MERCATUS WORKING PAPER

REGAUTHORITIES
THE REGULATIONS AUTHORITIES DATASET

Kofi Ampaabeng, Dustin Chambers, Ethan Greist, Patrick A. McLaughlin, Jonathan Nelson, and Thurston Powers

MERCATUS.ORG

MERCATUS CENTER
George Mason University

Kofi Ampaabeng, Dustin Chambers, Ethan Greist, Patrick A. McLaughlin, Jonathan Nelson, and Thurston Powers, "RegAuthorities: The Regulations Authorities Dataset," Mercatus Working Paper, Mercatus Center at George Mason University, Arlington, VA, July 2022.

Abstract

We introduce a new dataset that reveals the relationship between laws passed by the US Congress and regulations issued by federal regulatory agencies. This dataset, which we call the Regulatory Authorities Dataset (RegAuthorities), elucidates the scope of statutes as they manifest themselves as regulations, tracks the evolution of laws into tangible regulations, and quantifies the impact of statutory language on the volume of regulations.

JEL codes: C38, C55, K23, Z180

Keywords: administrative law, law, legal economic, legal institutions, regulation

Author Affiliation and Contact Information

Kofi Ampaabeng
Senior Research Fellow and Data Scientist
The Mercatus Center at George Mason University
kampaabeng@mercatus.gmu.edu

Dustin Chambers
Professor of Economics, Salisbury University
Senior Affiliated Scholar,
The Mercatus Center at George Mason University
dlchambers@salisbury.edu

Ethan Greist
Former Research Associate
The Mercatus Center at George Mason University

Patrick McLaughlin
Director of Policy Analytics and Senior Research Fellow
The Mercatus Center at George Mason University
pmclaughlin@mercatus.gmu.edu

Jonathan Nelson
Software Developer
The Mercatus Center at George Mason University
jnelson@mercatus.gmu.edu

Thurston Powers
Software Developer
The Mercatus Center at George Mason University
tpowers@mercatus.gmu.edu

Introduction

The *Code of Federal Regulations (CFR)* of the United States currently contains more than 100 million words, nearly tripling in size since 1970.[1] Because regulatory agencies in the United States promulgate and enforce regulations pursuant to mandates created by Congress, it is possible to map regulations back to their congressional origins. We describe such mapping with this new project, the Regulatory Authorities Dataset (RegAuthorities). Our intent is to measure the size and scope of federal laws by simultaneously examining both the statutes themselves and the regulations that are issued under those statutes' authority. Have agencies issued fewer or more regulations than expected? Do the linguistic characteristics of laws/statutes authorizing regulations influence the resulting regulatory output? What are the most influential statutes or laws in terms of regulatory output by agencies? Why do some statutes create more regulations than others? What are the statutes that offer broad authority to executive branch agencies to regulate? The dataset introduced in this paper provides interested researchers with the tools to answer these questions and more.

McLaughlin et al. show that the Dodd-Frank Wall Street Reform and Consumer Protection Act of 2010 (Pub. L. 111–203) produced more regulations than any other US law.[2] To arrive at this conclusion, McLaughlin et al. constructed a novel dataset, the Public Law Database (PLDB), which associates each CFR part with the specific statutes of the US Code (USC) that the CFR part cites as its authority. Using the PLDB as a starting point, the RegAuthorities dataset introduced in this paper systematically associates all statutes with their derivative regulations. Moreover, borrowing from the field of graph theory, this dataset introduces two

[1] Patrick A. McLaughlin, Oliver Sherouse, Mark Fabrizio, and M. Scott King, "Is Dodd-Frank the Biggest Law Ever?" *Journal of Financial Regulation* 7, no. 1 (March 2021): 149–74.

[2] McLaughlin, Sherouse, Fabrizio, and King, "Is Dodd-Frank the Biggest Law Ever."

new concepts useful to researchers: amplification and centrality. By *amplification* we mean the impact of a given unit of statutory law in terms of the amount of regulation it authorizes relative to its own verbosity. By *centrality* we mean the importance of a given unit of statutory law in terms of the number of regulations it authorizes regardless of its own verbosity. We discuss both concepts in detail in the methodology section.

