



A NEW WAY TO QUANTIFY REGULATION

The Industry-specific Regulatory Constraint Database (IRCD)

PATRICK A. MCLAUGHLIN

Senior Research Fellow

WHAT IS IRCD?

The Industry-specific Regulatory Constraint Database (IRCD) is a new database that quantifies federal regulation. IRCD offers a novel and objective measure of the accumulation of regulations in the economy overall and for all the different industries in the U.S. IRCD uses text analysis to count the number of binding constraints in the text of federal regulations, which are codified in the *Code of Federal Regulations (CFR)*. In addition, it measures the degree to which different groups of regulations target specific industries.

WHY QUANTIFY REGULATION?

In *Moneyball*, Oakland A's general manager Billy Beane used an improved set of analytical gauges to create a better baseball team. He was only able to do so because Bill James and other pioneers of baseball statistics had created new gauges of player performance, such as the on-base plus slugging statistic, that improved upon the traditional gauges, such as batting average and subjective assessments by old-school scouts.

Regulations are similar to baseball players in the sense that some fail while others succeed. However, unlike baseball, there are not many objective measures of regulations. IRCD represents the first step in taking a *Moneyball* approach to studying regulation: if we can objectively measure regulations, we can use this information to help determine whether regulations efficiently achieve their intended goals.

QUANTIFYING REGULATIONS WITH TEXT ANALYSIS

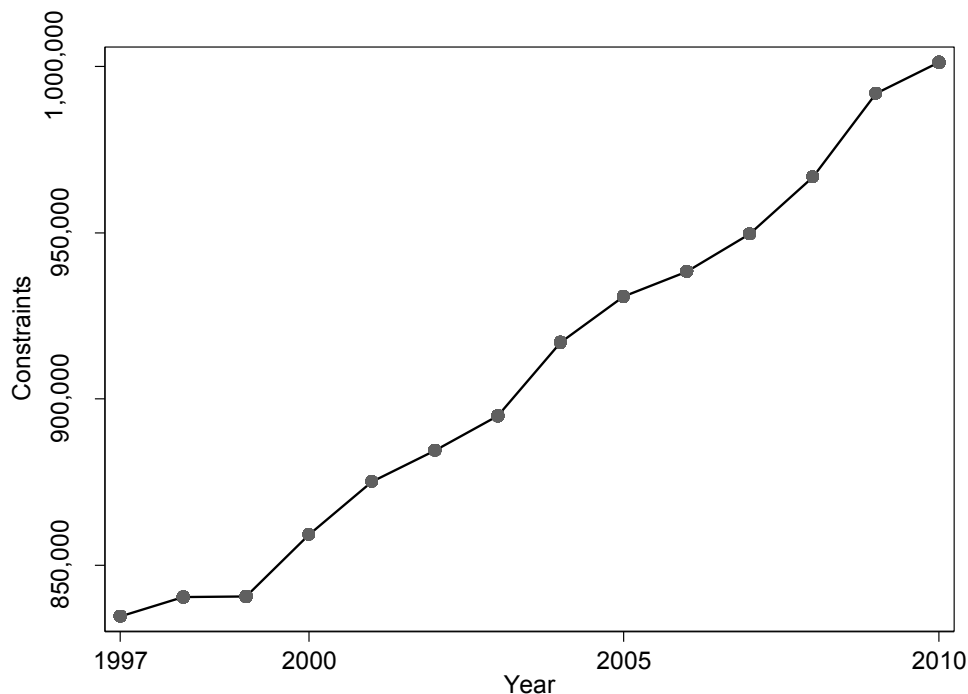
Previous efforts to assess the extent of regulation in the United States have used proxy variables designed to measure the quantity of federal regulation. For example, several studies count the number of pages published in the *Federal Register*, while others count the number of new rules promulgated annually. IRCD improves upon these studies in two principal ways. First, IRCD provides a novel measure that quantifies regulations based on the actual content of the *CFR*. Second, IRCD assesses the applicability of each *CFR* title to each industry. It uses the same industry classes as the North American Industrial Classification System (NAICS), which categorizes and describes each industry in the U.S. economy. The resulting database—IRCD—is the first industry-specific quantification of federal regulation, permitting within-industry and between-industry analyses of the causes and effects of federal regulations.

For more information or to meet with the scholars, contact
Robin Bowen, Regulatory Studies Program, (703) 993-8582, rbowen5@gmu.edu
Mercatus Center, 3301 Fairfax Drive, 4th Floor, Arlington, VA 22201

The ideas presented in this document do not represent official positions of the Mercatus Center or George Mason University.

Not all regulations are equal in their effect on the economy. Similarly, one page of regulatory text is often quite different from another page in content and consequence. For these reasons, IRCD relies on the content of regulatory text itself as a data source. IRCD parses the *CFR* to count the number of binding constraints—words that indicate an obligation to comply, such as “shall” or “must”—published annually from 1997 to 2010. This is important because the actual code of requirements of some regulations is hundreds of pages long, while other regulations have only a few paragraphs of requirements. Figure 1 shows that the total number of binding constraints in the entire *CFR* rose from about 835,000 in 1997 to just over 1 million in 2010.

