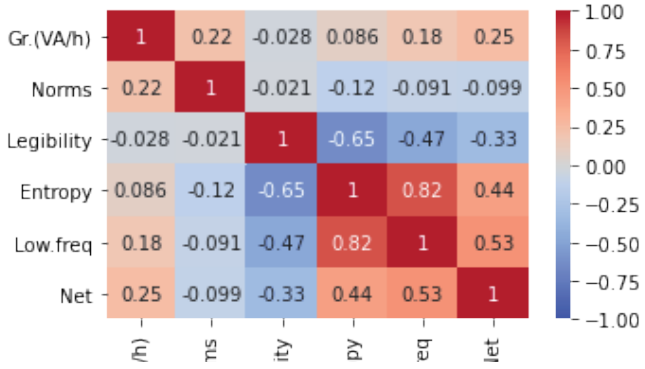
Appendix A

Figure 5: Complexity indicators. Regional distribution. 1978-2019



Source: Own elaboration.

Figure 6: Variable correlation matrix



BANCO DE ESPAÑA PUBLICATIONS

WORKING PAPERS

1940 MYROSLAV PIDKUYKO: Heterogeneous spillovers of housing credit policy.
1941 LAURA ÁLVAREZ ROMÁN and MIGUEL GARCÍA-POSADA GÓMEZ: Modelling regional housing prices in Spain.
1942 STÉPHANE DÉES and ALESSANDRO GALESI: The Global Financial Cycle and US monetary policy in an interconnected world.
1943 ANDRÉS EROSA and BEATRIZ GONZÁLEZ: Taxation and the life cycle of firms.
1944 MARIO ALLOZA, JESÚS GONZALO and CARLOS SANZ: Dynamic effects of persistent shocks.
1945 PABLO DE ANDRÉS, RICARDO GIMENO and RUTH MATEOS DE CABO: The gender gap in bank credit access.
1946 IRMA ALONSO and LUIS MOLINA: The SHERLOC: an EWS-based index of vulnerability for emerging economies.
1947 GERGELY GANICS, BARBARA ROSSI and TATEVIK SEKHPOSYAN: From Fixed-event to Fixed-horizon Density Forecasts: Obtaining Measures of Multi-horizon Uncertainty from Survey Density Forecasts.
1948 GERGELY GANICS and FLORENS ODENDAHL: Bayesian VAR Forecasts, Survey Information and Structural Change in the Euro Area.
2001 JAVIER ANDRÉS, PABLO BURRIEL and WENYI SHEN: Debt sustainability and fiscal space in a heterogeneous Monetary Union: normal times vs the zero lower bound.
2002 JUAN S. MORA-SANGUINETTI and RICARDO PÉREZ-VALLS: ¿Cómo afecta la complejidad de la regulación a la demografía empresarial? Evidencia para España.
2003 ALEJANDRO BUESA, FRANCISCO JAVIER POBLACIÓN GARCÍA and JAVIER TARANCÓN: Measuring the procyclicality of impairment accounting regimes: a comparison between IFRS 9 and US GAAP.
2004 HENRIQUE S. BASSO and JUAN F. JIMENO: From secular stagnation to robocalypse? Implications of demographic and technological changes.
2005 LEONARDO GAMBACORTA, SERGIO MAYORDOMO and JOSÉ MARÍA SERENA: Dollar borrowing, firm-characteristics, and FX-hedged funding opportunities.
2006 IRMA ALONSO ÁLVAREZ, VIRGINIA DI NINO and FABRIZIO VENDITTI: Strategic interactions and price dynamics in the global oil market.
2007 JORGE E. GALÁN: The benefits are at the tail: uncovering the impact of macroprudential policy on growth-at-risk.
2008 SVEN BLANK, MATHIAS HOFFMANN and MORITZ A. ROTH: Foreign direct investment and the equity home bias puzzle.
2009 AYMAN EL DAHRAWY SÁNCHEZ-ALBORNOZ and JACOPO TIMINI: Trade agreements and Latin American trade (creation and diversion) and welfare.
2010 ALFREDO GARCÍA-HIERNAUX, MARÍA T. GONZÁLEZ-PÉREZ and DAVID E. GUERRERO: Eurozone prices: a tale of convergence and divergence.
2011 ÁNGEL IVÁN MORENO BERNAL and CARLOS GONZÁLEZ PEDRAZ: Sentiment analysis of the Spanish Financial Stability Report. (There is a Spanish version of this edition with the same number).
2012 MARIAM CAMARERO, MARÍA DOLORES GADEA-RIVAS, ANA GÓMEZ-LOSCOS and CECILIO TAMARIT: External imbalances and recoveries.
2013 JESÚS FERNÁNDEZ-VILLAVERDE, SAMUEL HURTADO and GALO NUÑO: Financial frictions and the wealth distribution.
2014 RODRIGO BARBONE GONZALEZ, DMITRY KHAMETSHIN, JOSÉ-LUIS PEYDRÓ and ANDREA POLO: Hedger of last resort: evidence from Brazilian FX interventions, local credit, and global financial cycles.
2015 DANILO LEIVA-LEON, GABRIEL PEREZ-QUIROS and EYNO ROTS: Real-time weakness of the global economy: a first assessment of the coronavirus crisis.
2016 JAVIER ANDRÉS, ÓSCAR ARCE, JESÚS FERNÁNDEZ-VILLAVERDE and SAMUEL HURTADO: Deciphering the macroeconomic effects of internal devaluations in a monetary union.
2017 FERNANDO LÓPEZ-VICENTE, JACOPO TIMINI and NICOLA CORTINOVIS: Do trade agreements with labor provisions matter for emerging and developing economies' exports?
2018 EDDIE GERBA and DANILO LEIVA-LEON: Macro-financial interactions in a changing world.
2019 JAIME MARTÍNEZ-MARTÍN and ELENA RUSTICELLI: Keeping track of global trade in real time.
2020 VICTORIA IVASHINA, LUC LAEVEN and ENRIQUE MORAL-BENITO: Loan types and the bank lending channel.
2021 SERGIO MAYORDOMO, NICOLA PAVANINI and EMANUELE TARANTINO: The impact of alternative forms of bank consolidation on credit supply and financial stability.
2022 ALEX ARMAND, PEDRO CARNEIRO, FEDERICO TAGLIATI and YIMING XIA: Can subsidized employment tackle long-term unemployment? Experimental evidence from North Macedonia.

