NBER WORKING PAPER SERIES

THE COST OF REGULATORY COMPLIANCE IN THE UNITED STATES

Francesco Trebbi
Miao Ben Zhang

Working Paper 30691
http://www.nber.org/papers/w30691

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2022

We would like to thank Matt Backus, Matilde Bombardini, Bård Harstad, John Matsusaka, Amit Seru, Reed Walker, seminar participants at the Berkeley Political Economy Research Lab and other institutions for their comments. Niranjan Navaneethan provided excellent research assistance. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Cost of Regulatory Compliance in the United States
Francesco Trebbi and Miao Ben Zhang
NBER Working Paper No. 30691
November 2022
JEL No. P0,P48

ABSTRACT

We quantify firms' compliance costs of regulation from 2002 to 2014 in terms of their labor input expenditure to comply with government rules, a primary component of regulatory compliance spending for large portions of the U.S. economy. Detailed establishment-level occupation data, in combination with occupation-specific task information, allow us to recover the share of an establishment's wage bill owing to employees engaged in regulatory compliance. Regulatory costs account on average for 1.34 percent of the total wage bill of a firm, but vary substantially across and within industries, and have increased over time. We investigate the returns to scale in regulatory compliance and find an inverted-U shape, with the percentage regulatory spending peaking for an establishment size of around 500 employees. Finally, we develop an instrumental variable methodology for decoupling the role of regulatory requirements from that of enforcement in driving firms' compliance costs.

Francesco Trebbi
Haas School of Business
University of California, Berkeley
2220 Piedmont Ave
Berkeley, CA 94720
and CIFAR
and also NBER
ftrebbi@berkeley.edu

Miao Ben Zhang
701 Exposition Blvd
HOH-722
Los Angeles, CA 90089
United States
miao.zhang@marshall.usc.edu

1 Introduction

Cost-benefit analysis is a key tool for the study and practice of federal and state regulation (Arrow et al., 1996). In particular, assessing the costs and the unintended consequences of government rules is key to provisions of the 1946 U.S. Administrative Procedure Act aimed at avoiding "capricious or arbitrary" federal agency behavior (Adler and Posner, 1999; Sunstein, 2002, 2017).[1] Policy makers and economists alike, however, have struggled to obtain systematic measures of regulatory costs.[2] For example, the Office of Information and Regulatory Affairs (OIRA) 2020 reports that only 9.1 percent of all "significant" regulations in 2019 had clearly quantified costs and benefits.[3]

The lacking assessment of regulatory costs may be unsurprising to economists, given the heterogeneity and the complex technical requirements of industry regulations (Davis, 2017). It is indeed an area impervious to systematization, as government rules range from limiting the composition of agricultural herbicides in wheat production to COVID-19 vaccine approvals, from setting capital adequacy ratios for bank holding companies to safety standards in iron ore mining, to name a few examples, and are therefore very hard to compare. To complicate things further, identical rules may be enforced with different levels of stringency (Stiglitz, 2009), exceptions may be carved out to shield specific producers, or the true extent of rules may be hard to trace when enacted via various forms of "regulatory guidance" (see Davis (2017) for regulatory "dark matter").

While these difficulties should prompt caution, political debate over the costs of government red tape often offers none of the nuance. In most cases, deliberation evolves as if regulatory costs have been unambiguously quantified and assessed against the benefits (Sunstein, 2017).

This paper aims at systematically measuring the labor costs of regulation for establishments and firms in the United States. Our measure is made possible by merging occupational task data from O*NET (V23.0) to the Occupational Employment and Wage Statistics (OEWS) establishment-occupation level microdata from the U.S. Bureau of Labor Statistics (BLS), a large stratified survey covering about 1,200,000 establishments from all industries of the United States from 2002 to 2014.

We start by measuring the regulation relatedness of a labor *task*. Using a combination of keyword matching, manual assignment, and natural language processing methods, we assess the regulation relatedness of each of the 19,636 tasks in O*NET. O*NET also provides an importance rating of each task for each occupation, which in turn we use to aggregate the regulation-relatedness

[1] For examples of regulatory rulemaking where this argument was applied, see Business Roundtable v. SEC (647 F3d 1144, DC Cir. 2011) and Corrosion Proof Fittings v. EPA (947 F2d 1201, 5th Cir. 1991).

[2] For a discussion centered on financial regulation, see Posner and Weyl (2013, 2017).

[3] When looking at the sample of all "significant" regulations (i.e. rules deemed likely to have an annual impact on the economy of $100 million or more), the OIRA states that "*During FY 2019, executive agencies promulgated 55 major rules, over half of which were transfer rules*" (p.3), but only "*for five rules, we report the issuing agencies' quantification and monetization of both benefits and costs*" (p.4). 9.1 percent of all significant rules in 2019 had clearly quantified costs and benefits.

of tasks at the *occupation* level.[4] We then obtain our key measure of regulation intensity for each establishment by aggregating occupations' regulation-relatedness weighted by the establishment's labor spending on each occupation. We label this establishment-level measure *RegIndex*, a variable that indicates the percentage of an establishment's total labor spending ascribed to performing regulation-related tasks. Sections 2 and 3 contain details about data and the construction of our measure, respectively.

Our establishment RegIndex measure allows us to measure regulatory compliance costs at any broader level.[5] In particular, we aggregate establishment RegIndex at the firm level, where a firm is defined by the establishment's employer identification number (EIN) (Song et al., 2018). An average firm spends 1.34 percent of its total labor costs on performing regulation-related tasks. When aggregating RegIndex to the national level, we observe that real regulatory compliance costs of U.S. businesses have grown by about 1 percent each year from 2002 to 2014.

As for any new measure, validation is necessary. Section 3 presents a battery of test results assessing the informativeness of this approach. We start with several difference-in-differences estimations to demonstrate RegIndex's ability to correctly reflect major industry-specific regulatory changes during our sample period. These case studies are selected to cover diverse industries and include the 2009 Card Act on credit card issuers, the 2005 Energy Policy Act and the 2011 President executive order on the oil and gas industry, the 2010 Affordable Care Act on hospitals, and directives from the Department of Education on sexual assault investigations for U.S. colleges. Through these case studies, we also highlight our new approach's advantages in distinguishing regulation from deregulation, picking up regulations from outside formal codification, and measuring businesses' regulatory costs at a granular level.

For further validation of RegIndex, we also show that the aggregate RegIndex correlates with the aggregate projected compliance hours that agencies reported to the OIRA at the national level. When looking at cross-state differences, we find that, conditional on the industry of analysis, Republican-leaning states have lower levels of regulatory compliance costs.

The paper then proceeds to address a key question pertinent to the study of regulation. Section 4 asks how regulatory costs change with respect to establishments' (or firms') size and investigates the presence of increasing or decreasing returns to scale in regulatory compliance. Intuitively, a regulatory compliance cost schedule that is non-neutral to scale – as approximated by total employment – may distort incentives for producers and induce factor misallocation.

This is a key question, because extant literature offers no consensus on the shape of returns to scale in regulation for the entire U.S. economy overall. On the one hand, regulatory costs

[4]Examples of occupations that are at the top of our list of regulation relatedness include compliance officers, construction and building inspectors, financial examiners, agricultural inspectors, and industrial engineering technicians.

[5]See also Simkovic and Zhang (2020) for an early attempt to measure regulation at the broad industry level using partial, publicly available BLS data.

increasing with size imply incentives for firms to remain small (Garicano et al., 2016). This may arise, for instance, from government policies designed to support small businesses through lighter regulation (e.g. regulatory tiering), which implies increasing compliance requirements kicking in at progressively higher levels of employment (Brock and Evans, 1985; Brock et al., 1986).[6] On the other hand, regulatory costs decreasing with scale favor larger players vis-à-vis smaller competitors, quashing entry and favoring concentration (Philippon, 2019; Callander et al., 2021; Cowgill et al., 2021). This may arise naturally from economies of scale in regulatory compliance, due to the presence of fixed costs[7] or it may derive from regulatory capture and special deals for large players.[8]

Our main finding in Section 4 is a non-monotonic relation between an establishment's RegIndex and the establishment's size. In particular, establishments under 500 employees experience diminishing returns to scale, with the percentage of labor spending on regulatory compliance increasing with employment. This is consistent with the Code of Federal Regulation (CFR) having elements of tiering[9] and also with smaller firms having lower incentives to comply to regulations, given a higher likelihood of flying under the regulator's radar. For establishments above 500 employees, increasing returns to scale kick in, and the percentage of labor spending on regulatory compliance progressively decreases with employment. On average, RegIndex for mid-size establishments is about 47% greater than the smallest establishments' and 18% greater than the largest establishments'.[10]

The economic implications of this inverted-U shape are both static and dynamic. Statically, our finding has implications for the "missing middle" of the establishment size distribution.[11] Productivity of establishments with an efficient scale below 500 employees could be depressed,

[6]The distortive effects of regulatory tiering are well documented, especially in Europe, where firms below efficient scale appear over-represented and mid-size firms under-represented. See among the many Brock and Evans (1985); Evans (1986); Boeri and Jimeno (2005); Schivardi and Torrini (2008); Gourio and Roys (2014); Garicano et al. (2016).

[7]Regulatory environments where increasing returns to scale prevail are frequently documented in the literature. For instance, in the case of environmental regulation, List et al. (2003) examine the effects of air quality regulation on new plant formation finding large negative impacts in New York State in 1980-90. In the case of pharmaceutical companies and the FDA, Thomas (1990) also finds large negative effects on small firms. In their analysis of compliance risk Davis (2017); Calomiris et al. (2020) find larger corporations being at an advantage (while focusing only on public companies).

[8]This is relevant with respect to classic research on the political economy of regulation and capture, including Stigler (1971); Posner (1974); Peltzman (1976), but also Shleifer and Vishny (1998). For a more recent discussion, see Lancieri et al. (2022).

[9]Examples of related statutes include the Regulatory Flexibility Act of 1980 and the Small Business Regulatory Enforcement Fairness Act of 1996. OSHA, DOE, EPA all implement regulatory flexibility reviews focused on small businesses.

[10]Further robustness analyses show that the inverted-U relationship holds even within occupations. In particular, using data on 14 million job posting descriptions, we show that for job postings of the same occupation, mid-sized firms require more regulatory compliance skills than small and large firms.

[11]The discussion on the missing middle for the firm size distribution focuses on developing countries (Hsieh and Olken, 2014; Tybout, 2014), yet it is unclear whether a counterfactual firm size distribution absent certain equilibrium distortions would not be heavier in the middle for the United States as well.

given the incentives to stay under the efficient scale of production due to an increasing red tape burden. Dynamically, we find suggestive evidence that a higher RegIndex is correlated with a more negative long-term change in the share of establishments across employment bins within a sector. We note here that this evidence is not causal and further analysis is needed.

Finally, Section 5 proposes and implements a shift-share instrumental variable method for understanding whether the relation between RegIndex and establishment size may be driven by the de jure regulatory requirements on different establishments or by regulatory agencies' heterogeneous enforcement effort. This final step is useful in order to decompose how important the design of the rules is vis-à-vis the endogenous inspection likelihood, as these may vary with establishment size due to differential enforcement and supervision imposed on producers of different size. Across the board, we find that regulatory requirements play a greater role than enforcement for establishments and firms of all sizes. This is true especially for large establishments and firms. We also find that regulatory requirements have systematically more precisely estimated effects, while the role of enforcement is less precisely estimated. Overall, the evidence in Section 5 seems to reject the hypothesis that the results in Section 4 are solely driven by regulators' differential enforcement on businesses at different percentiles of the size distribution.

This paper confines its scope to the costs of regulation without addressing the benefits of regulation.[12] We further limit the analysis to assessing the labor costs of regulatory compliance in terms of wages paid to workers engaged in complying with any rule or standard required by the government. Hence, our work does not address other types of regulatory compliance costs, such as capital expenditure (e.g. air cleaners and filters for restaurants, pumping or draining systems for mine water, regtech software, etc.) and foregone investment opportunities and profits due to compliance risk.[13] This notwithstanding, we argue, labor compliance costs are a key dimension of the question. According to survey estimates from the Securities Industry Association (2006) in the U.S. financial sector 93.9 percent of regulatory compliance costs are labor related and 3.3 percent are physical capital related. According to survey estimates from the National Association of Manufacturers (2014) pertinent to the U.S. manufacturing sector, 68.4 percent of regulatory compliance costs are labor related, 13.4 percent capital related. More broadly, we find that over 60 percent of federal regulations approved after 1980 are "information collection regulations", for which labor costs are the main form of direct regulatory compliance costs.[14] A third important

[12] For a criticism, see Sunstein (2020).

[13] See, for example, Ryan (2012) for a case where such omissions are of first order.

[14] From 1980 to 2019, OIRA reviewed and approved 44,985 federal regulations (see https://www.reginfo.gov/public/do/eoCountsSearchInit?action=init). Following (Kalmenovitz, 2019), we identify 28,012 regulations (or 62.3 percent of total regulations) are "information collection" (IC) in nature because the agencies filed for Form 83-I when proposing these regulations. Examples of IC regulations include annual reports to shareholders, confirmation of electronic transactions, product labeling, labor regulation, etc. A main measure of regulatory compliance costs for IC regulations is the estimated total hours the regulated entities need to spend on complying with the regulation in each year, e.g., 14.2 million hours annually for complying with corporate 10-K filing.

limitation/caveat of our study is that we do not include in our measure costs of compliance that are borne by firms through outsourcing (e.g. external compliance and accounting services). The aforementioned surveys show that spending on outsider advisers accounts for only 2.8 percent and 8.7 percent of total compliance costs for the financial sector and the manufacturing sector, respectively. Based on the data, this would not appear to be a damning omission.

This paper's methodology is a partial, but relevant step in the direction of a more reliable granular assessment of the economic costs of regulation directly borne by producers. There are six underlying reasons justifying this statement. First, our approach has the advantage of being a firm-specific micro measure, which, as the same regulation may affect different firms differently even within the same sector, is crucial to capture heterogeneity in the data. Second, our measure focuses on actual compliance costs paid by firms and not on statements, expectations, or projections, either by management or by government agencies. Third, our measure looks at on-equilibrium effects of regulation, incorporating the endogenous response by firms to enforcement, to regulatory ambiguity, etc. Fourth, our measure covers both very small and very large establishments – which is an extremely important feature in the context of assessing the effects of government rules on firm entry and small businesses. Fifth, our approach captures regulations from all relevant sources, including federal, state, local, and industry privately-enforced regulations. Sixth, our approach is operational and reproducible by regulatory agencies in their ongoing assessments and validation of their own rules. It can be used as a straightforward empirical methodology for the regulatory agency, complementing their ex ante projections.

While we are not aware of any alternative measure of regulatory costs with the exact properties of RegIndex, it is important to mention that other methods designed to obtain a comparative perspective of regulatory requirements have been proposed in the past, at least since Morrall (1986), and that many have important advantages of their own. For instance, Al-Ubaydli and McLaughlin (2017)'s RegData, a data repository maintained by the Mercatus Center at George Mason University, is built using QuantGov, a library of machine learning algorithms and text analysis tools designed to collect information about the number of restrictions, rule complexity and industry incidence from the text in the Code of Federal Regulations. Al-Ubaydli and McLaughlin (2017) focus on the linguistic structure of the rule and statute text (average sentence length, number of restrictive words, etc.), rather than on any direct economic consequence. In this sense, our approach is complementary to theirs. However, RegIndex is also much more accurate. In the analysis of Section 3, we show that our methodology provides a systematically superior fit of the data relative to RegData. RegData fails to pick up several major industry regulatory changes in the case studies, while our measure identifies them accurately.

In another novel application, Kalmenovitz (2021) employs text data from Form 83-I filed to OIRA to indicate the expectations of regulators about the cost of compliance of each regulation. The 83-I forms include estimates of how many responses the regulator will receive per year, how many work hours firms will be required to dedicate to comply with the regulation, and the estimated dollar costs of compliance. As this information is regulation-specific, the author needs to match each regulation to industries via a natural language processing (NLP) step and then, through information on the set of industries in which each firm operates, ultimately retrace regulatory costs to firms. Relative to Al-Ubaydli and McLaughlin (2017), the method in Kalmenovitz (2021) has the major benefit of aggregating the expectations of experts. It is, however, a measure of regulatory burden as expected by regulators – potentially different from the rules' actual incidence, and not ex-post verified. It is also a measure that cannot be directly traced to establishments.[15]

Davis (2017) and Calomiris et al. (2020) employ an original linguistic approach to measuring regulatory exposure for large publicly traded companies. Davis (2017) focuses on Part 1A of forms 10-K to gauge firms' exposure to regulatory and policy risk.[16] Calomiris et al. (2020) focus on the transcripts of corporate earnings calls made by publicly traded corporations and show that their measure of increasing regulation is predictive of sales growth, asset growth, leverage, and other measures of firm performance. An important advantage of the Davis (2017) and Calomiris et al. (2020) approaches relative to the one in this paper is that they are apt at capturing future regulatory risk, both in terms of discretionary enforcement and of new rules affecting firms.[17]

More broadly, this paper is related to a newly revived literature on regulation and its political economy. Part of this literature focuses on the role of firms in influencing regulation, via lobbying and political influence (Blanes i Vidal et al., 2012; Bertrand et al., 2014; Drutman, 2015; Kang, 2016; Bombardini and Trebbi, 2020; Bertrand et al., 2021). Another part explores the political and policy risks to which firms are exposed (Julio and Yook, 2012; Baker et al., 2014, 2016; Hassan et al., 2019).

The rest of the paper proceeds as follows. Section 2 describes the data on tasks and employment. Section 3 presents the evidence on the validation of our approach, including several event studies around regulatory tightening and loosening in different industries. Section 4 focuses on the analysis of the returns to scale. Section 5 lays out our methodology on separating endogenous enforcement from the structure of regulatory requirements that we cannot fully observe. Section 6 concludes.

[15]See OTA (1995) and Harrington et al. (2000) for many examples of regulations in which *ex ante* estimates of compliance costs differ from *ex post* compliance costs due to firms' response to regulations.

[16]This is also an approach preceded by Baker et al. (2016) and followed subsequently by Hassan et al. (2019), who focus on firm's exposure to a broader political risk, as expressed in the 10K-forms of publicly traded firms.

[17]Indeed, Calomiris et al. (2020) underscore compliance risk as the most relevant channel behind their finding, rather than physical compliance costs.

2 Data

Our main source of information is the establishment-occupation level microdata from 2002 to 2014 provided by the OEWS program of the BLS. This data set covers surveys that track employment and wage rates for over 800 detailed occupations in approximately 1.2 million establishments over the course of a three-year period. The sample of establishments covers, on average, 62 percent of the nonfarm employment in the U.S. Within a three-year period, 400,000 establishments are surveyed during each year and therefore the same establishment is surveyed at most every three years (e.g., in t and $t+3$). The microdata provides each establishment's sampling weight (to recover economy-wide aggregates), NAICS 6-digit industry code, county code, government ownership indicators, and parent firm's employer identification number (EIN).

Defining the appropriate regulatory compliance entity is nontrivial. A firm may be considered the ultimate entity, in that it pays the fines and penalties if inspected at any constituting establishment and found in violation of agency rules. However, regulation varies across industries and states, inducing a firm to spend different amounts of resources on regulatory compliance across different constituting establishments. For these reasons, we conduct our main analysis at both the establishment level and the firm level.