The RegAuthorities dataset was created to better understand the link between law and regulation. It does this by connecting each part (unit of regulation) of the CFR to the sections of the USC it cites as rulemaking authority. These connections are then measured using network analysis techniques to approximate the regulatory impact of each section of the USC. Using these data, researchers can understand the patterns of regulatory authority and the characteristics of statutory law that result in different volumes of regulation.

The rest of this paper proceeds as follows. First, we describe the US regulatory process, starting from laws to regulations. We then highlight some of the possible uses of this dataset. Next, we describe the primary data sources for this dataset followed by a description of the methodology. We then present interesting insights derived from the data. Finally, we conclude and highlight some shortcomings of this paper and the methodology.

The US Regulatory Process

This dataset is focused on the connection between regulations and the laws that authorize them. We therefore begin by explaining the process by which laws are converted into regulations. Specifically, we discuss what regulations are, who writes them, from where they derive their authority, and how they are created.

Regulations (or rules, as they are alternatively known) are promulgated by regulatory agencies, who draft and enforce specific regulations based on the general authority granted to

them by acts of Congress. This notion of authority is crucial. When a regulatory agency issues a rule, that body must declare that Congress gave it the authority to write that rule. It does so by citing the section of the USC or the public law on which its authority to issue the rule in question is based. The RegAuthorities dataset traces these citations in the regulatory text back to the sections of the USC being cited by regulatory bodies. The specific mechanics of tracing authority's citations reflect the complexities of regulatory promulgation. There are three relevant processes involved in the promulgation: (1) the lawmaking process, (2) the law publication process, and (3) the regulatory rulemaking process. We will assume that most readers are familiar with the lawmaking process of the legislative branch.

The law publication process starts with an act of Congress. The act is initially published as either private or public law, depending on whether it affects only a small, specific group (private law) or the whole society (public law). All regulations ultimately derive their authority from public laws. At the end of each session of Congress, all the laws passed in that session are compiled into a volume of the *Statutes at Large*. Finally, all statutes pertaining to substantial public laws are eventually codified into the USC. Codification involves organizing the law by subject matter rather than date of passage and editing and formalizing the language of the law to conform to precise legal standards and terminology.[3]

The regulatory rulemaking process is governed by the Administrative Procedures Act, which describes the process for making, publishing, and enforcing rules. Generally, an agency first publishes the "proposed" rules in the *Federal Register (FR)*. These rules are then subject to public feedback, amendment, and review by other executive bodies before they are either rejected or published as "final" rules in the FR. From there, the rules are codified and published

[3] John V. Sullivan, *How Our Laws Are Made* (Washington, DC: US Government Printing Office, 2007), https://www.govinfo.gov/content/pkg/CDOC-110hdoc49/pdf/CDOC-110hdoc49.pdf.

in the CFR. The CFR is to the FR what the USC is to the statutes; it organizes the law by topic rather than passage date. There is one crucial difference that is key to understanding regulation: the rules in the CFR are based on the laws presented in the USC. As stated above, all parts of the CFR must cite their source of authority within the statutes passed by Congress. It is this mandatory citation of authority that we use to determine which sections of the USC are responsible for which parts of the CFR.[4]

Use Cases

The RegAuthorities dataset is a useful resource for various research topics. First, researchers can employ network analyses methods to examine the connection among regulations and statutes. This type of analyses could unearth relationships and patterns that are not readily visible from tabular data. Second, when combined with natural language programming and text analyses, researchers could examine the role of the language used in laws and statutes in regulatory output. Third, researchers can identify the most influential laws and statutes as they pertain to regulations. These are only a few of the possible uses of RegAuthorities, and we believe researchers will identify other creative ways to use this dataset.

This research is relatively unique, both within the field of network analysis and within the field of regulatory studies. When it comes to network analysis, our techniques are standard. Degree centrality is a straightforward measure of centrality that is well understood in the field of network analysis. However, the combination of degree centrality, node weighting, and total node weight ratios that forms our unique amplification centrality metric is less orthodox. Most weighted measures of degree centrality are modified by the size of links rather than the size of

[4] US Environmental Protection Agency, "The Basics of the Regulatory Process," accessed May 12, 2022, https://www.epa.gov/laws-regulations/basics-regulatory-process.

nodes. This may simply represent the reality that many of the networks that have been analyzed did not involve nodes with significant variability in their magnitude. Practicality suggests that node weighting is justified and even advantageous when modeling scenarios where the size of nodes significantly vary, and this conclusion is backed up by theoretical work.[5] Our decision to pay attention to node weights in our measure of amplification centrality reflects the fact that CFR parts and USC sections can vary greatly in their verbosity.