2023 JACOPO TIMINI and FRANCESCA VIANI: A highway across the Atlantic? Trade and welfare effects of the EU-Mercosur agreement.

2024 CORINNA GHIRELLI, JAVIER J. PÉREZ and ALBERTO URTASUN: Economic policy uncertainty in Latin America: measurement using Spanish newspapers and economic spillovers.

2025 MAR DELGADO-TÉLLEZ, ESTHER GORDO, IVÁN KATARYNIUK and JAVIER J. PÉREZ: The decline in public investment: "social dominance" or too-rigid fiscal rules?

2026 ELVIRA PRADES-ILLANES and PATROCINIO TELLO-CASAS: Spanish regions in Global Value Chains: How important? How different?

2027 PABLO AGUILAR, CORINNA GHIRELLI, MATÍAS PACCE and ALBERTO URTASUN: Can news help measure economic sentiment? An application in COVID-19 times.

2028 EDUARDO GUTIÉRREZ, ENRIQUE MORAL-BENITO, DANIEL OTO-PERALÍAS and ROBERTO RAMOS: The spatial distribution of population in Spain: an anomaly in European perspective.

2029 PABLO BURRIEL, CRISTINA CHECHERITA-WESTPHAL, PASCAL JACQUINOT, MATTHIAS SCHÖN and NIKOLAI STÄHLER: Economic consequences of high public debt: evidence from three large scale DSGE models.