An EIN will define the boundary of a firm in our analysis. This is because the EIN identifies a firm for tax purposes, and also because the EIN is commonly used to define a firm both in the academic literature (Chodorow-Reich, 2014; Song et al., 2018) and by government agencies, such as the BLS.[18] Following the convention recommended by the OEWS program, we aggregate establishments of an EIN surveyed in years $t-2$ to t to represent the occupational composition of the EIN in year t.[19] If a firm has establishments spanning multiple industries, we define the firm's major industry as the NAICS 6-digit code with the highest employment share.[20]

Our research makes use of the task statements for each occupation from the O*Net (V23.0) database. Each task statement is a single sentence and is pertinent to a unique occupation. An occupation is described by an average of 22 different task statements. O*Net also provides an average rating indicating how important the task is for the occupation rated by incumbent workers working in the occupation, occupational experts, and analysts. Occupations are categorized at the 8-digit standard occupational code (SOC) level. We match the 8-digit SOC in O*Net to the 6-digit

[18]See U.S. Department of Labor, Bureau of Labor Statistics, "Business Employment Dynamics Size Class Data: Questions and Answers," http://www.bls.gov/bdm/sizeclassqanda.htm, questions 3 and 5. Chodorow-Reich (2014); Song et al. (2018) describe different issues of using EINs to measure the economic entity of a firm.

[19]The BLS OEWS program uses a similar aggregation approach to publicize industry or geographical level statistics for each year. See technical notes at https://www.bls.gov/oes/oes_doc_arch.htm.

[20]Because the OEWS survey does not cover all establishments of a firm, our firm-level measures contain measure errors and our firm-level employment is under-estimated. Given that over 80 percent of establishments in our sample are standalone single-unit firms (Ayyagari and Maksimovic, 2017), the establishment-level and firm-level results are qualitatively the same, as we will show.

SOC occupations in the OEWS microdata, creating a characterization of all tasks performed at the establishment level, which allows for the construction of our main measure as described in the following section.

3 Measure and Validation

In this section, we first present our methodology for measuring regulatory compliance costs for establishments and firms. We then discuss a battery of validation results for our main regulatory index.

3.1 A Simple Theoretical Framework

We assume that a firm's labor includes workers performing regulatory compliance related tasks and workers performing other production tasks. Let L_{it} be the total number of production workers employed in establishment i at time t and w_{it} the average wage paid to the workers. Define R_{it} the total number of workers occupied in regulatory compliance related tasks. We allow the average wage of compliance workers to differ from that of production workers (due to specialization) and indicate it with w^r_{it}. We assume all wages to be taken as given from the establishment's perspective. R_{it} can be derived as the result of optimization on the part of the establishment owners, who maximize profits:

$$Y_{it} - w_{it}L_{it} - w^r_{it}R_{it} - p_{it} \times f_{it}$$

with respect to R_{it} and L_{it}, where establishment i faces a probability of inspection p_{it} at period t and fines $f_{it} = \frac{\tilde{R}_{it}}{R_{it}}$ are levied in case the establishment is found in less than full compliance with regulatory requirements $\tilde{R}_{it}$. We further assume a constant return to scale production function $Y_{it} = \phi_{it}L_{it}$, where ϕ_{it} indicates an establishment-specific productivity shock, and that $\tilde{R}_{it} = \tilde{R}(L_{it}) = \rho L^{\alpha}_{it}$ and $p_{it} = \pi L^{\beta}_{it}$, where α, β, ρ, π are policy parameters governing regulatory requirements and enforcement effort targeting establishment i. Simple static profit maximization implies:

$$
\begin{aligned}
R^*_{it} &= \left(\frac{p_{it}\tilde{R}_{it}}{w^r_{it}}\right)^{\frac{1}{2}} \\
&= \left(\frac{\pi\rho\left(\frac{\phi_{it}-w_{it}}{(\alpha+\beta)\left(w^r_{it}\pi\rho\right)^{\frac{1}{2}}}\right)^{\frac{2(\alpha+\beta)}{\alpha+\beta-2}}}{w^r_{it}}\right)^{\frac{1}{2}} \\
L^*_{it} &= \left(\frac{\phi_{it}-w_{it}}{(\alpha+\beta)\left(w^r_{it}\pi\rho\right)^{\frac{1}{2}}}\right)^{\frac{2}{\alpha+\beta-2}}.
\end{aligned}
\tag{1}
$$

Note that, intuitively, equilibrium compliance is an increasing function of regulatory requirements $\tilde{R}_{it}$, an increasing function of the enforcement activity of regulators via inspections, p_{it}, and a decreasing function of regulatory compliance labor wages, w^r_{it}.

We define the index of regulatory compliance costs as:

$$
\text{RegIndex}_{it} = \frac{w^r_{it}R_{it}}{W_{it}}, \tag{2}
$$

that is, the share of the total wage bill $W_{it} = w^r_{it}R_{it} + w_{it}L_{it}$ allocated to compliance labor related to regulation.

3.2 Construction of the Regulation Index

The construction of the regulatory index, RegIndex, starts with identifying which tasks are related to regulatory compliance. We achieve this goal by analyzing the texts of task statements in the O*Net data in two phases: a keyword matching phase and an annotation phase.

Keyword Matching Phase

We identify regulation-related tasks by matching the task statements to two different tiers of keywords. Two tiers are necessary in order to balance the rate of false positives and false negatives in the identification of tasks. The first tier of keywords includes the words *regulation*, *regulations*, and *regulatory*. These matches intuitively identify as a "regulation-related" task whose statement explicitly mentions regulation. These matches exhibit a low rate of false positives. Table 1 lists ten examples of regulation-related tasks identified by the first-tier keywords.

The first tier of keywords produces, however, an excessive rate of false negatives. Regulation can be described by various alternative words and phrases, and regulatory compliance behavior can be described without directly mentioning *regulations*. For this reason we employ a second tier of keywords for identifying regulation-related tasks. The second tier of keywords includes (A) alternative references to regulation: *Law, Laws, Legislation, Legislative, Statute, Statutes, Statutory, Ordinance, Ordinances, Code, Codes, Standards, Public Policy, Public Policies, License, Licenses, Licensing, Permit, Permits, Certification, Certifications*; (B) references to government agencies: *Government, Governments, Governmental, Federal, Legislature, Policy Maker, Policy Makers, Governing Agency, Governing Agencies, Public Agency, Public Agencies*; and (C) actions of compliance: *Compliance, Noncompliance.* These matches mitigate false negatives but may introduce false positives.

Annotation Phase

In this second phase, we employ two procedures to refine the quality of the keyword-matched regulation-related tasks. First, we manually annotate each matched task statement and exclude false positives, such as tasks in which the word "code" or "codes" means computer programming codes, tasks in which the word "permit" or "permits" is a verb instead of a noun, tasks which include the word "government", but are unrelated to government regulation, etc. This manual annotation procedure results in a final list of 829 regulation-related tasks out of a total of 19,636 tasks in the O*Net database.[21] Second, we assign a regulation-relevance value between 0 and 1 to account for the heterogeneity among regulation-related tasks. Specifically, tasks identified by the second-tier keywords are less informative of the tasks' relevance to government regulation (e.g., the task may mention "licenses", but may be unclear about whether the licenses are issued by the government or private entities). We thus assign tasks identified by the first-tier keywords a regulation-relevance value of 1, tasks identified only by the second-tier keywords a value of 0.75, and tasks not identified by any keywords a value of 0. Moreover, while a task statement is usually just one sentence, the statements may differ in their informativeness of the tasks' relevance to government regulation. Some tasks have only one focus, while others may have multiple focuses. For instance, the following task, "*Maintain awareness of advances in medicine, computerized diagnostic and treatment equipment, data processing technology, government regulations, health insurance changes, and financing options*" has six focuses, with only one of them related to government regulations. Thus, we compute the share of regulation-related focuses out of the total number of focuses in each statement and multiply this share with the regulation-relevance value.[22]

RTI and RegIndex Construction

Having measured the regulation relevance of each task, we next compute the regulation-task intensity (RTI) for each SOC 8-digit occupation by averaging its tasks' regulation-relevance values weighted by the tasks' importance ratings for that occupation from O*Net. We then aggregate the RTI to the SOC 6-digit level.

[21] The 829 regulation-related tasks can be downloaded at www.miaobenzhang.com/Regulation_Related_Tasks.xlsx.

[22] The regulation-relevance value for the 829 regulation-related tasks has a mean of 0.55 and a standard deviation of 0.31. In analyses available from the authors, we conduct robustness checks without accounting for the share of regulation-related focuses in task statements and find very similar results for all our regression analyses.

Table 2 lists the top 25 SOC 6-digit occupations with the highest RTI. For instance, compliance officers have the highest RTI of 0.343. We interpret this RTI as indicating that compliance officers on average spend 34.3 percent of their work hours on directly performing government regulation-related tasks (e.g. "evaluate applications, records, or documents to gather information about eligibility or liability issues" or "keep informed regarding pending industry changes, trends, or best practices").

Finally, we merge each SOC 6-digit occupation's RTI to the relevant establishments in the OEWS data, which provides each establishment's labor cost (employment times wage rate) for each occupation.[23] We define an establishment's regulation index (RegIndex) as the share of its total labor cost spent on performing regulation-related tasks. In particular, an establishment i's RegIndex is its occupations' average RTI weighted by its labor cost on each occupation j at time t:

$$\text{RegIndex}_{i,t} = \frac{\sum_j RTI_j \times emp_{i,j,t} \times wage_{i,j,t}}{\sum_j emp_{i,j,t} \times wage_{i,j,t}} \times 100. \tag{3}$$

Similarly, we define RegIndex for a firm where a firm's occupational labor costs are aggregated from those of establishments with the same EIN. Table 3 shows the top 25 NAICS 3-digit industries whose establishments have the highest average RegIndex.

This study focuses on establishments' and firms' in-house regulatory compliance cost. Hence, it excludes establishments and firms in industries where legal or compliance work is their primary function or source of revenue including legal services, accounting services, central banks, and public administration.[24]

3.3 Summary Statistics

The sample for the analysis in this paper includes 3.03 million U.S. firm-year observations and 3.36 million establishment-year observations surveyed by the OEWS program from 2002 to 2014. Table 4 provides the key summary statistics. To begin with, the unweighted average firm in our sample has 92 employees and the unweighted average establishment has 48 employees, both appear above the national average based on Census statistics. If one applies the sampling weights to the establishments assigned by the OEWS, the weighted average employment falls to 14, which is closer to the establishment-level average employment of 15.6 employees reported by the Census Statistics of U.S. Businesses (SUSB) during 2002-2014. The average annual wage per employee in our sample is $43,733 ($2.09 million divided by 47.79), which is in line with the average annual

[23]Wage rate in the OEWS survey includes "base rate pay, cost-of-living allowances, guaranteed pay, hazardous-duty pay, incentive pay such as commissions and production bonuses, and tips are included in a wage. Back pay, jury duty pay, overtime pay, severance pay, shift differentials, non-production bonuses, employer costs for supplementary benefits, and tuition reimbursements are excluded." See details on the technical notes of the OEWS at https://www.bls.gov/oes/oes_doc_arch.htm.

[24]Following Song et al. (2018), educational institutions are also excluded due to high government ownership.

wage per employee of $41,974 from the Census SUSB.

The key summary statistics in Table 4 focus on RegIndex, our regulatory-compliance cost index. The average establishment in our sample spends 1.31 percent of its total labor costs on regulation-related tasks in any given year. The average firm spends 1.34 percent. To be further noted, RegIndex varies substantially across firms and establishments, with the 0.5 percentile at 0, the median at 0.8-0.9 percent, and the 99.5 percentile above 10 percent. Further decomposition shows that year fixed effects explain merely 0.01 percent of the variation in establishments' RegIndex, state fixed effects explain 0.10 percent, NAICS 6-digit industry fixed effects explain 36.13 percent, and the residual 63.63 percent of the RegIndex variation is unexplained by the above.

To illustrate the recent time trends in regulatory compliance costs, Figure 1 plots the aggregate time-series of RegIndex. Following the aggregation method explicit in the OEWS program, Figure 1 aggregates RegIndex of all establishments in our sample surveyed between $t-2$ and t, weighted by a product of the establishments' sampling weights and their total labor cost, to obtain the average RegIndex of the U.S. economy at t.

From Figure 1, we observe an increase in aggregate RegIndex from 1.49 percent in 2002 to 1.59 percent in 2014. In economic terms, this increase in RegIndex corresponds to a $26.8 billion increase in nominal regulatory compliance costs from $51.9 billion in 2002 to $78.7 billion in 2014.[25] In real terms, from 2002 to 2014 the yearly growth rate of aggregate regulatory compliance costs averaged around 1 percent, which is about half of the 1.92 percent average yearly growth rate of the U.S. Real Gross Domestic Product over the same period. Overall, these patterns indicate a substantial economic magnitude of regulatory compliance, but a burden that has been growing at a slower rate relative to the rate of the growth of the U.S. economy.

3.4 Validation

3.4.1 Establishment RegIndex and Industry Regulatory Shocks

As a first validation exercise, this subsection illustrates the response of RegIndex to four major industry-level regulatory shocks. Before and after salient regulatory policy changes, we examine the RegIndex response by establishments within a treated industry (that is, the NAICS-6 industry affected directly by the regulation under analysis) relative to appropriately matched control industries. These case studies not only help demonstrate the effectiveness of our RegIndex in tracing major industry-level regulatory shocks, but also suggest that our RegIndex is able to overcome several limitations of existing regulation measures.

[25] If we use all private establishments (including establishments that earn revenue from regulatory compliance services such as legal services), the aggregate nominal regulatory compliance costs increased by $29 billion from $74 billion in 2002 to $103 billion in 2014.

Case 1: Regulation of the Credit Card Industry In our first case study, we examine the regulatory changes that the credit card issuing industry experienced due to the enactment of the Credit Card Accountability Responsibility and Disclosure Act of 2009 (CARD Act).

The CARD Act of 2009 was drafted to "implement needed reforms and help protect consumers by prohibiting various unfair, misleading and deceptive practices in the [U.S.] credit card market" (U.S. Senate, 2009). A key aspect of the CARD Act, in particular, was to impose regulatory limits on the ability of credit card issuers to charge certain types of credit card fees on customers, which became effective in February and August of 2010 (Agarwal et al., 2018). Accordingly, as the treatment group for this analysis we use establishments in the credit card issuing industry (NAICS 52221). As the control group we use establishments in all other nondepository credit intermediation industries narrowly defined (NAICS 5222x, except for 52221, including sales financing, consumer lending, and real estate credit).

Case 1 in Figure 2 plots the weighted average RegIndex for the treated credit card issuing industry and for the control industries in the years around the policy change (from 2005 to 2014). In the figure, establishment RegIndex is weighted by the product of each establishment's sampling weight and each establishment total labor costs. We can clearly observe in this figure that before the enactment of the CARD Act in 2009 the 95 percent confidence interval of RegIndex for the treated and the control groups appear to overlap and that they are statistically indistinguishable from one another. This feature of the data suggests the validity of the parallel trends assumption necessary for the consistency of the simple difference-in-differences estimator underlying this case study. We can see further that after the policy change, RegIndex in the credit card issuing industry rises dramatically, while the RegIndex average for the control group remains basically flat. While this graphical evidence underscores the ability of RegIndex to trace the heightened regulatory burden associated with the CARD Act on credit card issuing establishments during the post period of the analysis, measuring regulation intensity at such a granular level is proven to be challenging for supply-side approaches. Indeed, we show in Online Appendix Figure A.2 that the popular RegData measure, which counts restrictive words in the Code of Federal Regulations (CFR), cannot identify a similar effect to ours.

Case 2: Deregulation and Re-regulation of the Oil and Gas Industry In our second case study, we explore both (i) the deregulation of the oil and gas extraction industry by the Energy Policy Act of 2005 (EPAct), which exempted oil and gas facilities from environmental regulations to boost production under President George W. Bush, as well as (ii) the re-regulation of the industry by President Barak Obama's executive orders following the British Petroleum (BP) Deepwater Horizon oil spill in 2010. In response to the massive pollution deriving from the BP oil spill, which began on April 20, 2010, President Obama issued two executive orders on May 21, 2010 and on January 1, 2011. These orders led to several new regulatory policies, which were regarded as "the most aggressive and comprehensive reforms to offshore oil and gas regulation and oversight in U.S. history", according to the Bureau of Ocean Energy Management.[26]

Our treated industry for this case study is oil and gas extraction (NAICS 2111). We select downstream industries which use a significant amount of the treated industry's output as their inputs as the control industries. We assume that downstream industries share close economic ties with the upstream treated industry, but face a sufficiently different set of regulations to be not as strongly affected by EPAct and the Obama orders.

In fact, both control and treated industries may be affected by similar economic and regulatory forces prior to the shocks.[27] Using the BEA input-output table from 2007, we select the following three industries as the control group: petroleum and coal products manufacturing (NAICS 3241), natural gas distribution (NAICS 2212), and basic chemical manufacturing (NAICS 3251).[28]

Case 2 in Figure 2 shows parallel trends of RegIndex for the oil and gas extraction industry and the control industries before the enactment of the EPAct in 2005. After 2005, there appears to be a dramatic decline in the RegIndex for oil and gas extraction relative to the control industries, in line with the EPAct being a deregulatory policy change. This evidence highlights an important advantage of our measure, as separating text about regulation from text about deregulation is challenging for supply-side language-based measures. Consistent with this limitation, we show in the Online Appendix Figure A.2 that the supply-side RegData measure exhibits an increase in regulation of oil and gas extraction after the EPAct is signed into law in 2005. After the BP oil spill in 2010, consistently with the contemporary understanding that the industry faced heavy re-regulation, we observe a rapid increase in the RegIndex for the oil and gas extraction industry both in absolute terms and relative to control industries.

Case 3: Requirements for Colleges Our third case study focuses on U.S. colleges. In 2011, the Department of Education's Office of Civil Rights (OCR) issued a "Dear Colleague" letter directing all of the more than 7,000 colleges that receive federal money to adopt the lowest possible standard of proof in sexual assault investigations and to require additional training of university personnel to prevent assault-related incidents. OCR noted that it "may initiate proceedings to withdraw federal funding or refer the case to the U.S. Department of Justice for litigation" if universities fail to prevent sexual assault cases. This policy shift was believed at the time to have

[26]See https://www.boem.gov/regulatory-reform/.

[27]While intuitive, choosing control industries based on input-output relations offers no guarantee that the treated and control industries will exhibit parallel trends in the Regulation Index during the pre-treatment periods. We thus examine the parallel trends empirically when analyzing each regulatory shock. Another challenge with this approach is that the control group may also be affected by the new regulatory shocks. When treatment and control groups are both affected by the regulatory shock, we will be less likely to detect significant differences between the treated and control groups post-treatment. In this sense, our selection of control industries is conservative.

[28]The input-output account data from BEA provides information at the detailed industry level for only 2007 and 2012. See https://www.bea.gov/industry/input-output-accounts-data.

increased legal and compliance costs at colleges and universities around the country by tens of millions of dollars (Keehan, 2015).[29]

For this analysis we focus on colleges, universities, and professional schools (NAICS 6113) as the treated industry and on all other training school establishments (NAICS 6114, 6115, 6116, 6117) as the control industry. Case 3 in Figure 2 shows parallel trends of RegIndex for colleges and the control industry prior to 2011, the year of the policy enactment. However, after 2011, the RegIndex of colleges increases in both absolute terms and relative to the control group, again displaying the ability of this measure to trace the regulatory change. Such informal guidance-based regulations may not immediately enter formal regulatory texts such as the CFR, making supply-side measures difficult to pick them up. For instance, in the Online Appendix Figure A.2, we observe that the RegData measure cannot clearly identify the impact of the "Dear Colleague" letter.