The usage of node weighting to calculate importance of a connection as the ratio between a node's weight and the weight of all the nodes immediately connected to it is, as far as we are aware, a new development in network analysis. This is likely because most network analysis is focused on the overall structure of a network. Weight ratios are largely irrelevant to network structure.[6]

Our analysis, however, is far more focused on the influence of each individual node than the structure of the entire network. The ratio of the weight between a node and the sum weight of its immediate connections is far more relevant in this context. In other words, the size of a given USC section divided by the size of the entire USC is far less important for understanding the impact of that section than is the size of that section divided by the size of the CFR parts that cite it. Traditional network analysis would not consider the impact of a given USC section on the CFR parts that cite it but would instead look at its position within the network of USC-CFR citations. However, because lines of authority from the CFR to USC are direct, we have paid special attention to the impact of individual USC sections. Ratios between node weights are consequently far more important than they would be in traditional network analysis.

---

[5] J. Heitzig, J. F. Donges, Y. Zou, N. Marwan, and J. Kurths, "Node-Weighted Measures for Complex Networks with Spatially Embedded, Sampled, or Differently Sized Nodes," *European Physical Journal B* 85 (2012): 38; and M. Wiedermann, J. F. Donges, J. Heitzig, and J. Kurths, "Node-Weighted Interacting Network Measures Improves the Representation of Real-World Complex Systems," *EPL* 102 (2013): 28007.

[6] The ratio between the weight of a node and the sum of the weights of all nodes in a network is simply the percentage of total network weight that the node represents. This information by itself does not reveal much about the structure of the network.

In the context of the regulatory studies literature, this dataset represents a new contribution to an old line of inquiry. The relationship between law and regulation has long been an area of interest to regulatory studies scholars. Scholarship in this vein often focuses on the formation of new "regulatory regimes" based on a sequence of acts of Congress within a defined era that either create a new set of regulatory agencies or alter the level of oversight or authority given to existing agencies (sometimes both). So far, such research has tended to be more general. When specific, it focuses on individual laws as case studies.[7]

The amplification dataset represents the first attempt we know of to systematically measure the regulatory impact of statutory law and quantify the regulatory impact of each unit of statutory law. It is certainly the only "regulatory law to statutory law" citation dataset we have seen. Some static indexes, such as the *Parallel Table of Authorities*, do exist but not as parsable data that contain information about the documents they index. Finally, the application of network analysis techniques to law has some precedent. Past researchers have measured the network of internal references within the USC[8] and the network of several types of legislation within the European Union.[9] We believe that this dataset will facilitate similar research as applied to US regulation.

Methodology

The RegAuthorities dataset lists statutes and the regulations that cite them. In addition, the dataset includes the language features of both the statutes and regulations. The reported language features use the standard tool kit from the Mercatus Center's RegData project, including the number of restrictions in the text, the total number of words, and measures of language complexity. We

[7] McLaughlin, Sherouse, Fabrizio, and King, "Is Dodd-Frank the Biggest Law Ever?"

[8] Alexander Lyte, David Slater, and Shaun Michel, "Network Measures of the United States Code," MITRE, July 2015, 7.

[9] M. Koniaris, I. Anagnostopoulos, and Y. Vassiliou, "Network Analysis in the Legal Domain: A Complex Model for European Union Legal Sources," *Journal of Complex Networks* 6 (2018): 243–68.

proceed by defining the two new metrics reported in the RegAuthorities dataset and briefly describe the other data elements from the QuantGov platform. We then discuss the primary data sources followed by methodology used to generate the final dataset.

Definitions

In this dataset, we report two important metrics: amplification and centrality. *Amplification* refers to the magnification of statutes into regulations. We compute this measure using the number of words in the authorizing statute and in the published regulations. The metric, amplification factor, is the ratio of words in the regulation to the words in the statute. For example, if a statute contains 100 words and the citing regulations also contain 100 words, the amplification factor for this statute would be 1. The larger the amplification factor, the more amplified the statute. If a statute is cited by multiple regulations, the numerator of this ratio consists of the sum of words in all the citing regulations. Formally, we can write the amplification factor metric as

$$\Gamma_{USC_{i,t}} = \frac{\Sigma W_{CFR_{i,t}}}{W_{USC_{i,t}}},$$

where for a section i in the CFR in year t, Γ_{USC} is the amplification factor for that section; W_{CFR} is the total number of words in a CFR part that references a USC section; and W_{USC} is the total number of words in the referenced USC section.