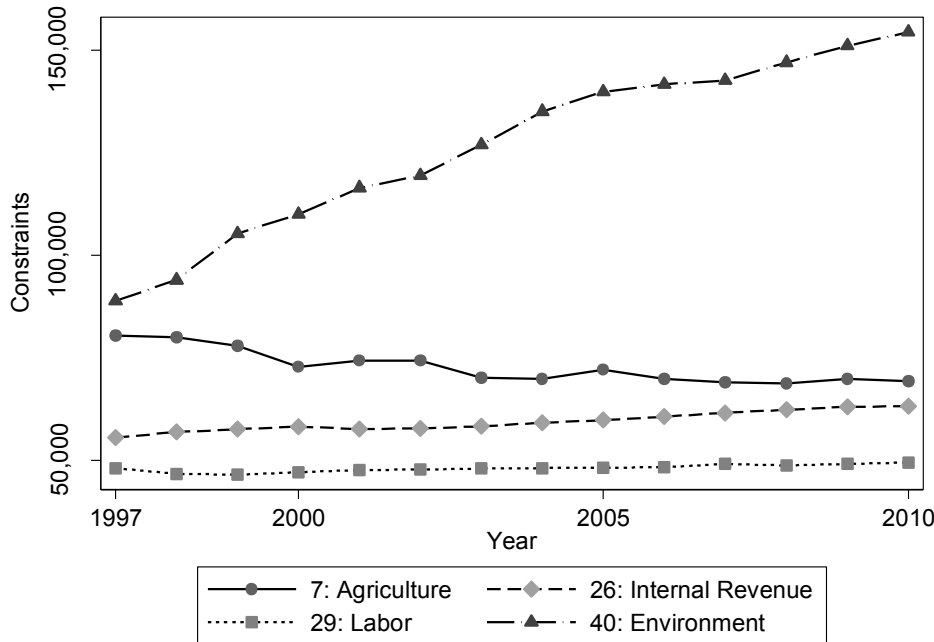
Figure 1: Total annual regulatory constraints, 1997–2010



QUANTIFYING BY CFR TITLE

Titles are arbitrary divisions of the *CFR*. In the time period examined, there were 50 numbered titles, each covering a broad subject area such as “Protection of the Environment” or “Labor.” By relying on the content of regulatory text as a data source, IRCD can separately parse each title and see how the growth of constraints differs across titles. Figure 2 shows the annual count of constraints for 4 of the 50 titles that comprise the *CFR*. These four titles were selected for this figure because they have the greatest number of constraints, on average, of any of the 50 titles. Figure 2 shows that there is substantial variation from title to title in how constraints accumulate over time. Environmental constraints, already the most numerous of any title in 1997, grew significantly through 2010, while agricultural constraints decreased slightly. Internal revenue and labor constraints grew over the time period, too, but at a much slower rate than environmental constraints.

Figure 2: Constraints of four selected CFR titles, 1997–2010



Similar graphs can be produced for each title using IRCD.

QUANTIFYING BY INDUSTRY

Another advantage of using text as data is the ability to flexibly assess which industries are targeted by regulations in each of the *CFR* titles. Typically, NAICS industry descriptions are simple and obvious, such as “chemical manufacturing” or “crop production.” Based on each NAICS industry description, we created various strings (combinations of words) that describe the industry (e.g., “crop producers” is one of several strings describing the “crop production” industry) and that can be used to gauge how heavily the industry is regulated in a particular *CFR* title or, ultimately, in a specific regulation.

We then developed a measure of how relevant each title is to specific industries based on the number of times these strings occur in each *CFR* title. The resulting dataset gives industry-specific measures of relevance—that is, measures of how heavily the regulations in a *CFR* title target a specific industry. By this measure, for example, the *CFR* titles that are the most relevant to the “oil and gas extraction” industry are “Title 18: Conservation of Power and Water Resources,” “Title 30: Mineral Resources,” and “Title 40: Protection of Environment.” By using this measure of the industry relevance of each *CFR* title along with the number of constraints in each title, one can create a measure of how regulated an industry is in in each year from 1997 to 2010.

There are several potential uses of a measure of how heavily regulated specific industries are. Both the causes and consequences of regulation are likely to differ from one industry to the next, and by quantifying regulations for all industries, scholars can test whether industry characteristics, such as dynamism, unionization, or a penchant for lobbying, are correlated with industry-specific regulation levels. The variety of industry-specific regulatory outcomes offered by IRCD permits researchers to compare effects across industries with greater statistical certainty. For example, if someone wanted to know whether high unionization rates are correlated with heavy regulation, he could compare our measure of industry-specific regulation for highly unionized industries to industries with little to no unionization.

COMBINING IRCD WITH OTHER DATASETS

IRCD's design allows users to easily combine regulatory data with many other datasets measuring possible causes and consequences of regulation. For example, the Bureau of Economic Analysis (BEA) produces data measuring annual GDP by industry, as well as several other measures of industry performance. Many BEA datasets on industry performance use the NAICS method to define industries, so it would be a simple matter to combine the two for statistical analysis. Thus, using IRCD data and GDP-by-industry data, one could test whether increases in regulation are correlated with decreases in overall industry output. Similarly, a researcher could use IRCD data together with the BEA's employment data by industry to see whether regulation leads to decreases in the number of jobs.

ABOUT THE MERCATUS CENTER

The Mercatus Center at George Mason University is a research, education, and outreach organization that works with scholars, policy experts, and government officials to connect academic learning and real-world practice.

The mission of Mercatus is to promote sound interdisciplinary research and application in the humane sciences that integrates theory and practice to produce solutions that advance in a sustainable way a free, prosperous, and civil society.

ABOUT THE AUTHOR

Patrick A. McLaughlin is a senior research fellow at the Mercatus Center at George Mason University. Before joining Mercatus, McLaughlin served as a senior economist at the Federal Railroad Administration in the U.S. Department of Transportation. He holds a Ph.D. in economics from Clemson University.