2030 BEATRIZ GONZÁLEZ: Macroeconomics, Firm Dynamics and IPOs.

2031 BRINDUSA ANGHEL, NÚRIA RODRÍGUEZ-PLANAS and ANNA SANZ-DE-GALDEANO: Gender Equality and the Math Gender Gap.

2032 ANDRÉS ALONSO and JOSÉ MANUEL CARBÓ: Machine learning in credit risk: measuring the dilemma between prediction and supervisory cost.

2033 PILAR GARCÍA-PEREA, AITOR LACUESTA and PAU ROLDAN-BLANCO: Raising Markups to Survive: Small Spanish Firms during the Great Recession.

2034 MÁXIMO CAMACHO, MATÍAS PACCE and GABRIEL PÉREZ-QUIRÓS: Spillover Effects in International Business Cycles.

2035 ÁNGEL IVÁN MORENO and TERESA CAMINERO: Application of text mining to the analysis of climate-related disclosures.

2036 EFFROSYNI ADAMOPOULOU and ERNESTO VILLANUEVA: Wage determination and the bite of collective contracts in Italy and Spain: evidence from the metal working industry.

2037 MIKEL BEDAYO, GABRIEL JIMÉNEZ, JOSÉ-LUIS PEYDRÓ and RAQUEL VEGAS: Screening and Loan Origination Time: Lending Standards, Loan Defaults and Bank Failures.

2038 BRINDUSA ANGHEL, PILAR CUADRADO and FEDERICO TAGLIATI: Why cognitive test scores of Spanish adults are so low? The role of schooling and socioeconomic background

2039 CHRISTOPH ALBERT, ANDREA CAGGESE and BEATRIZ GONZÁLEZ: The Short- and Long-run Employment Impact of COVID-19 through the Effects of Real and Financial Shocks on New Firms.

2040 GABRIEL JIMÉNEZ, DAVID MARTÍNEZ-MIERA and JOSÉ-LUIS PEYDRÓ: Who Truly Bears (Bank) Taxes? Evidence from Only Shifting Statutory Incidence.

2041 FELIX HOLUB, LAURA HOSPIDO and ULRICH J. WAGNER: Urban air pollution and sick leaves: evidence from social security data.

2042 NÉLIDA DÍAZ SOBRINO, CORINNA GHIRELLI, SAMUEL HURTADO, JAVIER J. PÉREZ and ALBERTO URTASUN: The narrative about the economy as a shadow forecast: an analysis using Banco de España quarterly reports.

2043 NEZIH GUNER, JAVIER LÓPEZ-SEGOVIA and ROBERTO RAMOS: Reforming the individual income tax in Spain.

2101 DARÍO SERRANO-PUENTE: Optimal progressivity of personal income tax: a general equilibrium evaluation for Spain.

2102 SANDRA GARCÍA-URIBE, HANNES MUELLER and CARLOS SANZ: Economic uncertainty and divisive politics: evidence from the *Dos Españas*.

2103 IVÁN KATARYNIUK, VÍCTOR MORA-BAJÉN and JAVIER J. PÉREZ: EMU deepening and sovereign debt spreads: using political space to achieve policy space.

2104 DARÍO SERRANO-PUENTE: Are we moving towards an energy-efficient low-carbon economy? An input-output LMDI decomposition of CO_2 emissions for Spain and the EU28.

2105 ANDRÉS ALONSO and JOSÉ MANUEL CARBÓ: Understanding the performance of machine learning models to predict credit default: a novel approach for supervisory evaluation.

2106 JAVIER ANDRÉS, ÓSCAR ARCE and PABLO BURRIEL: Market polarization and the Phillips curve.

2107 JUAN de LUCIO and JUAN S. MORA-SANGUINETTI: New dimensions of regulatory complexity and their economic cost. An analysis using text mining.