Centrality measures the importance of a section of the USC in published regulations. A large centrality value for a USC section means it has been used to authorize many regulations, and the reverse is also true. Graph theory is a subset of mathematics that examines the relationships between a series of objects. Within the graph theory literature, this measure is known as degree centrality and is defined as the number of links (citations) a given node (section of USC) possesses. Degree centrality alone, however, is conceptually too simplistic. When measuring the importance

of a section, we want to evaluate the size of the parts that cite it, not just their number. Therefore, our measure of centrality is "node weighted," ensuring that connections to larger nodes (defined as CFR parts with more words) are more important to our centrality metric.

These two measures, amplification and centrality, each tell a different story about authorizing statutes of regulations depending on the context. However, they are most informative when used together. In addition to determining the total words of regulation a given USC section is responsible for, it is also important to know how many words of regulation a given USC section has generated per word within that section. This final distinction is what makes amplification a relatively unique measure of centrality. Rather than identify the sections associated with the highest volume of regulation as most central, it identifies the sections associated with the highest volume of regulation *relative to their size* as most central. This allows for identifying the most potent and authoritative statutory language, which is our goal. We leave the task of using these data to estimate the dynamics between statutes and the regulations they spawn to future research.

In addition to these two metrics, the dataset also reports information from the Mercatus Center's QuantGov platform. QuantGov is a platform comprising machine learning and text analytic tools to generate quantitative data and insights from text data. The platform and its libraries can be applied to any selection of text to generate insights. In this dataset, we are interested in the following metrics: (1) the number of words, (2) the number of restrictions (i.e., the occurrence of the terms *shall*, *must*, *required*, *may not*, and *prohibited* in a body of text), (3) the complexity of the language, including Shannon entropy, (4) the average number of words in a sentence, and (5) the number of conditional clauses (i.e., the occurrence of phrases *if, unless*). Finally, we also include the federal agencies responsible for issuing the final regulations, enabling agency-level analysis of regulations.

Data Sources

We create the data using text from three main sources: the USC, the CFR, and public laws. Using the QuantGov platform, we extracted the text from the CFR and the USC from 1994 to 2018 and public laws from 1994 to 2020. The USC, the codification of permanent public laws of the United States, is divided into 53 titles that rarely change, each of which cover a broad topic. Generally, each title in the USC is further subdivided hierarchically into subtitle, chapter, subchapter, section, and subsection. Although some titles deviate from this convention, all titles have sections, and sections are how statutes are generally referenced. The CFR follows a similar hierarchy with 50 titles subdivided into subtitle, chapter, subchapter, part, subpart, and section.

For these data, and the QuantGov library in general, we need to define the unit of analysis. For practical purposes as well as by design, we consider a unit of regulation to be the CFR part. Similarly, we consider a unit of authority to be the USC section. The practical reason for choosing the CFR part as the unit of analysis for regulations and the USC section as the unit of analysis for statutes is that these two units are present for all titles, while all the other subdivisions are optional. Thus, for the rest of this paper, the term *CFR part* is our defined measure of a unit of regulation. Similarly, we refer to a unit of law or statute as a *section of the USC* or *USC section*.

Data Processing

For both the CFR and USC, we obtained the electronic versions stored in extensible markup language (XML) format.[10] XML uses tags to define properties of the data. Using these XML files, we extracted the metadata and the actual text from each USC section/CFR part. The metadata

[10] For more on the use of XML in documents, see https://www.loc.gov/preservation/digital/formats/fdd/fdd000075.shtml.

for the USC include the year, title, subtitle, chapter, subchapter, and section. Similarly, we processed the XML files for the CFR to obtain the statutory authority for each CFR part. Like the USC files, we extracted metadata and actual text from the XML-formatted CFR files. The metadata for the CFR include the CFR title number, part number, title name, part name, and the authority. The text contained in the authority tag of the CFR files is the main data for this project. We then applied tools from the QuantGov library to derive the language characteristics metrics—that is, the number of words, the number of restrictions, the quality of the text—and the agencies responsible for the regulations.