Case 4: Regulation of Hospitals The fourth and final case study used for the validation of RegIndex focuses on the healthcare industry. The Patient Protection and Affordable Care Act (ACA) is a landmark U.S. federal statute enacted in 2010 under President Barack Obama. The ACA represents the U.S. healthcare system's most significant regulatory overhaul and expansion of coverage since the enactment of Medicare and Medicaid in 1965. At the time of its passing, hospitals faced significant pressure in coping with the new regulatory changes and dealing with insurers, all features that should correspond with higher levels of RegIndex.

For this case study we regard hospitals (NAICS 622) as the treated group industry and use animal hospitals (NAICS 54194) as the control industry establishments. Case 4 in Figure 2 validates the parallel trends assumption for RegIndex in hospitals and animal hospitals prior to 2010, which is once again suggestive of this exercise being informative. As it can be seen from the figure, after 2010, the RegIndex of hospitals increases in both absolute terms and relative to the control group, appropriately tracing the heightened regulations imposed on the treated group relative to the controls. In this case, where ACA is clearly written in the CFR, imposes heightened regulation, and has treated and control industries very differently, supply-side measures are likely to also identify the regulation well. Indeed, we observe in the Online Appendix Figure A.2 that the RegData measure can also identify increased regulatory restrictions for hospitals from animal hospitals.

Finally, Table 5 reports the statistical significance of the graphical evidence for Cases 1-4 in Figure 2.

[29]Mandates under the new policy are discussed in greater detail in Appendix.

3.4.2 Time-Series Relation with Agency-Estimated Compliance Hours

As an additional validation exercise, we examine whether our approach for constructing RegIndex captures the time-series variation in the regulatory costs in the United States. Federal regulatory agencies are required to file Form 83-I to the OIRA in which the agency estimates firms' or individuals' compliance time for each regulation. We collect the estimates for each year from 2002 to 2014 from the "Information Collection Budget of the United States Government" from the White House website.[30]

We next compute counterpart compliance hours based on our approach for constructing RegIndex. Specifically, we regard an occupation's regulation-task intensity as the fraction of time an employee spends on regulation-related tasks in an hour. Assuming that all regulation-related occupations work 2,080 full-time hours in a year (noting that part-time workers are concentrated in the retail and restaurant industries), we estimate U.S. establishments' aggregate de facto regulatory compliance hours.

Figure 3 plots the time-series of annual aggregate regulatory compliance hours based on our approach and the compliance hours estimated by regulatory agencies. Our estimates account for about one-third of the hours estimated by federal agencies. There are at least two reasons for this discrepancy. First, our measure only accounts for businesses' compliance hours, but not households' regulatory compliance hours, which are instead factored in by OIRA. For instance, the annual compliance hours for individual income tax return accounts for about 31 percent of the total IRS tax compliance hours, which in turn represents 77 percent of the total estimated compliance hours by all agencies over our sample period. It follows that about 24 percent of the agency-estimated compliance hours concerns individual income tax returns, which cannot be accounted for by RegIndex.[31] Second, federal agencies are known to exercise systematic conservativeness when estimating the regulatory burden of rules. This is reflected in many cases in the OIRA reports showing that regulatory agencies retrospectively re-estimate lower compliance hours than their original estimates Office of Management and Budget (2005).

Despite the fact that the level of the aggregate regulatory compliance hours based on our approach is somewhat lower than the compliance hours estimated by OIRA, Figure 3 shows that the two estimates track each other robustly over time in terms of changes. The two time-series exhibit a statistically significant correlation of 67 percent.

[30]See https://www.whitehouse.gov/omb/information-regulatory-affairs/reports/. The filling for Form 83-I is mandated by the Paperwork Reduction Act (44 U.S.C. 3501). Federal agencies estimate three burden metrics for each regulatory regarding how many responses it will receive per year, how many hours it will take the public to comply with the regulation, and what would be the dollar costs of compliance. See Kalmenovitz (2019) for a review. Only estimated compliance time is consistently reported by the "Information Collection Budget" report by OIRA each year.

[31]While RegIndex does not capture households' individual tax filing costs, it may capture some pass-through business owners' individual tax filing costs. See the breakdown of IRS tax filing hours at https://taxfoundation.org/tax-compliance-costs-irs-regulations/.

3.4.3 State RegIndex and State Voting for Republican Party

More indirect dimensions of the data may also be informative of the validity of the RegIndex methodology. The political economy literature most obviously delineates clear party lines divisions regarding regulatory design and government intervention (Peltzman, 1998; Mian et al., 2010). In the U.S., Republican administrations make limiting the government burden on firms an explicit goal. Given the presence of substantial leeway at the state level in creating state-specific regulatory environments (e.g. in the case of the Insurance industry or state banking for instance[32]), one would expect to see lower levels of RegIndex in Republican-controlled or Republican-leaning constituencies. This sanity check is illustrated in this subsection.

We begin by estimating state-specific RegIndex averages, conditional on the state's industry composition. That is, we extract state fixed effects in each year controlling for industry fixed effects, and recover the conditional mean RegIndex for each one of the 50 states and the District of Columbia. All establishments are weighted by their total wage payment and the sampling weights assigned by the OEWS survey. Industry is defined at the NAICS 6-digit level.

Figure 4 reports the heat map of state-specific RegIndex averages in 2014. States with the highest RegIndex include Democratic party leaning Vermont, Connecticut, Delaware, Massachusetts, while states with the lowest RegIndex include Republican strongholds Alabama, Louisiana, North Dakota, Mississippi. Comparing our state RegIndex with the state RegData which counts for restrictive words in state regulatory texts since 2017,[33] a notable difference is that state RegData is heavily related to the number of businesses in the state. For instance, states with the highest RegData are California, New York, New Jersey, Ohio, Illinois, and Texas, while states with the lowest RegData are South Dakota, Idaho, North Dakota, Alaska, and Montana (see Online Appendix Figure A.1). States' number of establishments (from the Census SUSB) in 2017 explains 62 percent of the variation in the 2017 state RegData, where the coefficient has a t-statistics of 8.74. In contrast, states' number of establishments explains only 1 percent of our state RegIndex where the t-statistics of the coefficient is -0.74.

More systematically, state RegIndex averages correlate negatively with state political inclination to vote for the Republican Party. As an illustration, we consider states' Republican vote shares in the 2016 Presidential Election (Donald Trump vs. Hillary Clinton), the 2016 House elections, and the 2018 Senate elections. Table 6 shows that state-specific RegIndex is significantly and negatively related to the state Republican vote share in all three elections.[34]

[32] See Agarwal et al. (2014).

[33] RegData starts to count restrictive words in state regulatory texts for 16 states in 2017, 9 additional states in 2018, 18 additional states in 2018, 3 additional states in 2020, and 3 additional states in 2022. Analyzing RegData that covers the same state in multiple years reveals that state RegData is extremely stable over time, as state fixed effects explain over 99.6 percent of the variation of the pooled state RegData sample. Hence, for each state, we use the earliest available state RegData to represent the state's RegData in 2017. We download state RegData at https://quantgov-bulk-downloads.s3.amazonaws.com/State-RegData-Definitive-Edition_Regulations.zip.

[34] In Online Appendix Table A.1, we further control for state RegData and state number of establishments in the

Overall, the evidence points to RegIndex meeting certain intuitive criteria for validity. We proceed to explore some of the index's properties next.

4 Returns to Scale in Regulatory Compliance

This section examines a crucial property of RegIndex_{it} – the economies of scale in regulatory compliance costs. Specifically, we focus on the sign and magnitude of the derivative of the regulatory index with respect to establishment (or firm) employment, $\frac{\partial \text{RegIndex}_{it}}{\partial L_{it}}$.[35] We present both estimates for the whole U.S. economy and industry-specific estimates that account for heterogeneity in regulatory regimes across different sectors.

As discussed in the Introduction, economies of scale in regulatory compliance are a key feature of any regulatory architecture. Diseconomies of scale introduce a potential deterrent to firm growth, potentially pushing establishments and firms to operate below their efficient scale of production. Regulation may also introduce incentives toward concentration and may act as a barrier to entry, favoring large incumbents.[36] While the issue of returns to scale in regulation has received much attention in the Political Economy and Industrial Organization literature in some specific industries,[37] to the best of our knowledge, this paper is the first to provide a comprehensive set of facts representative of the entirety of the U.S. economy.

The simple framework in equation (1) allows one to explicitize several factors driving economies (or diseconomies) of scale under a limited set of assumptions. There are at least two factors. First, in the model, fines imposed by regulators are a function of requirements imposed by the rules, $\tilde{R}_{it}$. These standards are, in turn, function of size, $\tilde{R}_{it} = \tilde{R}(L_{it})$. Importantly, it is plausible to hypothesize $\frac{\partial \tilde{R}_{it}}{\partial L_{it}} \gtreqless 0$.[38] An example of a positive derivative is capital requirements imposed on large bank holding companies kicking in at several thresholds for total assets and as a function of the systemic importance of the financial institution (both measures correlate with employment). Another is the regulatory tiers discussed in Brock and Evans (1985). An example of a negative derivative is instead presented in Hopkins (1995), which shows that the smallest firms in his sample have paperwork and tax compliance costs (measured against turnover) about twice as

regression, and we observe very similar results to Table 6.

[35]For simplicity of exposition and with a limited abuse of notation, we will refer to L_{it} as "employment" (as opposed to the proper total employment given by the sum $L_{it} + R_{it}$). Given the magnitudes that we report in this article, this approximation is warranted and it makes both exposition and analysis much clearer. In the empirical analysis, for accuracy, we employ $L_{it} + R_{it}$.

[36]Classic references are Stigler (1971); Peltzman (1976).

[37]This dates back at least since Stigler (1971) and Peltzman (1976).

[38]In addition, note that fines may be a function of the establishment's covariates and size (in addition to the level of compliance exerted by the firm relative to a given standard required by the rules, $\tilde{R}_{it}$). That is, one could posit a general $f_{it} = f(L_{it}, R_{it}, \tilde{R}_{it})$ with $\frac{\partial f_{it}}{\partial L_{it}} \gtreqless 0$, where a positive derivative case can arise if fines are designed to be more than proportional to establishment size "to set an example", and negative if fines are capped by statutory limits/by the threat of litigation from large firms.

large as those of the largest firms.

As a second factor, scale matters through the probability of inspection p_{it}. The inspection probability is naturally driven by government enforcement effort, which can be a function of size, $E_{it} = E(L_{it})$. One may plausibly posit $p_{it} = p(E_{it})$, with $\frac{\partial p_{it}}{\partial E_{it}} \geq 0$, and $\frac{\partial E_{it}}{\partial L_{it}} \gtreqless 0$, where this derivative may be positive if larger establishments have more weight in inspection protocols or negative if, for instance, smaller plants are easier/faster to inspect.[39]

The considerations above suffice to illustrate a theoretical ambiguity in the relationship between regulatory costs and scale. Using equation (1) and the discussion above, we have:

$$R_{it} = \left(\frac{p(E_{it})\tilde{R}_{it}}{w_{it}^{r}} \right)^{\frac{1}{2}},$$

where $\frac{\partial R_{it}}{\partial L_{it}} \gtreqless 0$ depending on the dominating force. Importantly, using the definition (2), it follows that the sign of $\frac{\partial \text{RegIndex}_{it}}{\partial L_{it}}$ is also ambiguous.

This ambiguity is borne out by the data. Figure 5 presents firm and establishment-level non-parametric evidence of a non-monotonic relationship between total employment size and RegIndex. The dots in the graph represent averages for employment bins of [1, 2] employees, [3, 4], [5, 9], [10, 19], [20, 49], [50, 99], [100, 249], [250, 499], [500, 749], [750, 999], [1000, 1499], [1500, 2499], [2500, 3999], and above 4000. As firm or establishment employment increases regulatory costs per employee increase steadily until a firm size of around 500 workers or an establishment size of around 300 is achieved, then regulatory compliance costs per employee start falling rapidly, indicating economies of scale. We uncover the percentage of labor costs for regulatory compliance for mid-size businesses about 40-50 percent higher than that for the smallest businesses and about 10-20 percent higher than that for the largest ones.

The evidence from a parametric representation of the non-monotonicity is reported in Table 7, which includes the *max* and *argmax* of an inverted-U relationship between RegIndex and size for both firms and establishments. The coefficients of the parametric regression are significant and precisely estimated across all different specifications. The specifications in the table include different sets of fixed effects to assess the robustness of the finding: Year FE; Year×Industry FE; Year×Industry×State FE; Year FE + Firm FE. The table reports a range for *argmax* in panel A of about 499-511 total employees for firms at the peak of regulatory compliance costs and an *argmax* in panel B of about 309-344 for establishments. As the average firm in the United States includes only 1.26 establishments from 2002 to 2014 according to the Census SUSB, it is not surprising that *argmax* aligns between Panels A and B.

Figures 7 and 6 report the same information by large sector aggregates, highlighting a degree

[39] See Helland (1998) for a discussion related to EPA in the United States. See also Shimshack (2014) for a review of the evidence. Rousseau (2007) finds that routine environmental inspections are predicted by firm size in Belgium.

of heterogeneity in the presence of regulatory scale economies. The pictures present evidence both at the firm level in Figure 6 and the establishment level in Figure 7. Both figures show consistently that a non-monotonic relationship between RegIndex and size is evident within industries such as finance, other services, and retail. The non-monotonicity is less pronounced in manufacturing and utilities.

We further strengthen the inference of our finding on the inverted-U relationship between businesses' compliance costs and their size. In particular, one may be concerned that workers in the same occupation may in practice perform more regulation-related tasks in small firms than in mid-size firms. For instance, small firms may hire non-regulation-related occupations to cover non-regulation-related tasks and also some regulatory compliance tasks. To assess how firms' requirements on regulation-related tasks vary with their size for a given occupation, we investigate firms' skill requirements in their job posting descriptions using over 14 million job postings during 2010-2014 from the Burning Glass Technologies (BGT) data. The BGT data provide 17,420 skills extracted from the millions of job posting descriptions. For each skill, the data also provide a skill definition, usually one sentence like the task statements in the O*Net data. We next identify a BGT skill as "regulation-related" using the same procedure that identified regulation-related tasks in Section 3.2. This procedure yields 523 regulation-related skills in the BGT data. Averaging the regulation-related dummies for all skills within a job posting generates a continuous regulation-relatedness measure for each job posting. Finally, we name match BGT firms to OEWS firms and regress the regulation-related measure of job postings on the posting firms' employment and employment squared. Because the regulation-related measure is at the job posting level and varies within occupations, we can now control for Year×Occupation fixed effects.

Column (1) of Table 8 provides clear evidence that firms' demand for regulatory compliance tasks also exhibits an inverted-U relationship with firm size, with the peak at around 800 employees. That is, for the same occupation, mid-size firms require more regulatory compliance skills than small and large firms. Columns (2) and (3) further show that mid-size firms require more regulatory compliance skills than small and large firms regardless of hiring regulation-related occupations (RTI > 0) or non-regulation-related occupations (RTI $= 0$). These results suggest that our finding on the inverted-U relationship between RegIndex and firm size is a conservative estimate. The relationship between firms' actual percentage regulatory compliance costs and their size may show an even more pronounced inverted-U shape.

Some Implications

Potential distortions to the establishment and firm size distributions ensue from the presence of the economies of scale in regulation that we just documented. A full analysis of the dynamic implications of the returns to scale in regulation in the U.S. economy which we have just characterized is better left to future research. We do not possess an appropriate identification strategy to address this matter at this point and our evidence is rather suggestive than conclusive. We offer some implications nonetheless.

In the case $\tilde{R}_{it} = \rho L_{it}^{\alpha}$ and $p_{it} = \pi L_{it}^{\beta}$, we have $L_{it}^{*} = \left(\frac{\phi_i - w_{it}}{(\alpha+\beta)\left(w_{it}^{r}\pi\rho\right)^{\frac{1}{2}}} \right)^{\frac{2}{\alpha+\beta-2}}$ from the first order conditions of the establishment optimization problem (1). This, in turn, implies that policy parameters α, β, ρ, π directly affect the optimal establishment size distribution. For this purpose, Figures 7 and 6 report the change in firms' size distribution from 2002 to 2014 within each industry and for the entire economy.[40]

There is suggestive evidence that negative changes in firm mass occur in bins of the firm size distribution where regulatory compliance costs are higher, suggesting the presence of distortions. Particularly, we estimate the following regression using RegIndex and changes in the mass of firms within 14 employment bins (l) in each of 8 different industries (j) between the years 2002 and 2014:

$$\Delta EstShare_{l,j} = \beta RegIndex_{l,j} + \delta_j + \epsilon_{l,j}.$$

In this regression, a negative β estimate would indicate a hollowing out of an industry-employment bin when the level of RegIndex in that bin is higher.

We obtain negative and statistically significant estimates of $\beta = -0.017$ ($t = -3.16$) based on standard errors clustered at the sector level. These estimates further imply that the substantially higher regulatory burden falling on mid-size firms (and establishments) correlates strongly with a loss of mass in the middle of the firm (and establishment) size distribution over time. This hollowing out in the middle of the size distribution should not be uniquely ascribed to regulatory requirements, as many omitted confounding factors may affect this negative correlation, but it should be taken as suggestive evidence of the broader implications of our approach.

5 Decoupling Regulatory Requirements from Enforcement Effort

As discussed in Section 3, the share of regulatory compliance costs RegIndex$_{it}$ can be driven by both the extent of regulatory agencies' regulatory requirements applied onto establishments, $\tilde{R}_{it}$, and also the extent of agencies' enforcement effort, E_{it}, which in turn affects the likelihood of inspection p_{it}. These competing drivers introduce a challenge in interpreting the findings in Section 4. For example, a small establishment may exhibit low RegIndex either because the agency rules are designed to be lighter on smaller establishments (proxied by lower $\tilde{R}_{it}$) or because small establishments are more likely to fly under the radar of regulators (proxied by lower p_{it}) and therefore they choose to comply less to rules that otherwise would apply equally to all establishments.

[40] In Figure 6 we compute a firm's sampling weight as the minimum sampling weight of its establishments.

This section presents a method to analyze the impact of the two main factors on driving the RegIndex of an establishment (or firm) of a given size. In order to identify the differential role of regulatory requirements and of enforcement in driving regulatory compliance, we add additional information originating from the supply side, i.e. from the regulatory agencies who design and enforce the rules. In what follows, we describe and implement two shift-share regulatory supply shocks to help inspect the importance of regulatory requirements versus enforcement effort in driving the inversed-U shape of RegIndex with respect to size.

For each regulatory agency k, let us assume at time t one is able to measure both its regulatory requirement, reg_{kt}, and the extent, for given reg_{kt}, of the enforcement and supervision effort, enf_{kt}. The measurement of reg_{kt} and enf_{kt} is discussed in subsection 5.1 below and, for now, we solely clarify how shocks in reg_{kt} and enf_{kt} end up differentially affecting an establishment i's RegIndex through $\tilde{R}_{it}$ and p_{it}.