We retrieved the cited authorities from the CFR files for each available year. This was the most laborious part of the dataset creation. The CFR authorities contained in the XML files do not follow a consistent format for each CFR part. We therefore developed a series of algorithms (using regular expressions) to extract the cited authorities for each part. The four snippets below provide a few examples of the different ways in which agencies and the writers of the CFR indicate the authorities for their regulations:

> *Authority: (Secs. 2–4, Pub. L. 91–601, 84 Stat. 1670, 1671 (15 U.S.C. 1471–1473); sec. 701(a), 52 Stat. 1055 (21 U.S.C. 371(a)),(Secs. 2, 5, 7, 9, Pub. L. 91–601; 94 Stat. 1670–1674 (15 U.S.C. 1471, 1474, 1476, 1478); sec. 30(a), Pub. L. 92–573, 86 Stat. 1231 (15 U.S.C. 2079(a))*
>
> *Authority:7 U.S.C. 1a, 2, 5, 6, 6a, 6b, 6c, 6d, 6e, 6f, 6g, 6h, 6i, 6k, 6l, 6m, 6n, 6o, 6p, 6r, 6s, 7, 7a-1, 7a-2, 7b, 7b-3, 8, 9, 10a, 12, 12a,12c, 13a, 13a-1, 16, 16a, 19, 21, 23, and 24 (2012).(Secs. 4g, 5, 42 Stat. 1000, 49 Stat. 1496; 7 U.S.C. 6g, 7; secs. 4g, 5, 8a; 7 U.S.C. 6g, 7, 12a),(Sec. 411, Pub. L. 93–463, 88 Stat. 1414; 7 U.S.C. 4a note)*
>
> *Authority: 5 U.S.C. 1008; 49 U.S.C. Chapters 401, 411, 413, 415, 417.*
>
> *Authority: 39 U.S.C. 5402; 42 U.S.C., 4321, 49 U.S.C. Subtitle I and Chapters 401, 411, 413, 415, 417, 419, 461, 463, 471.*

In each of these examples, our regular expressions were able to correctly identify the USC sections. If a citation is especially atypical, it is possible that our algorithm failed to extract the correct citation. However, our quality checks lead us to believe that these errors—including

missed citations and incorrectly formatted citations—are rather rare (less than 0.5 percent). We will continue to improve the algorithm over time with the goal of eliminating all errors.

Summary of Data

Below we provide a few highlights of the RegAuthorities data. First, we summarize the data and then identify the most important statutes, as reflected by our centrality and amplification metrics. For simplicity, we present data for only the 2019 CFR. The full dataset contains data for the years 2007–2019.

Characteristics of Cited Authorities

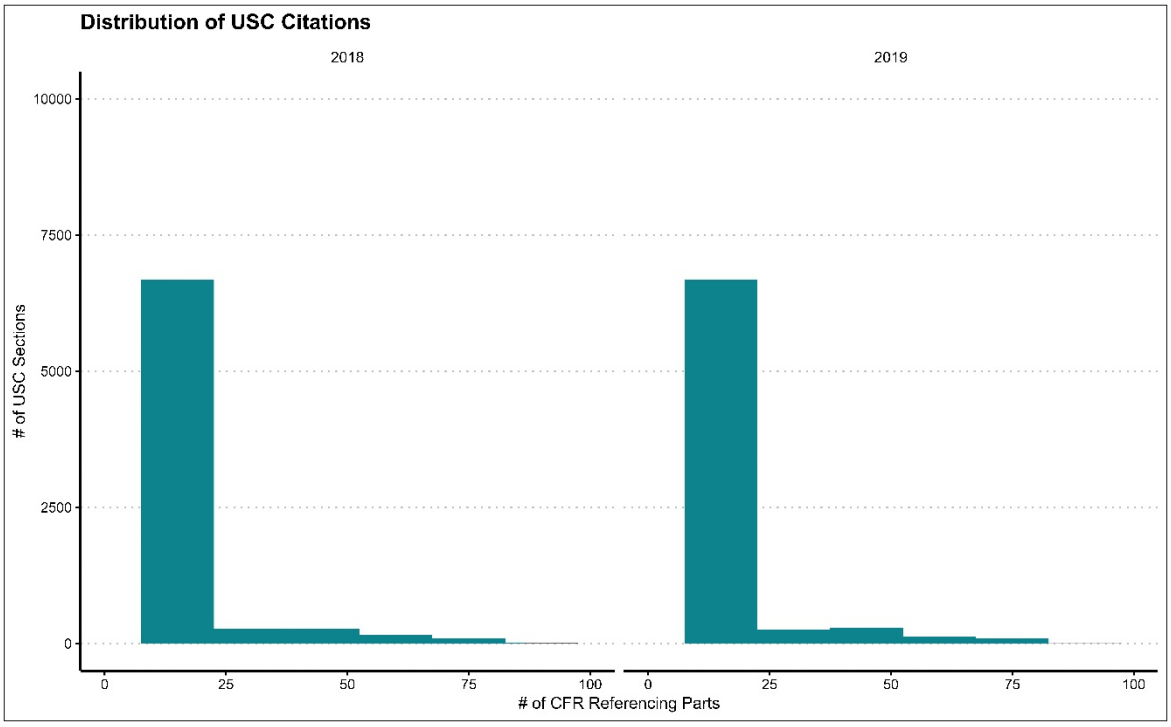
Table 1 (and figure 1) shows the summary of citations to the USC. From the 2018 (2019) edition of the USC, we identified 44,131 (44,419) sections. Of these, nearly half (48 percent) are not cited by any regulation, while 40 percent are the authority for fewer than ten regulations, suggesting that statutes generally have a narrow impact on regulations.

Table 1 Summary of CFR References to US Statutes

	# of Referencing Parts		% of USC Sections		Cumulative References	
# of References	2018	2019	2018	2019	2018	2019
None	23,096	23,358	47.99	48.24	23,096	23,358
Only 1	6,805	6,746	14.14	13.93	29,901	30,104
2–10	12,285	12,394	25.52	25.60	42,186	42,498
11–100	5,924	5,900	12.31	12.19	48,110	48,398
101–200	16	16	0.03	0.03	48,126	48,414
201–500	4	4	0.01	0.01	48,130	48,418
More than 500	1	1	0.00	0.00	48,131	48,419

Source: Authors' calculations from Ampaabeng et al., RegAuthorities: The Regulations Authorities Dataset, QuantGov, Mercatus Center at George Mason University, Arlington, VA, 2022.

Figure 1 Distribution of the Number of Citations



Source: Authors' calculations from Ampaabeng et al., RegAuthorities (dataset).

Centrality

Table 2 shows the 20 most frequently cited statutes by regulations for 2019 and 2018. In 2019 (2018), 5 USC 301 was the most referenced statute, with 856 (868) different regulations citing this section as its authority. This represents more than 10 percent of all regulations (there are 8,530 CFR parts as of 2019). In addition, 40 out of 50 CFR parts cite 5 USC 301 either in part or as their sole authority. 5 USC 301 reads as follows:

> The head of an Executive department or military department may prescribe regulations for the government of his department, the conduct of its employees, the distribution and performance of its business, and the custody, use, and preservation of its records, papers, and property. This section does not authorize withholding information from the public or limiting the availability of records to the public.

Table 2 Most Frequently Cited USC Sections

				Referencing CFR Parts		Referencing CFR Words	
USC Section	Title	Chapter Title	Section Title	2018	2019	2018	2019
5 USC 301	Government Organization and Employees	Powers	Departmental regulations	868	856	16,672,866	7,675,511
5 USC 552	Government Organization and Employees	Administrative Procedure	Public information; agency rules, opinions, orders, records, and proceedings	303	295	13,767,256	4,576,076
21 USC 371	Food and Drugs	Federal Food, Drug, and Cosmetic Act	Regulations and hearings	224	224	2,405,120	2,454,861
40 USC 121	Public Buildings, Property, and Works	General	Administrative	206	211	1,311,546	1,328,752
42 USC 3535	Public Health and Welfare	Department of Housing and Urban Development	Administrative provisions	201	201	2,676,807	2,676,467
28 USC 2461	Judiciary and Judicial Procedure	Fines, Penalties, and Forfeitures	Mode of recovery	162	163	3,476,221	3,489,519
42 USC 7101	The Public Health and Welfare	Department of Energy	Definitions	159	159	2,212,128	2,223,208
5 USC 552a	Government Organization and Employees	Administrative Procedure	Records maintained on individuals	156	157	2,139,154	2,266,002
21 USC 321	Food and Drugs	Federal Food, Drug, and Cosmetic Act	Definitions, generally	152	152	1,877,385	1,883,580
31 USC 9701	Money and Finance	Miscellaneous	Fees and charges for Government services and things of value	142	142	3,085,369	3,085,913
42 USC 2201	The Public Health and Welfare	Development and Control of Atomic Energy	General duties of Commission	128	127	2,533,377	2,522,556
46 USC 3306	Shipping	Inspections Generally	Regulations	120	120	1,274,068	1,274,128
8 USC 1103	Aliens and Nationality	Immigration and Nationality	Powers and duties of the Secretary, the Under Secretary, and the Attorney General	120	121	987,801	1,011,034