We posit that different industries have different exposure to regulations falling under agency k's oversight. Assume one can measure industry j's RegIndex originating from regulatory agency k, r_{jkt}. Then, one can, in turn, create two shift-share instrumental variables tracing changes in the regulatory requirements and in the enforcement pertinent to establishment i from industry j:

$$\begin{aligned} iv(\Delta\text{log}(\tilde{R}_{it})) &= \sum_k \Delta \log reg_{kt} \times r_{jk0} + \nu_{it} \\ iv(\Delta\text{log}(p_{it})) &= \sum_k \Delta \log enf_{kt} \times r_{jk0} + \upsilon_{it}, \end{aligned} \quad (4)$$

where one employs, for both regulatory requirements and enforcement effort, establishment-agency weights at a predetermined initial time 0, ω_{jk0}. This latter step is useful to avoid contamination stemming from endogenous shifts in product or process to reduce exposure to certain rules associated with k. ν_{it} and υ_{it} are establishment-specific i.i.d. shocks.

One can now use $iv(\Delta\text{log}(\tilde{R}_{it}))$ and $iv(\Delta\text{log}(p_{it}))$ to instrument log changes in establishment i's regulatory requirements, $\Delta\text{log}(\tilde{R}_{it})$ and enforcement effort, $\Delta\text{log}(p_{it})$, from pertinent regulatory agencies. For simplicity, let us assume a flexible, linear reduced-form instrumental variable model with establishment-level controls X_{it} and i.i.d. shocks:[41]

$$\begin{aligned} \Delta\text{log}(\tilde{R}_{it}) &= \beta_0 + \beta_1 iv(\Delta\text{log}(\tilde{R}_{it})) + \beta_2 X_{it} + \varepsilon_{it} \\ \Delta\text{log}\,(p_{it}) &= \gamma_0 + \gamma_1 iv(\Delta\text{log}(p_{it})) + \gamma_2 X_{it} + \eta_{it}. \end{aligned} \quad (5)$$

[41] As a caveat, these equations should be intended as approximations, rather than structural relationships, as we do not set to fully represent the response (and expectations) of individual firms with respect to the behavior of their pertinent regulatory agencies.

In our estimation, we allow β_1 and γ_1 to vary by establishment size, so as to highlight how regulatory requirements and enforcement effort contribute to the inverted-U shape relation between RegIndex and size in Section 4. In equation (5), regulatory requirements for the establishment go up when regulatory agencies overseeing the establishment increase their regulatory requirements (first equation) and compliance behavior of establishments increases as agencies invest more in enforcement (second equation). Notice further that equations in (5) are expressed in first differences and consequently they partial out establishment-specific fixed effects, which allow this representation to capture a vast array of time-invariant characteristics, including industry and geographic location.

Using the definition of RegIndex in equation (2) and taking log differences of R_{it} in the equilibrium condition (1), it follows that:

$$\begin{aligned}\Delta\log(\text{RegIndex}_{it}) &= \frac{1}{2}\Delta\log(\tilde{R}_{it}) + \frac{1}{2}\Delta\log(p_{it}) + \frac{1}{2}\Delta\log(w^r_{it}) - \Delta\log W_{it} \\ &= \delta_0 + \frac{1}{2}\beta_1 iv(\Delta\log(\tilde{R}_{it})) + \frac{1}{2}\gamma_1 iv(\Delta\log(p_{it})) + \frac{1}{2}\Delta\log(w^r_{it}) - \Delta\log W_{it} + \delta_2 X_{it} + \xi_{it} \qquad (6)\end{aligned}$$

which is a testable reduced-form equation where $\delta_0 = \frac{\beta_0+\gamma_0}{2}$, $\delta_2 = \frac{\beta_2+\gamma_2}{2}$, and $\xi_{it} = \varepsilon_{it} + \eta_{it} + \upsilon_{it} + \nu_{it}$. The estimation delivers the parameters β_1, γ_1 that express the responses of establishments' RegIndex to regulatory requirements shocks and enforcement effort shocks. Based on these estimates, we inspect how regulatory requirements and enforcement effort contribute to the inverted-U shape relation between RegIndex and size in Section 5.2.

5.1 Measuring Shocks to Regulatory Requirements and Enforcement

We measure regulatory requirements originating from each major agency in the U.S., reg_{kt}, by using the number of new regulations originated from k each year. For each fiscal year, major agencies need to file to the OIRA their estimates of the changes in regulatory compliance hours for regulations under their oversight.[42] Such changes are further decomposed into changes in enactment and retirement of regulations and in changes in agencies' re-estimation of regulatory compliance hours. We use 3-year log differences in compliance hours due to changes in enactment and retirement of regulations to measure regulation-requirement shocks of the agency to firms, i.e. $\Delta\log reg_{kt}$, in equation (4).

We measure enforcement shocks enf_{kt} for each major agency k using panel data of U.S. federal government employees from 2002 to 2014. This individual-level database is compiled by the U.S.

[42] Major agencies' estimated changes in regulatory compliance hours can be downloaded from the "Information Collection Budget of the United States Government" reports at https://www.whitehouse.gov/omb/information-regulatory-affairs/reports/#ICB.

Office of Personnel Management's (OPM) Enterprise Human Resources Integration System. The data are made available by BuzzFeed News through a Freedom of Information Act request and have found growing application in the economics of the U.S. public administration (Spenkuch et al., 2021). The data cover detailed information of all federal employees, except for the Department of Defense, at a quarterly frequency.[43] Variables crucial for our study include the employee's agency, occupation, and full time/part time employment status. Importantly, while the data adopt a different occupation classification system from SOC, we are able to obtain each federal employee occupation's task description from the "Handbook of Occupational Group and Families" on the OPM website.[44] We thus identify each occupation as "regulation-related" using the same list of keywords and following the same procedure as in Section 3. We use 3-year log differences in regulation-related full time employment in each agency to measure enforcement shocks of the agency to firms, i.e., $\Delta \log enf_{kt}$ in equation (4).[45]

Finally, we select 12 agencies that have both regulation-requirement shocks and enforcement shocks from 2002 to 2014. These agencies include Department of Labor (DOL), Department of Transportation (DOT), Department of Education (ED), Environmental Protection Agency (EPA), Federal Communications Commission (FCC), Federal Deposit Insurance Corporation (FDIC), Federal Trade Commission (FTC), Department of Health and Human Services (HHS), Department of Housing and Urban Development (HUD), Nuclear Regulatory Commission (NRC), Securities and Exchange Commission (SEC), and Department of Agriculture (USDA).[46] These 12 agencies account for 81 percent of non-Treasury regulatory compliance hours for an average year from 2002 to 2014.[47] Figure 8 plots the time series for $\Delta \log enf_{kt}$ and $\Delta \log reg_{kt}$ for each main agency in our sample.

Our shift-share instrumental variables in equation (4) also require measuring each industry's exposure to the 12 regulatory agencies' shocks, r_{jk0}. To do so, we extract the top 50 identifying keywords for each regulatory agency using natural language processing of the CFR text. In particular, we first select all the CFR volumes related to the 12 agencies and to their subagencies.[48]

[43] The data includes 206 million observations and can be downloaded at https://archive.org/download/opm-federal-employment-data. We thank Joe Raffiee for introducing this data to us.

[44] The OPM has its own definitions for government occupations that are different from the SOC system. The handbook for OPM occupation description can be downloaded at https://www.opm.gov/policy-data-oversight/classification-qualifications/classifying-general-schedule-positions/occupationalhandbook.pdf.

[45] Our procedure identifies 227 out of 612 occupations of federal employees as "regulation-related." In Online Appendix A, we conduct a robustness check by identifying a narrower list of 105 enforcement-focused regulation-related occupations, and we re-estimate our model using this alternative measure of enforcement shocks which is potentially a more volatile supply-side instrument for regulatory enforcement. Online Appendix Tables A.2-A.5 show virtually no substantive differences between results using the alternative shocks from our main results.

[46] The employment data are from the detailed subagencies. We aggregate regulation-related occupations for all subagencies that exist throughout 2002-2014 using the OPM links between agency and subagencies at https://www.opm.gov/about-us/open-government/Data/Apps/Agencies/.

[47] About 99.4 percent of the regulatory compliance hours from the Department of Treasury are from the Internal Revenue Service (IRS).

[48] CFR texts are available at https://www.govinfo.gov/app/collection/cfr/2021/.

Next, we count the term frequency of each word in the overall selected CFR text ($count_{all}$) and the also in each agency k's texts ($count_k$). We then identify keywords as specific to agency k if the keywords appear over 50 percent of the time in agency k's texts alone, i.e., $count_k/count_{all} \geq 0.5$. Third, we further remove uninformative words from the list, such as abbreviations by computing the similarity of the keywords to its agency's name using the Google BERT. Our final keywords are the top 50 keywords that have the highest BERT similarity score.[49] We compute a Google BERT similarity between each regulation-related tasks in Section 3 and an agency's keywords, and we standardize the 12 similarities for each task to sum up to 1. These standardized similarities capture a regulation-related task's exposure to the agencies.

Multiplying a regulation-related occupation's regulatory-task intensity with the shares results in the occupation's compliance intensity related to each of the 12 agencies, r_{jk0} in equation 4. Following our procedure in Section 3, we compute each NAICS 6-digit industry j's percentage labor spending towards compliance with agency k's regulations in year t, i.e., r_{jk0} in equation (4). With all elements in equation (4) now constructed, we compute the two instrumental variables, $iv(\Delta\log(p_{it}))$ and $iv(\Delta\log(\tilde{R}_{it}))$, following equation (4).

Our instrumental variables exhibit desirable properties that are intuitive. First, Table 9 reports the top 3 industries for each regulatory agency. The table shows intuitive profiles of oversight, which supports by and large the intuitive validity of our approach based on keywords. For instance, the Environmental Protection Agency (EPA) reports Waste Management and Remediation Services; Petroleum and Coal Products Manufacturing; and Construction of Buildings as its top industries under oversight. The Securities and Exchange Commission (SEC) reports Securities, Commodity Contracts, and Other Financial Investments and Related Activities; Credit Intermediation and Related Activities; Funds, Trusts, and Other Financial Vehicles as its top industries, and so on.

Second, we explore how regulatory and enforcement shocks co-move over time. Figure 8 traces the time series of the two shocks for each main regulatory agency. We observe that there is substantial independent variation in each of the two shock series across all regulators, although for some agencies the separation is starker. Dynamics also differ. For instance, SEC sharply accelerates hiring in the aftermath of the financial crisis of 2008, while other agencies, like FCC, do not. Finally, Table 10 shows that the overall correlation between the instrumental variables is 12 percent at the establishment level.

5.2 Regulatory Requirements and Enforcement Effort

We can now present the estimation of equation (6). Table 11 reports our results for the entire U.S. sample of firms and establishments over the entire sample period. Coefficients are reported

[49] Appendix Table A.2 lists the 50 keywords for each of the 12 agencies.

for standardized variables, so that they can be interpreted as in units of standard deviations. We begin by investigating the roles of $iv(\Delta\log(p_{it}))$ and $iv(\Delta\log(\tilde{R}_{it}))$ separately.

Columns (1) and (4) examine the response of firms' and establishments' share of regulatory costs ($\Delta\log(\text{RegIndex}_{it})$), respectively, to enforcement shocks, $iv(\Delta\log(p_{it}))$, accounting for controls required by (6), particularly, $\Delta\log W_{it}$ and $\Delta\log(w^r_{it})$. In both columns, we report positive and statistically significant coefficients, confirming the intuitive conditional correlation between increases in agency regulatory hires translating into more enforcement and consequently higher regulatory compliance expenditure. The estimated coefficients further indicate a quantitatively meaningful relationship, as a one standard deviation increase in $iv(\Delta\log(p_{it}))$ produces a 0.21 of a standard deviation increase in the change in regulatory compliance costs index for a firm (0.23 for establishments).

Columns (2) and (5) examine the response of firms' and establishments' share of regulatory costs ($\Delta\log(\text{RegIndex}_{it})$), respectively, to regulation requirements shocks, $iv(\Delta\log(\tilde{R}_{it}))$. The conditional correlation for $iv(\Delta\log(\tilde{R}_{it}))$ is higher than for the case of $iv(\Delta\log(p_{it}))$. A one standard deviation increase in $iv(\Delta\log(\tilde{R}_{it}))$ produces a 0.26 of a standard deviation increase in the change in regulatory compliance costs index for a firm (0.27 for establishments). This appears to suggest a preponderance of regulatory requirements in driving compliance costs relative to the role of enforcement.

Columns (3) and (6) include both shocks in the estimation of equation (6) to inspect the marginal effects of changes in regulatory requirements, $iv(\Delta\log(\tilde{R}_{it}))$, from that of changes in enforcement effort, $iv(\Delta\log(p_{it}))$. The estimation reveals a stronger role for regulatory requirements than for enforcement effort in driving changes in RegIndex. The effect of regulatory requirements is estimated at 0.23 and statistically significant, while that of enforcement effort is at 0.11 and statistically insignificant for firms (and similarly for establishments).

How can these instrumental variables be informative about the relation between the level of RegIndex and size? We inspect the connection between the level and log changes in RegIndex. Note that $\Delta\log(\text{RegIndex}_{it}) = \log(\text{RegIndex}_{it}) - \log(\text{RegIndex}_{i,t-3})$. Panel A of Figure 9 shows the average RegIndex at $t-3$ and t for firms in four employment bins: between 1 and 19 employees, between 20 and 399, between 400 and 749, and above 750 employees. We observe that over time, the inverted-U shape relation between RegIndex and firm size became stronger. In particular, much of the changes in RegIndex come from firms with a medium and high level of employment, while there is little change in RegIndex for small firms. Importantly, this enhanced inversed-U relationship between RegIndex and size maps to a greater average log change in RegIndex for larger firms, as is evident in Panel B.[50]

Table 12 reports that much of the increase in $\log(\text{RegIndex}_{it})$ for larger firms originates from changes in regulatory requirements, $iv(\Delta\log(\tilde{R}_{it}))$, rather than enforcement, $iv(\Delta\log(p_{it}))$. In

[50]Figure 10 shows a very similar pattern at the establishment level.

particular, Columns (1)-(4) show that enforcement shocks do not appear to have a statistically significant effect on $\Delta\log(\text{RegIndex}_{it})$. In contrast, regulatory-requirement shocks show a significant impact on $\Delta\log(\text{RegIndex}_{it})$, particularly for larger firms. A test of the coefficients for $iv(\Delta\log(\tilde{R}_{it}))$ and $iv(\Delta\log(p_{it}))$ shows that the effect of $iv(\Delta\log(\tilde{R}_{it}))$ is significantly greater than the effect of $iv(\Delta\log(p_{it}))$ with p-value of 0.09 and 0.04 for firms in the 400-749 employment bin and in the ≥750 employment bin, respectively. Columns (5)-(8) show similar but slightly weaker results at the establishment level, with the coefficients for $iv(\Delta\log(\tilde{R}_{it}))$ and $iv(\Delta\log(p_{it}))$ not statistically different for establishments in the 400-749 employment bin (p-value = 0.22) and significantly different for establishments in the ≥750 employment bin (p-value = 0.01). In sum, the evidence in this section suggests that the inverted-U shape relation between RegIndex and size has become sharper over time, and mostly from the fact that medium and large firms have experienced greater increases in RegIndex. Changes in regulatory requirements appear to have contributed significantly to such increases, while the contribution from changes in enforcement is rather limited.

Finally, Tables 13 and 14 display variations of the above results across broad sectors of firms and establishments, respectively. First of all, changes in log RegIndex correlate highly with changes in regulatory requirements, $iv(\Delta\log(\tilde{R}_{it}))$ and changes in enforcement, $iv(\Delta\log(p_{it}))$ individually across all sectors. These high correlations across the board further validate our RegIndex measure in that $iv(\Delta\log(\tilde{R}_{it}))$ and $iv(\Delta\log(p_{it}))$ are driven by dynamics from outside our main OEWS data, one from agencies' enaction and retirement of regulations filed to OIRA and the other from changes in agencies' regulation-related employment. When employing both instruments in the analysis, we observe that across all sectors regulatory requirements appear to play a more important and statistically precise role in driving business regulatory compliance costs. Enforcement effort plays an important role for businesses in the manufacturing sector and for smaller businesses in Retail, Wholesale, Utilities, and Finance.

6 Conclusion

This paper presents a new approach to estimating the compliance costs of regulation in the United States. Based on establishment-occupation level data, we quantify the total labor costs paid by businesses to employees engaging in safety, compliance, monitoring, and other regulation-related tasks in order to meet federal, state, and local regulatory requirements. A typical U.S. establishment spends about 1.31 percent of its total wage bills on employees for performing regulatory compliance tasks.

We show that the percentage of establishments' labor costs paid for regulatory compliance first increases with establishment size, measured by of total employment, and then decreases, exhibiting an inverted-U shape. This inversed-U shape suggests that for small businesses regulation is tiered and tends to be lighter, while red tape increases as employment reaches 500 workers. Beyond this threshold, regulatory costs tend to decrease, indicating economies of scale kicking in for regulatory compliance. Identifying the presence of increasing returns to scale in regulatory compliance for medium-large firms is an important step in the direction of assessing equilibrium distortions due to the design of the U.S. administrative system, which is a prominent area of future investigation.

We further design and implement a shift-share instrumental variable method to identify establishment responses to regulation requirements versus enforcement. Using this design, we argue that changes in regulatory requirements appear to contribute significantly to the enhanced inversed-U relationship between establishments' regulatory compliance cost and size in our sample period, while the contribution from changes in enforcement is rather limited.

Future research may extend the use of our methodology to other high-income countries, where similar microdata are available. Such extensions can provide a comparative perspective on the costs of regulation and allow to further assess the external validity of the approach presented in this paper. Quantifying the productivity losses (or gains) to firms due to the extent of the regulatory burden also appears to be a fruitful direction of future inquiry.

References

Adler, M.D., Posner, E.A., 1999. Rethinking cost-benefit analysis. Yale LJ 109, 165.

Agarwal, S., Chomsisengphet, S., Mahoney, N., Stroebel, J., 2018. Do banks pass through credit expansions to consumers who want to borrow? The Quarterly Journal of Economics 133, 129–190.

Agarwal, S., Lucca, D., Seru, A., Trebbi, F., 2014. Inconsistent regulators: Evidence from banking. The Quarterly Journal of Economics 129, 889–938.

Al-Ubaydli, O., McLaughlin, P.A., 2017. Regdata: A numerical database on industry-specific regulations for all united states industries and federal regulations, 1997–2012. Regulation & Governance 11, 109–123.

Arrow, K.J., Cropper, M.L., Eads, G.C., Hahn, R.W., Lave, L.B., Noll, R.G., Portney, P.R., Russell, M., Schmalensee, R., Smith, V.K., et al., 1996. Is there a role for benefit-cost analysis in environmental, health, and safety regulation? Science 272, 221–222.

Ayyagari, M., Maksimovic, V., 2017. Fewer and less skilled? human capital, competition, and entrepreneurial success in manufacturing. Human Capital, Competition, and Entrepreneurial Success in Manufacturing (December 26, 2017) .

Baker, S.R., Bloom, N., Canes-Wrone, B., Davis, S.J., Rodden, J., 2014. Why has us policy uncertainty risen since 1960? American Economic Review P&P 104, 56–60.

Baker, S.R., Bloom, N., Davis, S.J., 2016. Measuring economic policy uncertainty. The quarterly journal of economics 131, 1593–1636.

Bertrand, M., Bombardini, M., Fisman, R., Hackinen, B., Trebbi, F., 2021. Hall of mirrors: Corporate philanthropy and strategic advocacy. The Quarterly Journal of Economics 136, 2413–2465.

Bertrand, M., Bombardini, M., Trebbi, F., 2014. Is it whom you know or what you know? an empirical assessment of the lobbying process. The American Economic Review 104, 3885–3920.