				Referencing CFR Parts		Referencing CFR Words	
USC Section	Title	Chapter Title	Section Title	2018	2019	2018	2019
30 USC 1201	Mineral and Lands Mining	Surface Mining Control and Reclamation	Congressional findings	115	115	478,676	479,456
29 USC 794	Labor	Vocational Rehabilitation and Other Rehabilitation Services	Nondiscrimination under Federal grants and programs	114	111	1,362,824	1,333,722
21 USC 351	Food and Drugs	Federal Food, Drug, and Cosmetic Act	Adulterated drugs and devices	112	112	1,070,598	1,084,646
5 USC 553	Government Organization and Employees	Administrative Procedure	Rule making	109	107	4,254,407	4,287,391
42 USC 1302	The Public Welfare	Social Security	Rules and regulations; impact analyses of Medicare and Medicaid rules and regulations on small rural hospitals	108	110	2,786,264	2,861,364
21 USC 352	Food and Drugs	Federal Food, Drug, and Cosmetic Act	Misbranded drugs and devices	107	107	939,672	940,767
51 USC 20113	National and Commercial Space Programs	National Aeronautics and Space Program	Powers of the Administration in performance of functions	107	106	1,225,355	1,228,381

Source: Ampaabeng et al., RegAuthorities (dataset).

It is therefore not surprising that this statute is the source from which most regulations derive their authority.

Amplification

Table 3 contains the same columns as table 2 but reports the most amplified statutes. 26 USC 1502 is the most amplified statute as of 2019 and 2018. It contains a mere 112 words, yet it is associated with 73 million words of regulatory text. As reflected in table 3, only one regulation (34 CFR 682; again, we define regulation as a single CFR part) cites 26 USC 1502. However, the CFR part that references this statute—26 CFR 1—cites 280 different statutes as its authority.

Table 3 Most Amplified USC Sections

USC Section	USC Title	Amplification		# Referencing CFR Parts		# Referencing CFR Words		# of Words in USC	
		2018	2019	2018	2019	2018	2019	2018	2019
16 USC 1543	Commerce and Trade	180,430.72	181,066.25	19	20	6,495,506	6,518,385	36	36
16 USC 4201	Commerce and Trade	498,215.38	498,881.85	7	7	6,476,800	6,485,464	13	13
16 USC 4203	Commerce and Trade	157,970.73	158,182.05	7	7	6,476,800	6,485,464	41	41
16 USC 4212	Commerce and Trade	150,623.26	150,824.74	7	7	6,476,800	6,485,464	43	43
16 USC 4241	Commerce and Trade	177,944.38	178,188.90	8	8	12,989,940	13,007,790	73	73
26 USC 107	Internal Revenue Code	220,135.33	214,279.25	1	1	18,271,232	17,785,178	83	83
26 USC 109	Internal Revenue Code	468,493.13	456,030.21	1	1	18,271,232	17,785,178	39	39
26 USC 115	Internal Revenue Code	338,356.15	329,355.15	1	1	18,271,232	17,785,178	54	54
26 USC 1502	Internal Revenue Code	652,544.00	635,184.93	1	1	73,084,928	71,140,712	112	112
26 USC 161	Internal Revenue Code	253,767.11	247,016.36	1	1	9,135,616	8,892,589	36	36
26 USC 211	Internal Revenue Code	254,016.89	247,266.14	2	2	9,144,608	8,901,581	36	36
26 USC 6001	Internal Revenue Code	270,646.77	263,458.53	10	10	37,078,608	36,093,818	137	137
26 USC 6065	Internal Revenue Code	207,065.28	201,781.09	14	14	9,525,003	9,281,930	46	46
26 USC 701	Internal Revenue Code	522,035.20	508,147.94	1	1	18,271,232	17,785,178	35	35
26 USC 723	Internal Revenue Code	186,441.14	181,481.41	1	1	9,135,616	8,892,589	49	49
26 USC 741	Internal Revenue Code	157,510.62	153,320.50	1	1	9,135,616	8,892,589	58	58
26 USC 742	Internal Revenue Code	351,369.85	342,022.65	1	1	9,135,616	8,892,589	26	26
26 USC 753	Internal Revenue Code	285,488.00	277,893.41	1	1	9,135,616	8,892,589	32	32
26 USC 7805	Internal Revenue Code	225,416.22	219,509.79	87	84	140,434,302	136,754,600	623	623
42 USC 7574	The Public Health and Welfare	182,964.88	184,064.08	41	41	4,391,157	4,417,538	24	24

Source: Ampaabeng et al., RegAuthorities (dataset).

The large amplification factor of 26 USC 1502 is therefore a consequence of the fact that it only contains 112 words but is cited by a single incredibly lengthy regulation.

Discussion

What are the implications of centrality and amplification? From table 2, it is clear that the sections with the highest centrality measures deal with four main concerns: definitions, powers, rules, and procedures. It is therefore unsurprising that many regulations will derive their authorities from such sections of the USC. High-centrality statutes that do not pertain to these four concerns demand further analyses and discussion. For example, 21 USC 351 (Adulterated Drugs and Devices) is among the top 10 most referenced authorities—referenced by 112 CFR parts. It is possible that how the regulations are written would affect the degree of centrality. If the citing regulations are spread over multiple CFR parts—RegData's unit of regulation—having a high degree of centrality might not be very meaningful.

The degree of amplification presents more interesting questions for research. What are the characteristics of the most amplified statutes? Do they provide broad regulatory authority to an agency? Do the most amplified statutes pertain to highly specialized topics such as taxes, environmental protection, or transportation? The most amplified section (26 USC 1502) pertains to income taxes, and it is unsurprising that the agency charged with creating the regulations for the statutes would issue many regulations. The 20 most amplified statutes pertain to just three titles in the USC: Internal Revenue (26), Public Health (42), and Commerce (16).

Challenges

Although the authority citations in the CFR are informationally rich and detailed, the unstandardized nature of the citations presents some challenges—namely, some citations may be erroneous, overly broad, or missing entirely. The problem of overly broad citations is not common but encountered with noticeable frequency in CFR data. Normally, a USC section is

the appropriate unit to cite as a statutory authority for regulation. Sections are usually specific enough that citing one as a source of authority is relatively precise. Unfortunately, some CFR parts cite entire USC chapters rather than specific sections. The chapter is the unit of text hierarchically one level up (or sometimes more) from the section. Chapters frequently include hundreds or even thousands of individual sections that span a wide variety of topics. Not only are citations to entire chapters legally vague, but they also have the potential to skew the dataset.

To make comparison meaningful, the consistent use of units is important. A section is comparable to a section and a chapter is comparable to a chapter, but a chapter is not comparable to a section. Any full chapter citation would therefore have to be interpreted as a citation to each section within that chapter. Such an interpretation makes precise analysis of the data difficult. Therefore, we have made the judgment that any full-chapter citation likely indicates an improper and overbroad citation made by the regulator. Consequently, we have removed from the dataset any CFR part that cites a full USC chapter. In total, these parts represent a small portion of the overall dataset and only number in the hundreds. They are provided in a separate dataset for those who wish to use them.

Conclusion and Next Steps

This paper introduces the RegAuthorities dataset, which quantifies both the relationship between laws and statutes, and the regulations they generate. Using these data, we demonstrate that relatively few sections of the US statutes generate a significant number of regulations. In addition, we show that many statutes fail to spawn any regulations. These findings underscore the value of the RegAuthorities dataset as a tool to identify outdated or irrelevant statutes. Like all data products from the Mercatus Center's Policy Analytics team, we intend to frequently update the RegAuthorities dataset. These updates include the analysis of new policy data as they become available and continuous improvements to the underlying algorithms.