Boeri, T., Jimeno, J.F., 2005. The effects of employment protection: Learning from variable enforcement. European Economic Review 49, 2057–2077.

Bombardini, M., Trebbi, F., 2020. Empirical models of lobbying. Annual Review of Economics , 391–413.

Brock, W.A., Evans, D.S., 1985. The economics of regulatory tiering. The Rand Journal of Economics , 398–409.

Brock, W.A., Evans, D.S., Phillips, B.D., 1986. The economics of small businesses: Their role and regulation in the US economy. Holmes & Meier New York.

Callander, S., Foarta, O., Sugaya, T., 2021. Market Competition and Political Influence: An Integrated Approach. Technical Report. Centre for Economic Policy Research.

Calomiris, C.W., Mamaysky, H., Yang, R., 2020. Measuring the cost of regulation: A text-based approach. Technical Report. National Bureau of Economic Research.

Chodorow-Reich, G., 2014. The employment effects of credit market disruptions: Firm-level evidence from the 2008–9 financial crisis. The Quarterly Journal of Economics 129, 1–59.

Cowgill, B., Prat, A., Valletti, T., 2021. Political power and market power. arXiv preprint arXiv:2106.13612 .

Davis, S.J., 2017. Regulatory complexity and policy uncertainty: headwinds of our own making. Becker Friedman Institute for Research in economics working paper .

Drutman, L., 2015. The business of America is lobbying: How corporations became politicized and politics became more corporate. Oxford University Press.

Evans, D.S., 1986. The differential effect of regulation across plant size: Comment on pashigian. The Journal of Law and Economics 29, 187–200.

Garicano, L., Lelarge, C., Van Reenen, J., 2016. Firm size distortions and the productivity distribution: Evidence from france. American Economic Review 106, 3439–79.

Gourio, F., Roys, N., 2014. Size-dependent regulations, firm size distribution, and reallocation. Quantitative Economics 5, 377–416.

Harrington, W., Morgenstern, R.D., Nelson, P., 2000. On the accuracy of regulatory cost estimates. Journal of Policy Analysis and Management 19, 297–322. doi:10.1002/(SICI)1520-6688(200021)19:2<297::AID-PAM7>3.0.CO;2-X.

Hassan, T.A., Hollander, S., Van Lent, L., Tahoun, A., 2019. Firm-level political risk: Measurement and effects. The Quarterly Journal of Economics 134, 2135–2202.

Helland, E., 1998. The enforcement of pollution control laws: Inspections, violations, and self-reporting. Review of Economics and Statistics 80, 141–153.

Hopkins, T.D., 1995. Profiles of regulatory costs. Report to the US Small Business Administration, Office of Advocacy .

Hsieh, C.T., Olken, B.A., 2014. The missing" missing middle". Journal of Economic Perspectives 28, 89–108.

Julio, B., Yook, Y., 2012. Political uncertainty and corporate investment cycles. The Journal of Finance 67, 45–83.

Kalmenovitz, J., 2019. Regulatory intensity. Available at SSRN 3451344.

Kalmenovitz, J., 2021. Regulatory intensity and firm-specific exposure. mimeograph Drexel University .

Kang, K., 2016. Policy influence and private returns from lobbying in the energy sector. Review of Economic Studies 83, 269–305.

Keehan, A., 2015. Confronting campus sexual assault claims: An examination of higher education claims. United Educators Report .

Lancieri, F., Posner, E.A., Zingales, L., 2022. The political economy of the decline in antitrust enforcement in the united states. Available at SSRN .

List, J.A., Millimet, D.L., Fredriksson, P.G., McHone, W.W., 2003. Effects of environmental regulations on manufacturing plant births: evidence from a propensity score matching estimator. Review of Economics and Statistics 85, 944–952.

Mian, A., Sufi, A., Trebbi, F., 2010. The political economy of the us mortgage default crisis. American Economic Review 100, 1967–98.

Morrall, J.F., 1986. A review of the record. Regulation 10, 25.

OTA, 1995. Gauging Control Technology and Regulatory Impacts in Occupational Safety and Health: An Appraisal of OSHA's Analytic Approach. US Government Printing Office.

Peltzman, S., 1976. Toward a more general theory of regulation. The Journal of Law and Economics 19, 211–240.

Peltzman, S., 1998. Political participation and government regulation. University of Chicago Press.

Philippon, T., 2019. The great reversal: How America gave up on free markets. Harvard University Press.

Posner, E., Weyl, E.G., 2013. Benefit-cost analysis for financial regulation. American Economic Review 103, 393–97.

Posner, E.A., Weyl, E.G., 2017. Property is only another name for monopoly. Journal of Legal Analysis 9, 51–123.

Posner, R.A., 1974. Theories of economic regulation. Bell Journal of Economics 5, 335–358.

Rousseau, S., 2007. Timing of environmental inspections: Survival of the compliant. Journal of regulatory economics 32, 17–36.

Ryan, S.P., 2012. The costs of environmental regulation in a concentrated industry. Econometrica 80, 1019–1061.

Schivardi, F., Torrini, R., 2008. Identifying the effects of firing restrictions through size-contingent differences in regulation. Labour Economics 15, 482–511.

Shimshack, J.P., 2014. The economics of environmental monitoring and enforcement. Annu. Rev. Resour. Econ. 6, 339–360.

Shleifer, A., Vishny, R.W., 1998. The grabbing hand: Government pathologies and their cures. Harvard University Press.

Simkovic, M., Zhang, M.B., 2020. Regulation Intensity and Technology-Driven Entry: Measurement and Micro-Evidence. SSRN Scholarly Paper ID 3205589. Rochester, NY. doi:10.2139/ssrn.3205589.

Song, J., Price, D.J., Guvenen, F., Bloom, N., Von Wachter, T., 2018. Firming up inequality. The Quarterly Journal of Economics 134, 1–50.

Spenkuch, J.L., Teso, E., Xu, G., 2021. Ideology and performance in public organizations. Technical Report. National Bureau of Economic Research.

Stigler, G.J., 1971. The theory of economic regulation. The Bell journal of economics and management science , 3–21.

Stiglitz, J., 2009. Regulation and Failure. New Perspectives on Regulation , 13–25.

Sunstein, C.R., 2002. The cost-benefit state: the future of regulatory protection, American Bar Association.

Sunstein, C.R., 2017. Cost-benefit analysis and arbitrariness review. Harv. Envtl. L. Rev. 41, 1.

Sunstein, C.R., 2020. On neglecting regulatory benefits. Available at SSRN 3541782 .

National Association of Manufacturers, 2014. The cost of federal regulation to the u.s. economy, manufacturing and small business .

Office of Information and Regulatory Affairs (OIRA), 2020. 2018, 2019, and 2020 report to congress on the benefits and costs of federal regulations and agency compliance with the unfunded mandates reform act .

Office of Management and Budget, 2005. Validating Regulatory Analysis: Report to Congress on the Costs and Benefits of Federal Regulations and Unfunded Mandates on State, Local, and Tribal Entities. Washington DC: Government Printing Office.

Securities Industry Association, 2006. The costs of compliance in the u.s. securities industry: Survey report .

Thomas, L.G., 1990. Regulation and firm size: Fda impacts on innovation. The RAND Journal of Economics , 497–517.

Tybout, J., 2014. The missing middle, revisited. Journal of Economic Perspectives 28, 235–36.

Blanes i Vidal, J., Draca, M., Fons-Rosen, C., 2012. Revolving door lobbyists. The American Economic Review 102, 3731–3748.

Figure 1: **Aggregate Series of Regulation Index**

This figure plots the aggregate Regulation Index (RegIndex) from 2002 to 2014. *RegIndex* is the percentage of an establishment's annual labor spending on performing regulation-related tasks (see Section 3). We focus on non-government/non-education industries (Song et al., 2018) and exclude industry categories which provide legal or compliance work as their primary function or source of revenue: legal services, accounting firms, government administration, courts, and central banking.

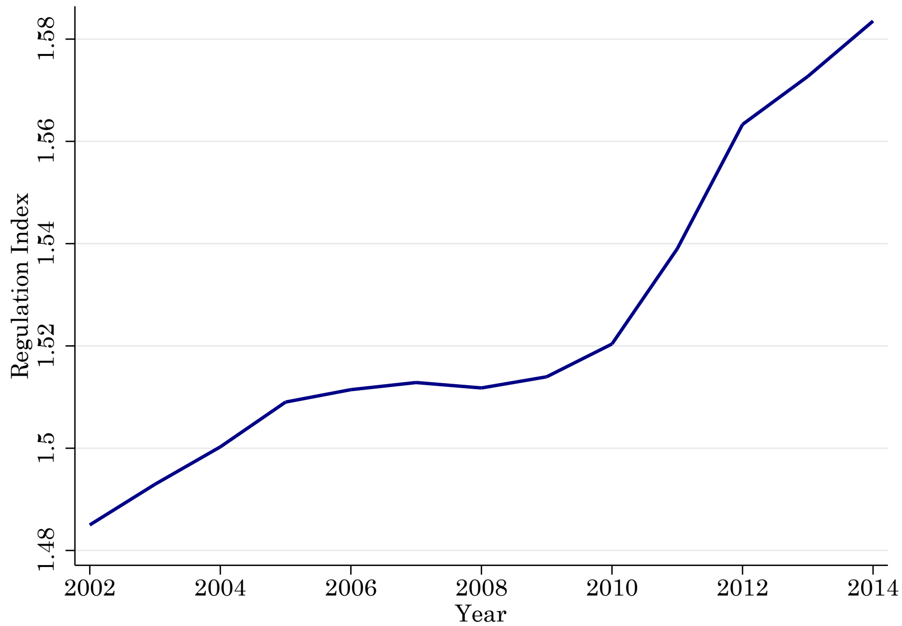


Figure 2: **Validation of RegIndex Using Industry Regulatory Shocks**

This figure plots the response of industries' Regulation Index (RegIndex) to five industry-level regulatory shocks. RegIndex is the percentage of an industry's annual labor spending on performing regulation-related tasks. Section 3 provides details of the industry shocks and discusses the classification of treated and control groups. To ease the comparison, we shift the lines vertically so that they have the same value in the year before the treatment. The value in the year before the treatment is the average of the regulation measures across the treated and control industries in that year. The difference between the two lines after the treatment, minus the difference between the two lines before the treatment reflects the difference-in-differences estimation. The shaded areas indicate the 95% confidence interval of the industries' average RegIndex.

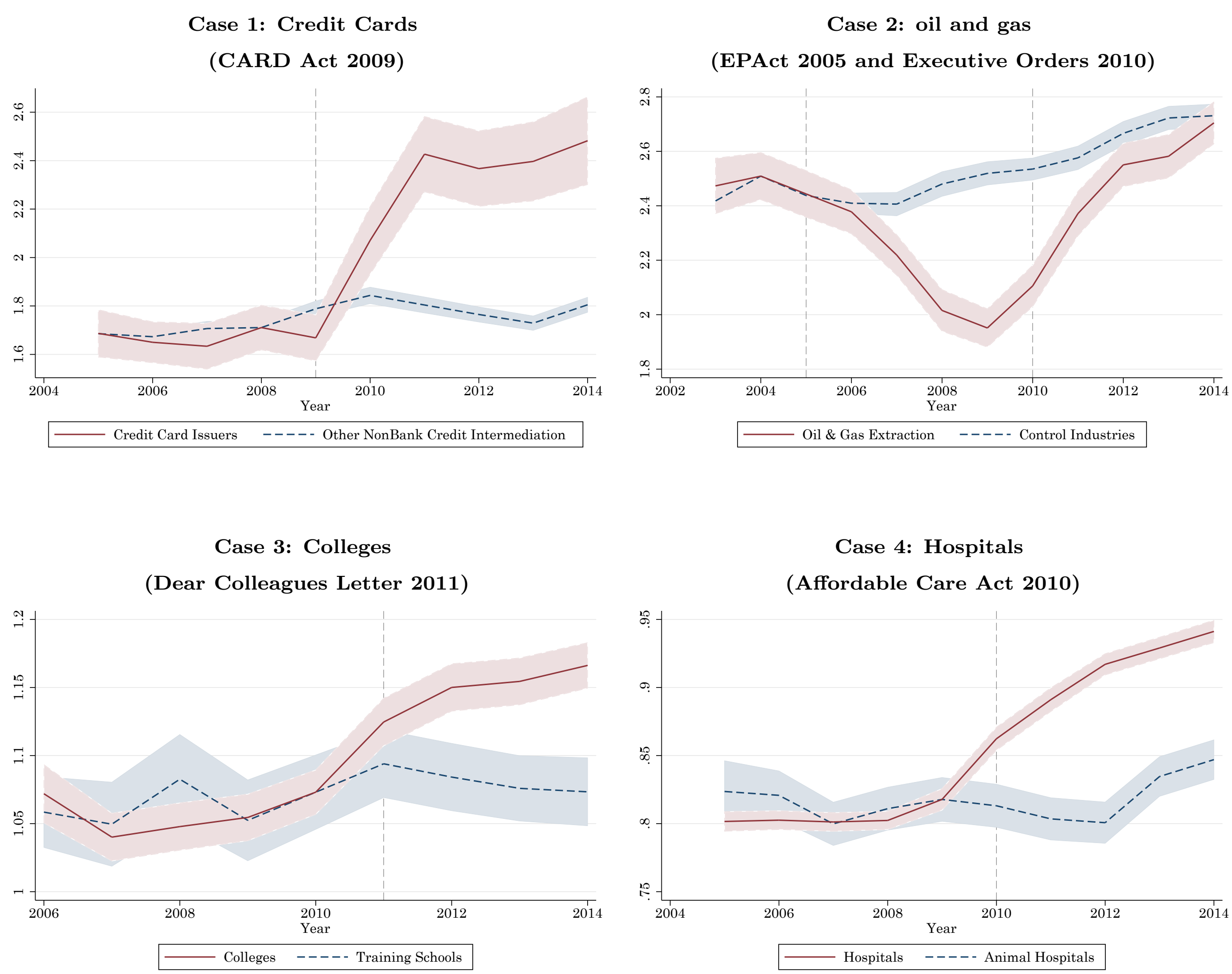


Figure 3: **Validation of RegIndex Using Agency-Estimated Compliance Hours**

This figure plots the aggregate annual compliance hours (in billions of hours) identified by our RegIndex measure and the estimated annual compliance hours (in billions of hours) submitted by various regulatory agencies to the White House Office of Information and Regulatory Affairs (OIRA).

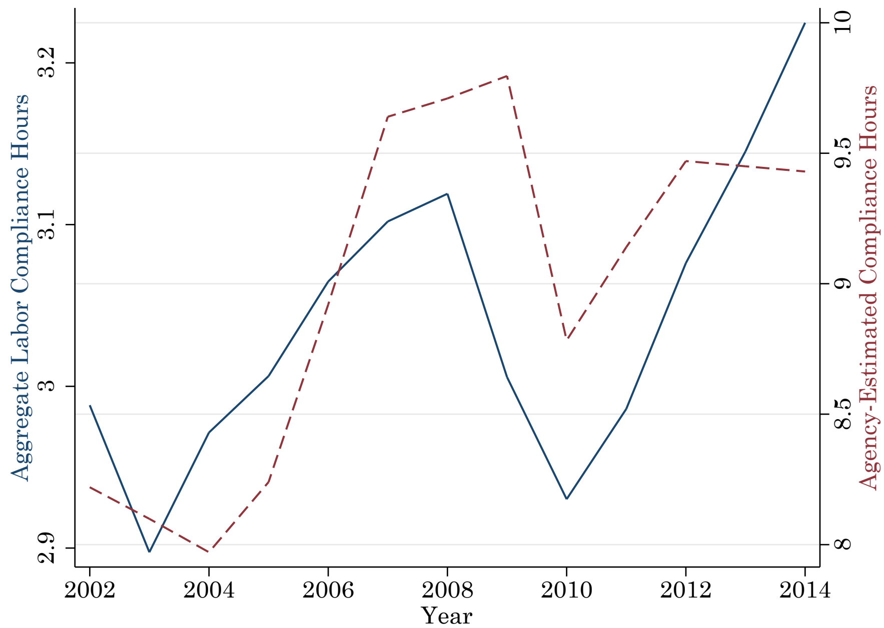


Figure 4: **Regulation Index Across States**

This figure plots the coefficients on state dummies in the following regressions based on about 1 million private establishments in the 2014 OEWS universe: $RegIndex_{i,t} = \alpha + \sum_{s \in States} \beta_s \times State_s + FE_{Ind} + \epsilon_{i,t}$ The coefficient α shows the RegIndex for the benchmark state "Alabama." The sum of coefficients $\alpha + \beta_s$ shows the RegIndex for the other 50 states (including the District of Columbia). All establishments are weighted by their total wage payment and their sampling weights assigned by the OEWS survey. Industry is at the NAICS 6-digit level.

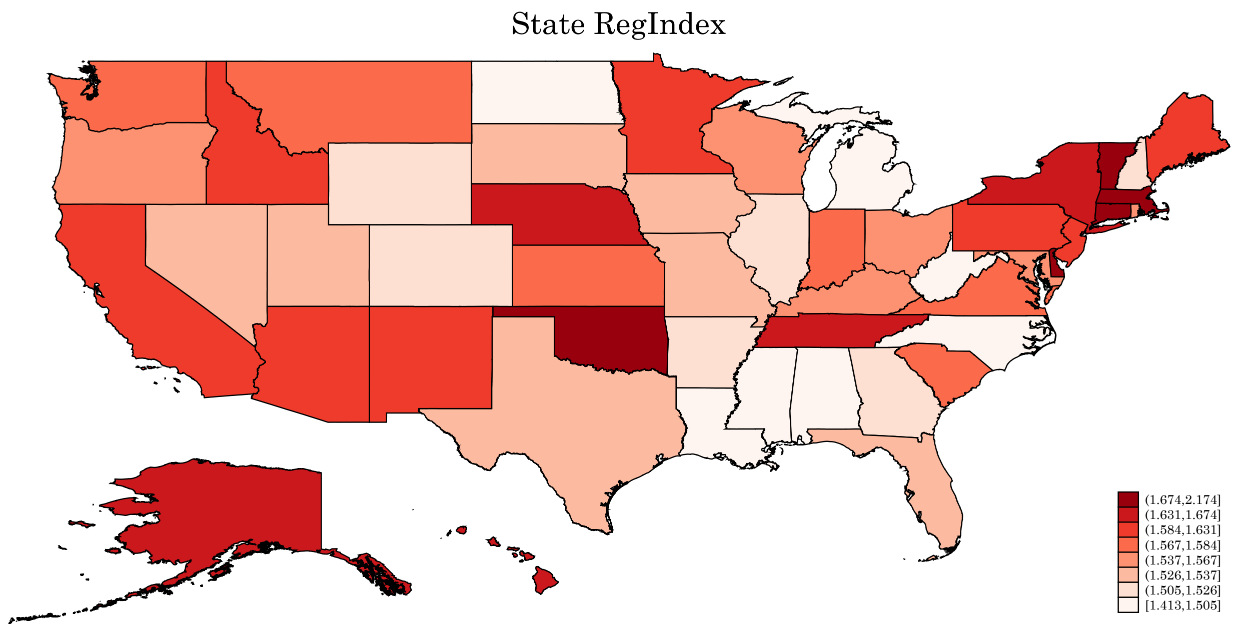
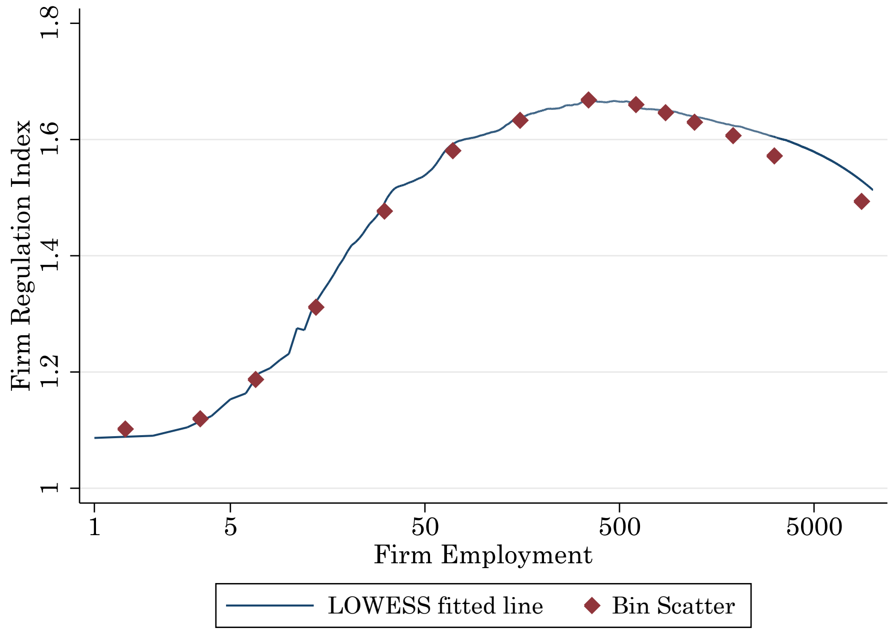


Figure 5: **RegIndex and Employment**

This figure plots the relation of RegIndex and employment for firms in Panel A and establishments in Panel B. The dots represent the average RegIndex in each employment bin, where the bins are [1, 2], [3, 4], [5, 9], [10, 19], [20, 49], [50, 99], [100, 249], [250, 499], [500, 749], [750, 999], [1000, 1499], [1500, 2499], [2500, 3999], and above 4000. The line represents the LOWESS smoothed fitted curve using the bandwidth of 0.05. The x-axis is in log scales.

Panel A: Firm-Level



Panel B: Establishment-Level

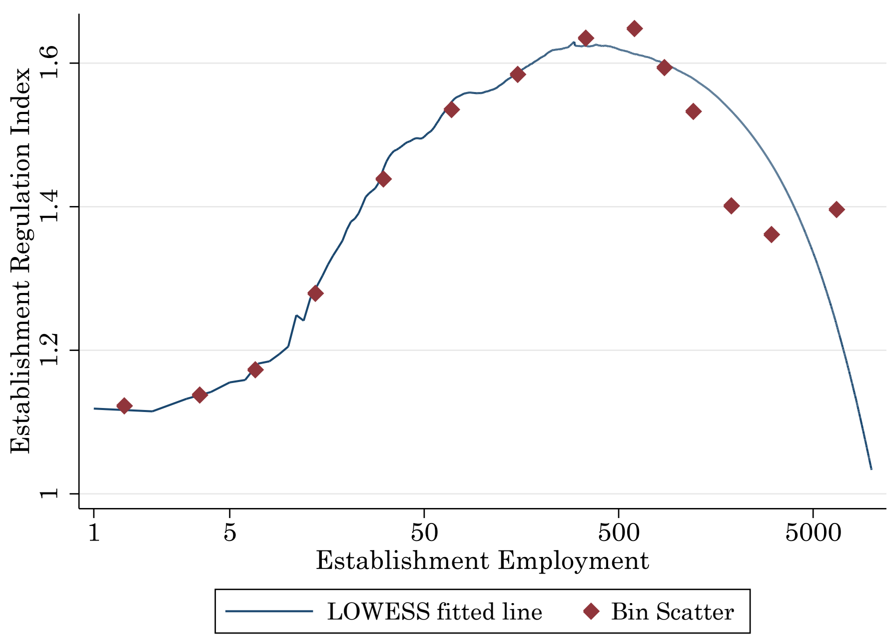


Figure 6: **Firm RegIndex and Changes in Firm Size Distribution by Sector**

The scatter plot shows the relation of RegIndex and firm employment in each NAICS 1-digit sector. Each dot represents the average RegIndex in each employment bin, where the bins are [1, 2], [3, 4], [5, 9], [10, 19], [20, 49], [50, 99], [100, 249], [250, 499], [500, 749], [750, 999], [1000, 1499], [1500, 2499], [2500, 3999], and above 4000. The x-axis is in log scales. The navy line which corresponds to the right y-axis shows the change in firms' size distribution from 2002 to 2014. We compute a firm's sampling weight as the minimum sampling weight of its establishments. We further estimate the following regression using RegIndex and changes in distribution in 14 employment bins (l) of the 8 sectors (j): $\Delta FirmShare_{l,j} = \beta RegIndex_{l,j} + SectorFE + \epsilon_{l,j}$, and we obtain an estimate of $\beta = -0.017$ with a t-statistics of -3.16 based on standard errors clustered at the sector level.

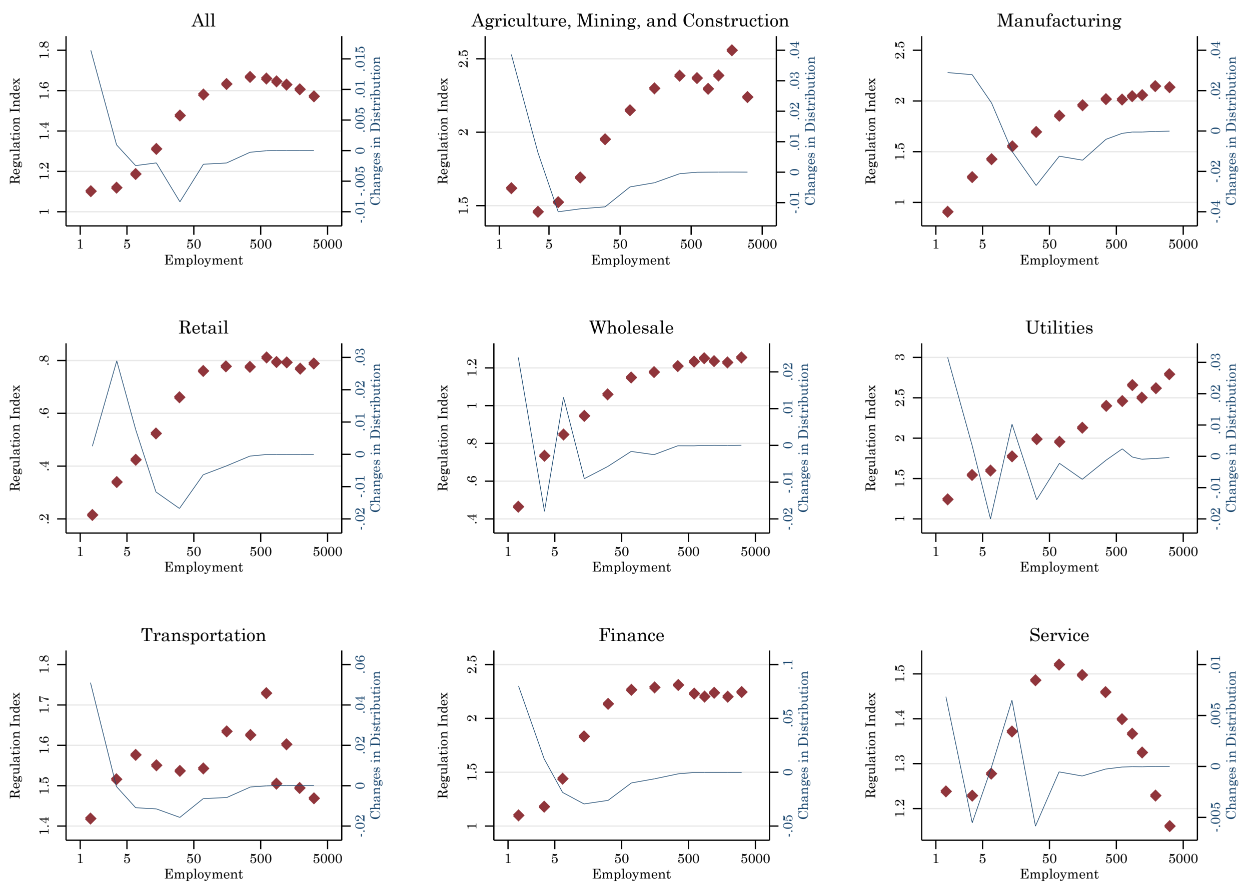


Figure 7: **Establishment RegIndex and Changes in Establishment Size Distribution by Sector**

The scatter plot shows the relation of RegIndex and establishment employment in each NAICS 1-digit sector. Each dot represents the average RegIndex in each employment bin, where the bins are [1, 2], [3, 4], [5, 9], [10, 19], [20, 49], [50, 99], [100, 249], [250, 499], [500, 749], [750, 999], [1000, 1499], [1500, 2499], [2500, 3999], and above 4000. The x-axis is in log scales. The navy line which corresponds to the right y-axis shows the change in establishments' size distribution from 2002 to 2014. We further estimate the following regression using RegIndex and changes in distribution in 14 employment bins (l) of the 8 sectors (j): $\Delta EstShare_{l,j} = \beta RegIndex_{l,j} + SectorFE + \epsilon_{l,j}$, and we obtain an estimate of $\beta = -0.016$ with a t-statistics of -3.67 based on standard errors clustered at the sector level.

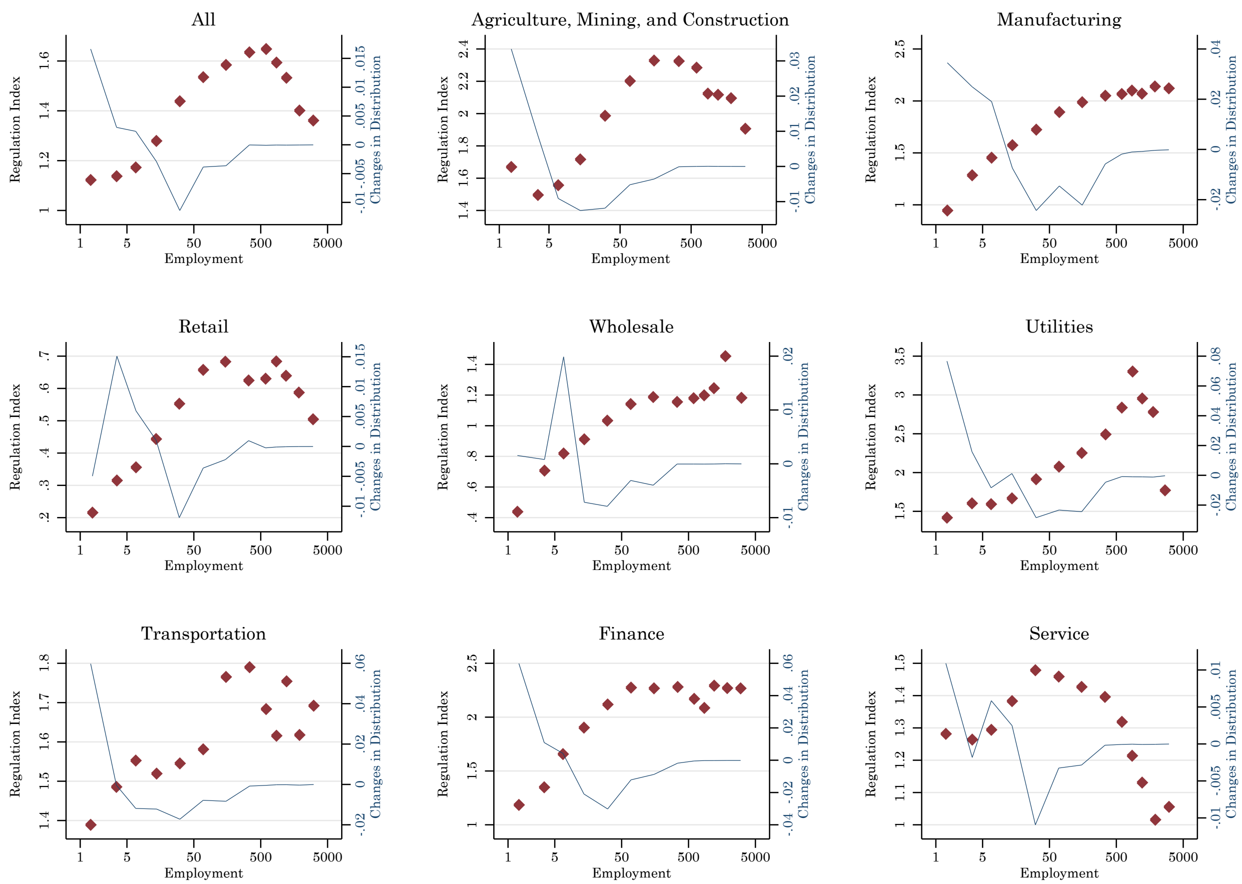


Figure 8: **Shocks to Enforcement and Regulatory Requirements by Agency**

This figure plots the shocks to an agency's regulatory enforcement and regulation requirements. Enforcement shocks are measured by the 3-year growth rate of the agency's regulation-related employment. Regulation-requirement shocks are measured by the 3-year growth rate of the agency's estimated compliance hours of its regulations excluding adjustments. See Section 5 for more details.

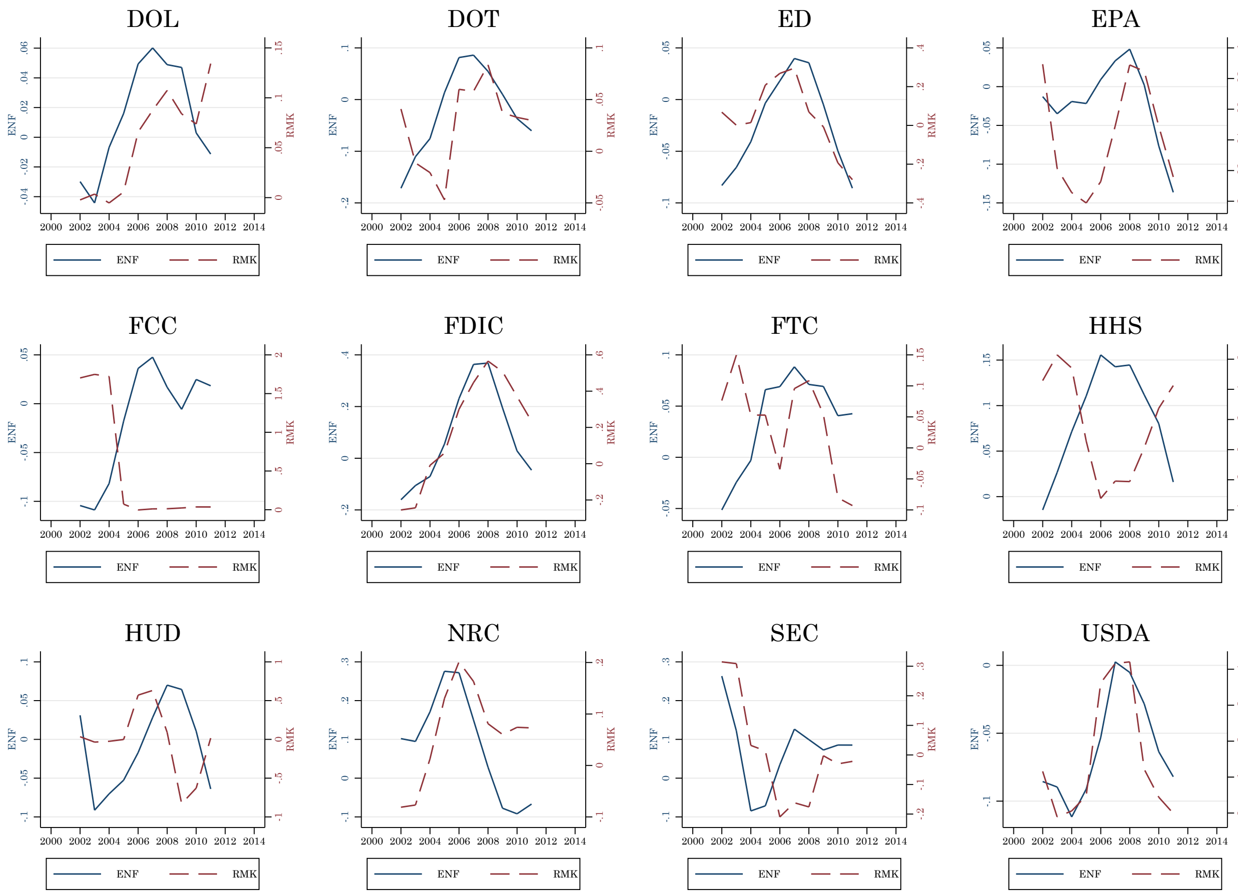
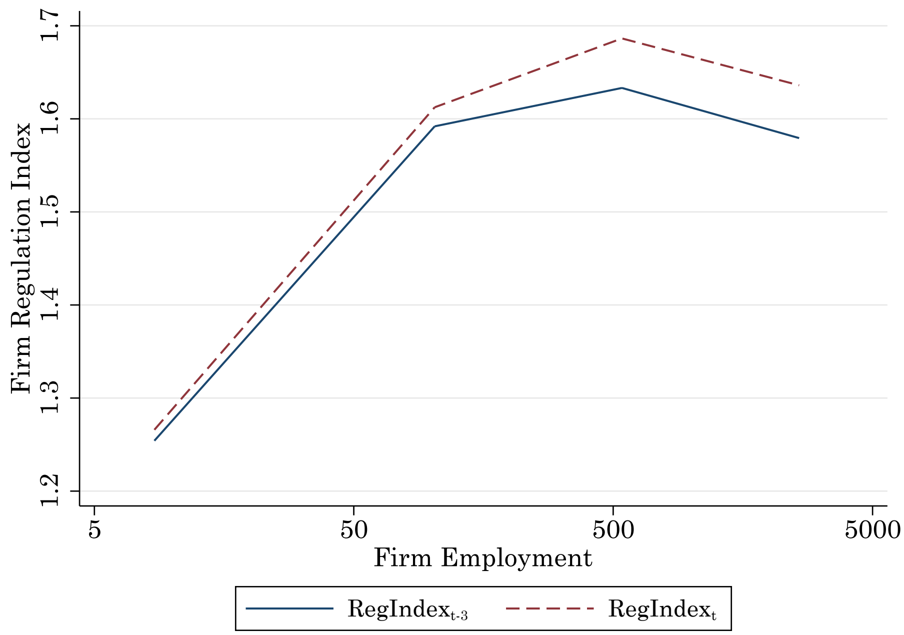


Figure 9: **Changes in Firm RegIndex by Employment Bins**

Panel A plots the average RegIndex for same set of firms at $t-3$ and t, where firms are grouped by their employment at $t-3$ into four bins: [1, 19], [20, 399], [400, 749], and above 750. Panel B plots the average of 3-year changes in log RegIndex of the firms. The sample includes firms that have occupational employment at both $t-3$ and t from 2005 to 2014. The x-axis is in log scales.

Panel A: $RegIndex_{t-3}$ **and** $RegIndex_t$



Panel B: $\Delta Log(RegIndex_t)$

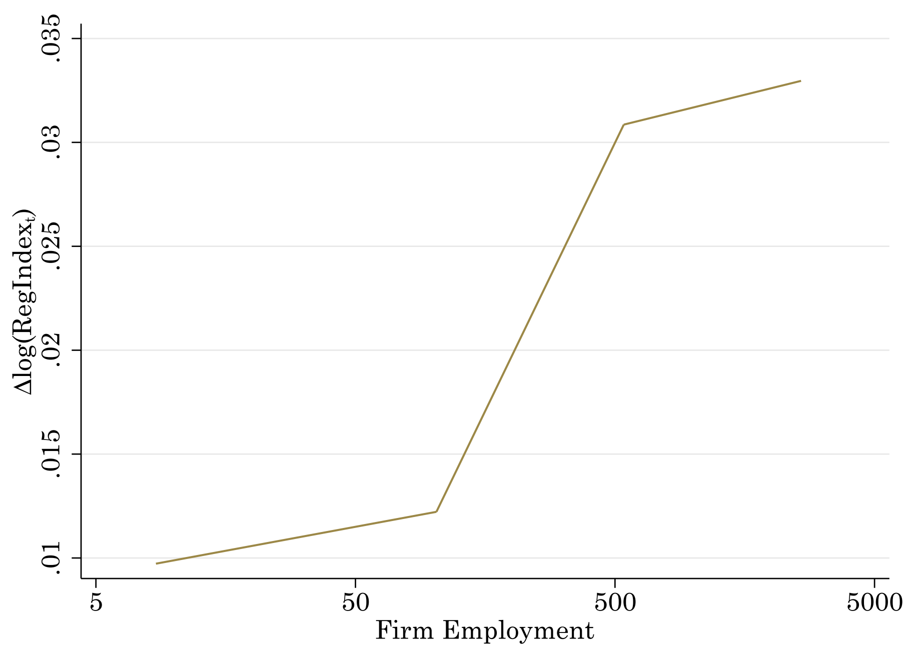
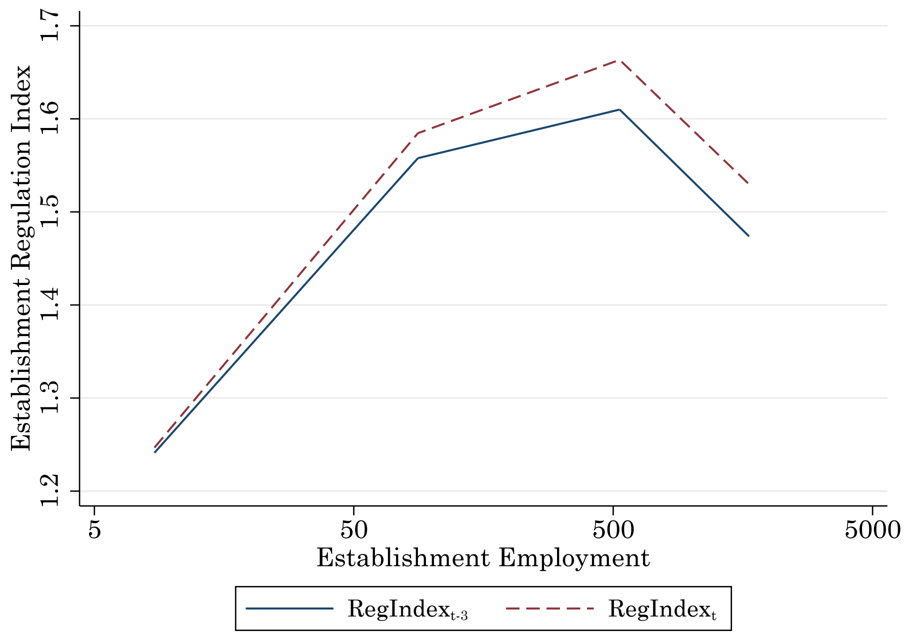


Figure 10: **Changes in Establishment RegIndex by Employment Bins**

Panel A plots the average RegIndex for same set of establishments at $t-3$ and t, where establishments are grouped by their employment at $t-3$ into four bins: [1, 19], [20, 399], [400, 749], and above 750. Panel B plots the average of 3-year changes in log RegIndex of the establishments. The sample includes establishments that have occupational employment at both $t-3$ and t from 2005 to 2014. The x-axis is in log scales.

Panel A: $RegIndex_{t-3}$ **and** $RegIndex_t$



Panel B: $\Delta Log(RegIndex_t)$

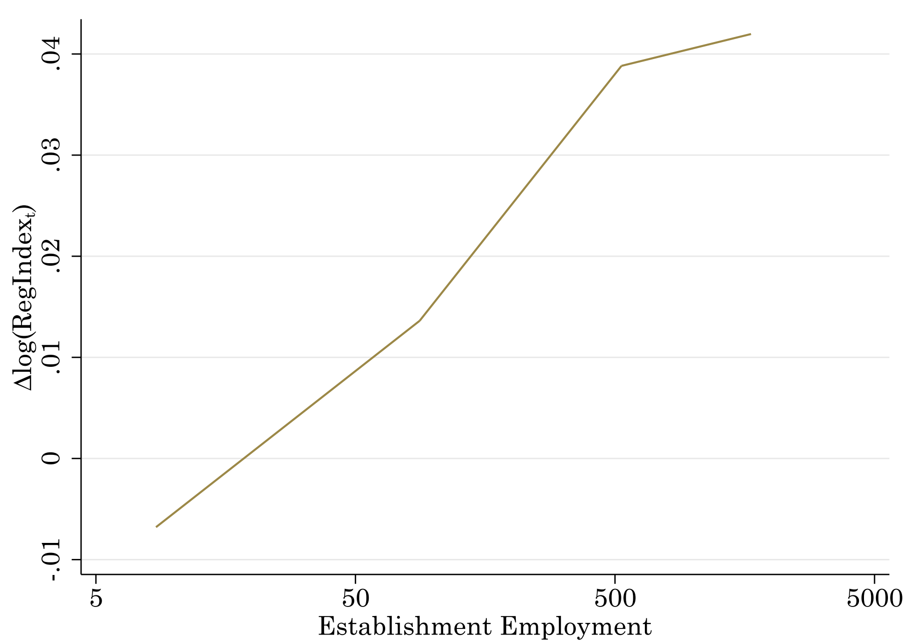


Table 1: **Examples of Regulation-related Tasks**

This table lists 10 regulation-related tasks from the O*Net database. See Section 3 for our definition of regulation-related tasks. *Import.* is the importance rating of the task for the occupation ranging from 1 to 5 provided by the O*Net database.

Occupation and Task	Import.
Construction Managers	
Inspect or review projects to monitor compliance with building and safety codes or other regulations.	3.91
Agricultural Inspectors	
Inspect agricultural commodities or related operations, as well as fish or logging operations, for compliance with laws and regulations governing health, quality, and safety.	4.59
Construction and Building Inspectors	
Evaluate project details to ensure adherence to environmental regulations.	4.12
Financial Examiners	
Establish guidelines for procedures and policies that comply with new and revised regulations and direct their implementation.	3.69
Industrial Engineering Technologists and Technicians	
Monitor environmental management systems for compliance with environmental policies, programs, or regulations.	2.67
Occupational Health and Safety Specialists	
Inspect or evaluate workplace environments, equipment, or practices to ensure compliance with safety standards and government regulations.	4.21
Urban and Regional Planners	
Determine the effects of regulatory limitations on land use projects.	4.00
Aircraft Mechanics and Service Technicians	
Conduct routine and special inspections as required by regulations.	4.49
Food Service Managers	
Monitor compliance with health and fire regulations regarding food preparation and serving, and building maintenance in lodging and dining facilities.	4.45
Compensation and Benefits Managers	
Fulfill all reporting requirements of all relevant government rules and regulations, including the Employee Retirement Income Security Act (ERISA).	4.35

Table 2: **Top 25 Occupations with the Highest Regulation-Task Intensity**

This table reports the top 25 SOC 6-digit occupations with the highest regulation-task intensity (RTI). See Section 3 for the construction of RTI.

SOC	Occupation Title	RTI
13-1041	Compliance Officers	0.343
47-4011	Construction and Building Inspectors	0.340
45-2011	Agricultural Inspectors	0.278
17-3026	Industrial Engineering Technicians	0.262
13-2061	Financial Examiners	0.256
19-3051	Urban and Regional Planners	0.229
33-2021	Fire Inspectors and Investigators	0.223
23-1011	Lawyers	0.204
17-2081	Environmental Engineers	0.189
19-2041	Environmental Scientists and Specialists, Including Health	0.183
19-3011	Economists	0.180
19-1012	Food Scientists and Technologists	0.176
43-4031	Court, Municipal, and License Clerks	0.156
33-1021	First-Line Supervisors of Fire Fighting and Prevention Workers	0.154
17-2111	Health and Safety Engineers, Ex. Mining Safety Engineers & Inspectors	0.152
53-6051	Transportation Inspectors	0.140
35-1011	Chefs and Head Cooks	0.134
19-3094	Political Scientists	0.132
13-2082	Tax Preparers	0.130
29-9012	Occupational Health and Safety Technicians	0.129
33-3051	Police and Sheriff's Patrol Officers	0.121
33-9091	Crossing Guards	0.119
11-9151	Social and Community Service Managers	0.119
11-9021	Construction Managers	0.117
33-3041	Parking Enforcement Workers	0.117

Table 3: **Top 25 Industries with the Highest Regulation Index**

This table reports the top 25 NAICS 3-digit industries with the highest regulation-index (RegIndex). See Section 3 for the construction of RegIndex.

NAICS	Industry Title	RegIndex (%)
485	Transit and Ground Passenger Transportation	3.930
525	Funds, Trusts, and Other Financial Vehicles	3.359
325	Chemical Manufacturing	3.274
324	Petroleum and Coal Products Manufacturing	2.992
551	Management of Companies and Enterprises	2.882
523	Securities, Commodity Contracts, and Other Financial Related Activities	2.734
221	Utilities	2.733
211	Oil and Gas Extraction	2.705
483	Water Transportation	2.628
236	Construction of Buildings	2.624
334	Computer and Electronic Product Manufacturing	2.621
486	Pipeline Transportation	2.594
813	Religious, Grantmaking, Civic, Professional, and Similar Organizations	2.565
238	Specialty Trade Contractors	2.511
336	Transportation Equipment Manufacturing	2.470
533	Lessors of Nonfinancial Intangible Assets (except Copyrighted Works)	2.460
237	Heavy and Civil Engineering Construction	2.452
531	Real Estate	2.430
522	Credit Intermediation and Related Activities	2.414
482	Rail Transportation	2.230
327	Nonmetallic Mineral Product Manufacturing	2.180
212	Mining (except Oil and Gas)	2.164
313	Textile Mills	2.153
326	Plastics and Rubber Products Manufacturing	2.139
562	Waste Management and Remediation Services	2.114

Table 4: **Summary Statistics**

This table provides the summary statistics of firms, establishments, and industries in our analyses. Firms are defined by employer identification numbers (EINs) following Song et al. (2018). We bundle establishments of an EIN surveyed in year $t-2$ to t as a firm in year t following the convention of the Bureau of Labor Statistics. Industries are defined at the NAICS 6-digit level. We aggregate establishments of an industry surveyed in year $t-2$ to t weighted by the establishments' sampling weights to compute the industry-level metrics in year t. RegIndex is the ratio of labor spending on regulation-related tasks and total labor spending in percentage. See Section 3 for the construction of RegIndex. The sample period is from 2002 to 2014.

Variable	Mean	SD	P0.5	Median	P99.5	Obs.
				Panel A: Firms		
Employment	92.16	617.16	1.00	13.00	2,465.00	3,027,680
Annual Wage ($ mn)	4.07	31.30	0.02	0.46	115.48	3,027,680
RegIndex	1.34	1.88	0	0.86	10.46	3,027,680
				Panel B: Establishments		
Employment	47.79	192.45	1.00	13.00	875.00	3,364,336
Annual Wage ($ mn)	2.09	11.73	0.02	0.44	43.31	3,364,336
RegIndex	1.31	1.90	0	0.80	10.57	3,364,336
				Panel C: Industry		
Employment (1,000)	90.66	285.44	0.01	25.13	2,041.20	15,159
Annual Wage ($ mn)	3,611.91	11,112.21	0.12	1,001.32	67,466.05	15,159
RegIndex	1.66	1.02	0	1.60	5.58	15,159

Table 5: **Validation: Case Studies of Industry Regulatory Shocks**

This table reports the response of establishments' RegIndex to major industry regulatory shocks in five case studies. Section 3 provides the details of each case study. *Treated* is a dummy variable that equals 1 if the industry is treated by the shock and 0 if not. *Post* is a dummy variable that equals 1 if the year is after the law enactment year and 0 if prior to. We exclude the law enactment year. All standard errors are double clustered at year and NAICS 6-digit industry. Each observation is weighted by a product of the establishment's weight assigned by the OEWS survey and the establishment's total annual wage payment. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Treated Ind:	Credit Cards	Oil and Gas	Oil and Gas	Colleges	Hospitals
Case:	CARD Act 2009	EPAct 2005	Executive Order 2010	Dear Colleague 2011	ACA 2010
	(1)	(2)	(3)	(4)	(5)
Treated × Post	0.423***	-0.521***	0.569***	0.119*	0.062***
	(0.116)	(0.090)	(0.153)	(0.053)	(0.006)
Treated	0.257	1.413**	-1.839***	-0.286**	0.606***
	(0.195)	(0.419)	(0.330)	(0.107)	(0.004)
Year FE	Yes	Yes	Yes	Yes	Yes
Establishment FE	Yes	Yes	Yes	Yes	Yes
Observations	10,082	3,140	5,877	23,319	25,043
Adjusted R^2	0.364	0.441	0.267	0.427	0.398

Table 6: **Validation: State Voting Share for Republican Party and RegIndex**

This table reports the results of regressing of state voting share for Republican Party in the 2016 presidential election, the 2016 house delegation elections, and the 2017-18 senate elections on states' 2014 RegIndex. See Figure 4 for the estimation of states' 2014 RegIndex. Column (1) includes 50 states and the District of Columbia while Columns (2) and (3) only include the 50 states. Robust standard errors are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

	State Voting Share for Republican Party in 2016 Presidential Election (1)	State Voting Share for Republican Party in 2016 House Delegation Elections (2)	State Voting Share for Republican Party in 2017-18 Senate Elections (3)
State RegIndex	-0.640***	-0.966***	-1.828***
	(0.083)	(0.242)	(0.467)
Constant	1.502***	2.031***	3.408***
	(0.132)	(0.376)	(0.741)
Observations	51	50	50
Adjusted R^2	0.411	0.315	0.119

Table 7: **Economies of Scale for RegIndex**

Panel A reports the results of regressing firms' RegIndex on their employment and employment squared, where RegIndex is at percentage and employment is at 1,000s. Panel B reports the results at the establishment level. All standard errors are double clustered at year and firm in Panel A and at year and establishment in Panel B. Industry is defined at NAICS6. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively. Given a quadratic formula $y = ax^2 + bx + c$, the *max* is computed as $c - \frac{b^2}{4a}$, while the *argmax* is computed as $-\frac{b}{2a}$. The sample period is from 2002 to 2014.

Panel A: Firm-Level					
	(1)	(2)	(3)	(4)	(5)
Emp	2.897***	2.920***	2.008***	1.935***	0.544***
	(0.065)	(0.065)	(0.076)	(0.068)	(0.074)
Emp^2	-2.902***	-2.927***	-1.963***	-1.909***	-0.542***
	(0.068)	(0.068)	(0.073)	(0.064)	(0.069)
max	1.961***	1.965***	1.782***	1.755***	1.517***
	(0.028)	(0.015)	(0.018)	(0.016)	(0.018)
argmax	0.499***	0.499***	0.511***	0.507***	0.501***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.018)
Year FE	-	Yes	-	-	Yes
Year-Ind FE	-	-	Yes	-	-
Year-Ind-State FE	-	-	-	Yes	-
Firm FE	-	-	-	-	Yes
Observations	3,027,680	3,027,680	3,027,241	2,918,296	2,162,080
Adjusted R^2	0.007	0.007	0.378	0.412	0.597

Panel B: Establishment-Level					
	(1)	(2)	(3)	(4)	(5)
Emp	3.607***	3.788***	2.985***	2.902***	0.612***
	(0.265)	(0.217)	(0.148)	(0.134)	(0.125)
Emp^2	-5.255***	-5.510***	-4.452***	-4.397***	-0.992***
	(0.486)	(0.430)	(0.284)	(0.259)	(0.197)
max	1.823***	1.850***	1.724***	1.699***	1.444***
	(0.041)	(0.019)	(0.016)	(0.014)	(0.016)
argmax	0.343***	0.344***	0.335***	0.330***	0.309***
	(0.007)	(0.007)	(0.006)	(0.005)	(0.012)
Year FE	-	Yes	-	-	Yes
Year-Ind FE	-	-	Yes	-	-
Year-Ind-State FE	-	-	-	Yes	-
Establishment FE	-	-	-	-	Yes
Observations	3,362,824	3,362,824	3,362,418	3,255,415	2,194,239
Adjusted R^2	0.005	0.006	0.371	0.408	0.534

Table 8: **Economies of Scale for Regulation Intensity Within Occupation**

This table reports the robustness of Table 7 by showing the relationship between firms' requirements on regulatory compliance skills and their size *within an occupation.* To do so, we name match firms in our OEWS sample to firms in the Burning Glass Technologies (BGT) data, which provide 14 million job postings from 2010 to 2014 (the overlapping period between Burning Glass and our OEWS sample). We next measure each job posting's textual content's requirement on regulation-related tasks by applying our exact methodology in Section 4 on BGT's definition of "skills." We take a simple average of the regulation-related dummy for each skill within a job posting to measure the job posting's average regulation relatedness (in percentage). We then regress the job posting's regulation relatedness on the firm's employment (in thousands) and the squared of the employment, while controlling for year-occupation fixed effects. Column (1) reports results for all SOC 6-digit occupations. Column (2) requires results for occupations that have positive regulation-task intensity (RTI) in our definition in Section 3 (see Table 2), while Column (3) reports the results for occupations with RTI equal to 0. All standard errors are clustered at the firm (EIN) level. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively. Given a quadratic formula $y = ax^2 + bx + c$, the *max* is computed as $c - \frac{b^2}{4a}$, while the *argmax* is computed as $-\frac{b}{2a}$.

	All Occ (1)	Occ (RTI>0) (2)	Occ (RTI = 0) (3)
Emp	0.722***	0.536***	0.858***
	(0.204)	(0.168)	(0.290)
Emp^2	-0.445***	-0.341***	-0.520***
	(0.131)	(0.102)	(0.181)
max	1.558***	2.073***	1.188***
	(0.092)	(0.071)	(0.128)
argmax	0.811***	0.787***	0.825***
	(0.048)	(0.048)	(0.059)
Year-OCC FE	Yes	Yes	Yes
Observations	14,052,988	5,878,039	8,174,949
Adjusted R^2	0.139	0.138	0.119

Table 9: **Top 3 Industries for Each Regulatory Agency**

This table reports each regulatory agency's top 3 most exposed industries, where industry is defined at the NAICS 3-digit level. r_k is the ratio of the industry's labor spending on agency k's regulation-related tasks and total labor spending. RegIndex is the ratio of the industry's labor spending on all regulation-related tasks and total labor spending. See Section 5 for details.

Agency	Rank	NAICS3	Title	r_k/RegIndex
USDA	1	722	Food Services and Drinking Places	0.0707
USDA	2	311	Food Manufacturing	0.0634
USDA	3	115	Support Activities for Agriculture and Forestry	0.0621
DOT	1	492	Couriers and Messengers	0.2298
DOT	2	485	Transit and Ground Passenger Transportation	0.2196
DOT	3	482	Rail Transportation	0.1721
EPA	1	562	Waste Management and Remediation Services	0.1775
EPA	2	324	Petroleum and Coal Products Manufacturing	0.1736
EPA	3	236	Construction of Buildings	0.1651
FCC	1	512	Motion Picture and Sound Recording Industries	0.1220
FCC	2	492	Couriers and Messengers	0.1064
FCC	3	515	Broadcasting (except Internet)	0.1054
FDIC	1	523	Securities, Commodity Contracts, and Other Financial Activities	0.1738
FDIC	2	522	Credit Intermediation and Related Activities	0.1726
FDIC	3	525	Funds, Trusts, and Other Financial Vehicles	0.1490
HHS	1	446	Health and Personal Care Stores	0.2442
HHS	2	621	Ambulatory Health Care Services	0.2400
HHS	3	622	Hospitals	0.2166
HUD	1	531	Real Estate	0.1808
HUD	2	236	Construction of Buildings	0.1562
HUD	3	238	Specialty Trade Contractors	0.1361
FTC	1	313	Textile Mills	0.1213
FTC	2	315	Apparel Manufacturing	0.1205
FTC	3	314	Textile Product Mills	0.1176
NRC	1	221	Utilities	0.1041
NRC	2	325	Chemical Manufacturing	0.1006
NRC	3	562	Waste Management and Remediation Services	0.0928
ED	1	624	Social Assistance	0.1172
ED	2	485	Transit and Ground Passenger Transportation	0.1107
ED	3	713	Amusement, Gambling, and Recreation Industries	0.1050
DOL	1	113	Forestry and Logging	0.1487
DOL	2	813	Religious, Grant-making, Civic, Professional, Similar Organizations	0.1480
DOL	3	448	Clothing and Clothing Accessories Stores	0.1477
SEC	1	523	Securities, Commodity Contracts, and Other Financial Activities	0.1963
SEC	2	522	Credit Intermediation and Related Activities	0.1786
SEC	3	525	Funds, Trusts, and Other Financial Vehicles	0.1638

Table 10: **Correlation of Instrumental Variables**

This table reports the Pearson correlation between the instrumental variable for enforcement shocks, $iv(\Delta\log(p_{it}))$, and the instrumental variable for regulatory-requirement shocks, $iv(\Delta\log(\tilde{R}_{it}))$, in full sample and in each NAICS 1-digit sector. The full firm-level sample includes 608,500 observations that have 3-year changes in log RegIndex. The full establishment-level sample includes 628,733 observations that have 3-year changes in log RegIndex. Section 5 provides more details on the construction of the instrumental variables.

	Firm-Level Sample	Establishment-Level Sample
All Sectors	0.104	0.116
Agriculture, Mining, and Construction	0.159	0.157
Manufacturing	0.007	-0.084
Retail	-0.183	-0.219
Wholesale	-0.127	-0.131
Utilities	0.024	-0.049
Transportation	-0.092	-0.139
Finance	0.167	0.139
Service	0.077	0.123

Table 11: **Enforcement vs. Regulatory Requirement: An Instrumental Variable Approach**

This table reports the results of regressing 3-year changes in a firm's log RegIndex on instrumental variables. Equation (6) provides the regression specification. Industry is defined at the NAICS 6-digit level. $iv(\Delta\log(p_{it}))$ and $iv(\Delta\log(\tilde{R}_{it}))$ are the instrumental variables for the industry's enforcement shocks and regulatory-requirement shocks. Section 5 provides more details on the construction of the instrumental variables. $\Delta\log$(Wage) is the 3-year changes in the log total wage spending of the firm. $\Delta\log(w^r)$ is the 3-year changes in the log wage rate of the regulation-related tasks. Columns (1)-(3) show results at the firm level, while Columns (4)-(6) show results at the establishment level. All regressions control for year fixed effects and beginning of period log RegIndex, log(RegIndex), to account for mean-reversion of firms' or establishments' RegIndex. All variables are standardized to have mean 0 and variance of 1 for the ease of interpretation. Standard errors are double clustered by industry and year and presented in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively. The sample period is from 2002 to 2014.

	Firm-Level			Establishment-Level		
	(1)	(2)	(3)	(4)	(5)	(6)
$iv(\Delta\log(p_{it}))$	0.209**		0.112	0.234***		0.128*
	(0.077)		(0.075)	(0.063)		(0.065)
$iv(\Delta\log(\tilde{R}_{it}))$		0.257***	0.228***		0.269***	0.231***
		(0.038)	(0.054)		(0.036)	(0.052)
$\Delta\log$(Wage)	-0.051***	-0.049***	-0.049***	-0.083***	-0.081***	-0.080***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)	(0.008)
$\Delta\log(w^r)$	0.290***	0.282***	0.280***	0.277***	0.270***	0.268***
	(0.010)	(0.012)	(0.012)	(0.011)	(0.013)	(0.013)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	608,500	608,500	608,500	628,733	628,733	628,733
Adjusted R^2	0.322	0.340	0.344	0.323	0.340	0.345

Table 12: **Enforcement vs. Regulatory Requirements in Subsamples by Size**

This table reports the results of regressing 3-year changes in a firm's log RegIndex on two instrumental variables, $iv(\Delta\log(p_{it}))$ and $iv(\Delta\log(\tilde{R}_{it}))$ in four subsamples of firms by employment. Equation (6) provides the regression specification. See Table 11 for variable definitions. Section 5 provides more details on the construction of the instrumental variables. Columns (1)-(4) show results at the firm level while Columns (5)-(8) show results at the establishment level. All regressions control for Δlog(Wage), $\Delta\log(w^r)$, log(RegIndex), and year fixed effects. All variables are standardized to have mean 0 and variance of 1 for the ease of interpretation. Standard errors are double clustered by industry and year and presented in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively. The sample period is from 2002 to 2014.

	Firm-Level				Establishment-Level			
	1-19	20-399	400-749	$\geq$ 750	1-19	20-399	400-749	$\geq$ 750
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$iv(\Delta\log(p_{it}))$	0.102	0.120	0.099	0.083	0.115*	0.140*	0.168**	0.077
	(0.075)	(0.078)	(0.076)	(0.064)	(0.062)	(0.066)	(0.074)	(0.066)
$iv(\Delta\log(\tilde{R}_{it}))$	0.169**	0.258***	0.289***	0.309***	0.170**	0.271***	0.313***	0.367***
	(0.056)	(0.054)	(0.054)	(0.046)	(0.053)	(0.050)	(0.053)	(0.055)
Δlog(Wage)	-0.139***	0.002	0.047***	0.028**	-0.164***	-0.024**	0.030**	0.001
	(0.008)	(0.008)	(0.010)	(0.012)	(0.007)	(0.008)	(0.013)	(0.015)
$\Delta\log(w^r)$	0.353***	0.237***	0.182***	0.230***	0.355***	0.215***	0.107***	0.122***
	(0.011)	(0.013)	(0.017)	(0.020)	(0.011)	(0.013)	(0.021)	(0.026)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	189,404	352,779	29,482	36,835	220,464	375,485	19,622	13,162
Adjusted R^2	0.400	0.324	0.243	0.214	0.397	0.321	0.265	0.233

Table 13: **Enforcement vs. Regulatory Requirements for Firms in Each Sector**

This table reports the results of regressing 3-year changes in a firm's log RegIndex on instrumental variables in Table 11 and 12 in each NAICS 1-digit sector. Equation (6) provides the regression specification. See Table 11 and 12 for variable definitions. All regressions control for Δlog(Wage), $\Delta\log(w^r)$, log(RegIndex), and year fixed effects. All variables are standardized to have mean 0 and variance of 1 for the ease of interpretation. Standard errors are double clustered by industry and year and presented in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively. The sample period is from 2002 to 2014.

	All Sizes			Subsample by Firm Size			
				1-19	20-399	400-749	$\geq$ 750
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Agriculture, Mining, and Construction			
$iv(\Delta\log(p_{it}))$	0.190**		0.106	0.103	0.106	0.068	0.213**
	(0.078)		(0.078)	(0.081)	(0.075)	(0.103)	(0.082)
$iv(\Delta\log(\tilde{R}_{it}))$		0.269***	0.243***	0.196**	0.276***	0.279**	0.207*
		(0.044)	(0.059)	(0.061)	(0.059)	(0.094)	(0.103)
				Manufacturing			
$iv(\Delta\log(p_{it}))$	0.349***		0.183*	0.120	0.190*	0.243**	0.234*
	(0.087)		(0.087)	(0.068)	(0.094)	(0.104)	(0.103)
$iv(\Delta\log(\tilde{R}_{it}))$		0.331***	0.285***	0.241***	0.307***	0.323***	0.411***
		(0.057)	(0.047)	(0.044)	(0.052)	(0.048)	(0.043)
				Retail			
$iv(\Delta\log(p_{it}))$	0.132***		0.097**	0.088*	0.127**	0.062	0.033
	(0.040)		(0.033)	(0.039)	(0.044)	(0.047)	(0.038)
$iv(\Delta\log(\tilde{R}_{it}))$		0.134***	0.120***	0.122***	0.128***	0.099**	0.103***
		(0.028)	(0.025)	(0.022)	(0.032)	(0.034)	(0.022)
				Wholesale			
$iv(\Delta\log(p_{it}))$	0.214***		0.125**	0.076*	0.164**	0.096	0.200
	(0.057)		(0.042)	(0.034)	(0.059)	(0.103)	(0.110)
$iv(\Delta\log(\tilde{R}_{it}))$		0.194***	0.169***	0.125***	0.194***	0.285***	0.272***
		(0.038)	(0.032)	(0.035)	(0.039)	(0.074)	(0.070)

Table 13: **Enforcement vs. Regulatory Requirements for Firms by Sector** *–Continued*

This table reports the results of regressing 3-year changes in a firm's log RegIndex on instrumental variables in Table 11 and 12 in each NAICS 1-digit sector. Equation (6) provides the regression specification. See Table 11 and 12 for variable definitions. All regressions control for $\Delta\log(\text{Wage})$, $\Delta\log(w^r)$, log(RegIndex), and year fixed effects. All variables are standardized to have mean 0 and variance of 1 for the ease of interpretation. Standard errors are double clustered by industry and year and presented in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively. The sample period is from 2002 to 2014.

	All Sizes			Subsample by Firm Size			
				1-19	20-399	400-749	$\geq$ 750
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Utilities			
$iv(\Delta\log(p_{it}))$	0.435***		0.196*	0.123*	0.210	-0.053	0.140
	(0.127)		(0.103)	(0.055)	(0.167)	(0.207)	(0.116)
$iv(\Delta\log(\tilde{R}_{it}))$		0.443***	0.373**	0.544***	0.254	0.539***	0.378***
		(0.130)	(0.120)	(0.120)	(0.201)	(0.044)	(0.046)
				Transportation			
$iv(\Delta\log(p_{it}))$	0.167		0.100	0.059	0.105	0.129*	0.184**
	(0.101)		(0.067)	(0.062)	(0.071)	(0.058)	(0.071)
$iv(\Delta\log(\tilde{R}_{it}))$		0.374***	0.360***	0.310***	0.367***	0.448***	0.514***
		(0.037)	(0.030)	(0.037)	(0.028)	(0.065)	(0.058)
				Finance			
$iv(\Delta\log(p_{it}))$	0.380***		0.190**	0.191*	0.192**	0.211	0.088
	(0.053)		(0.075)	(0.093)	(0.080)	(0.143)	(0.108)
$iv(\Delta\log(\tilde{R}_{it}))$		0.304***	0.239**	0.206*	0.275***	0.222***	0.336**
		(0.067)	(0.079)	(0.093)	(0.067)	(0.061)	(0.112)
				Service			
$iv(\Delta\log(p_{it}))$	0.196**		0.097	0.101	0.096	0.055	0.034
	(0.078)		(0.082)	(0.081)	(0.081)	(0.066)	(0.062)
$iv(\Delta\log(\tilde{R}_{it}))$		0.251***	0.222***	0.169**	0.258***	0.269***	0.279***
		(0.040)	(0.058)	(0.058)	(0.057)	(0.045)	(0.047)

Table 14: **Enforcement vs. Regulatory Requirements for Establishments by Sector**

This table reports the results of regressing 3-year changes in an establishment's log RegIndex on instrumental variables in Table 11 and 12 in each NAICS 1-digit sector. Equation (6) provides the regression specification. See Table 11 and 12 for variable definitions. All regressions control for Δlog(Wage), $\Delta\log(w^r)$, log(RegIndex), and year fixed effects. All variables are standardized to have mean 0 and variance of 1 for the ease of interpretation. Standard errors are double clustered by industry and year and presented in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively. The sample period is from 2002 to 2014.

	All Sizes			Subsample by Establishment Size			
				1-19	20-399	400-749	$\geq$ 750
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Agriculture, Mining, and Construction			
$iv(\Delta\log(p_{it}))$	0.207**		0.118	0.127*	0.112	0.170	0.037
	(0.065)		(0.069)	(0.068)	(0.068)	(0.095)	(0.123)
$iv(\Delta\log(\tilde{R}_{it}))$		0.275***	0.243***	0.188***	0.283***	0.241*	0.162
		(0.041)	(0.054)	(0.055)	(0.053)	(0.119)	(0.112)
				Manufacturing			
$iv(\Delta\log(p_{it}))$	0.388***		0.229**	0.154**	0.242***	0.304**	0.349***
	(0.073)		(0.072)	(0.066)	(0.072)	(0.096)	(0.103)
$iv(\Delta\log(\tilde{R}_{it}))$		0.342***	0.288***	0.238***	0.312***	0.331***	0.460***
		(0.054)	(0.028)	(0.028)	(0.030)	(0.036)	(0.039)
				Retail			
$iv(\Delta\log(p_{it}))$	0.195***		0.154***	0.116**	0.188***	0.193**	0.008
	(0.054)		(0.040)	(0.041)	(0.044)	(0.071)	(0.138)
$iv(\Delta\log(\tilde{R}_{it}))$		0.189***	0.171***	0.122***	0.202***	0.287***	0.413***
		(0.049)	(0.045)	(0.031)	(0.057)	(0.072)	(0.104)
				Wholesale			
$iv(\Delta\log(p_{it}))$	0.252***		0.164***	0.112**	0.205***	0.464***	0.504***
	(0.052)		(0.039)	(0.042)	(0.048)	(0.085)	(0.120)
$iv(\Delta\log(\tilde{R}_{it}))$		0.191***	0.155***	0.128***	0.182***	0.230**	0.089**
		(0.035)	(0.025)	(0.031)	(0.027)	(0.082)	(0.032)

Table 14: **Enforcement vs. Regulatory Requirements for Establishments by Sector**–*Continued*

This table reports the results of regressing 3-year changes in an establishment's log RegIndex on instrumental variables in Table 11 and 12 in each NAICS 1-digit sector. Equation (6) provides the regression specification. See Table 11 and 12 for variable definitions. All regressions control for Δlog(Wage), $\Delta\log(w^r)$, log(RegIndex), and year fixed effects. All variables are standardized to have mean 0 and variance of 1 for the ease of interpretation. Standard errors are double clustered by industry and year and presented in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively. The sample period is from 2002 to 2014.

	All Sizes			Subsample by Establishment Size			
				1-19	20-399	400-749	$\geq$ 750
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Utilities			
$iv(\Delta\log(p_{it}))$	0.362***		0.150	0.233	0.047	0.161	0.078
	(0.075)		(0.086)	(0.141)	(0.095)	(0.272)	(0.286)
$iv(\Delta\log(\tilde{R}_{it}))$		0.373***	0.319***	0.313*	0.293***	0.058	0.293
		(0.081)	(0.075)	(0.165)	(0.068)	(0.232)	(0.223)
				Transportation			
$iv(\Delta\log(p_{it}))$	0.192*		0.134*	0.086*	0.143*	0.193*	0.284
	(0.086)		(0.062)	(0.045)	(0.068)	(0.086)	(0.167)
$iv(\Delta\log(\tilde{R}_{it}))$		0.366***	0.350***	0.294***	0.373***	0.393***	0.526***
		(0.033)	(0.023)	(0.034)	(0.023)	(0.064)	(0.066)
				Finance			
$iv(\Delta\log(p_{it}))$	0.377***		0.196**	0.199**	0.207***	0.161***	-0.073
	(0.039)		(0.062)	(0.081)	(0.063)	(0.045)	(0.132)
$iv(\Delta\log(\tilde{R}_{it}))$		0.294***	0.219**	0.216**	0.234**	0.298***	0.285***
		(0.063)	(0.080)	(0.089)	(0.078)	(0.067)	(0.075)
				Service			
$iv(\Delta\log(p_{it}))$	0.232***		0.121	0.117	0.127	0.129**	0.064
	(0.062)		(0.073)	(0.071)	(0.071)	(0.052)	(0.060)
$iv(\Delta\log(\tilde{R}_{it}))$		0.260***	0.218***	0.167**	0.259***	0.264***	0.285***
		(0.040)	(0.058)	(0.057)	(0.056)	(0.063)	(0.074)

Online Appendix to

"The Cost of Regulatory Compliance in the United States"

Francesco Trebbi Miao Ben Zhang

- Not for Publication -

A Measuring Enforcement Employment

As a robustness check to using our main measure of enforcement shock in Section 5.1, which is the based on major regulatory agencies' employment of regulation-related occupations, we construct a refined measure based on major regulatory agencies' employment of enforcement-focused regulation-related occupations. Our data as described in Section 5.1 cover detailed information of federal government employees such as their agency, occupation, and fulltime/parttime status from 2002 and 2014. Section 5.1 has described our method of identifying regulation-related occupations. We further identify "enforcement" occupations among regulation-related occupations. To do so, we first obtain each federal employee occupation's task description from "Handbook of Occupational Group and Families" at the US OPM website.[51] Then we identify an occupation as enforcement-focused if its task description includes the following keywords: "*enforcement, enforce, enforces, supervision, supervisory, monitor, monitors, oversight, oversee, oversees, sanctions, sanction, penalty, penalties, fine, fines, inspect, inspects, inspection, inspections, investigate, investigates, investigation, investigations, examine, examines, examination, examinations.*" These procedures identify 105 "enforcement" occupations out of a total of 227 regulation-related occupations.

Finally, we apply our definitions of enforcement-related regulatory occupations to the 12 agency's employment, and compute the 3-year log differences for each agency's enforcement employment, which is an alternative measure of $\Delta \log enf_{kt}$ in equation (4). Tables A.2-A.5 present results using enforcement shocks, $iv(\Delta\log(p_{it}))$, based on this alternative measure.

B Additional Tables and Figures

[51] The OPM has its own definitions for government occupations that are different from the SOC system. The handbook for OPM occupation description can be downloaded at https://www.opm.gov/policy-data-oversight/classification-qualifications/classifying-general-schedule-positions/occupationalhandbook.pdf.

Table A.1: **Robustness—Validation: State Voting Share for Republican Party and RegIndex**

This table reports the robustness check of Table 6 by controlling for states' 2017 RegData measure, and states' number of establishments in 2017. See Figure 4 for the estimation of states' 2014 RegIndex. State RegData is from QuantGov.org. States' number of establishment is from the Census Statistics of U.S. Businesses (SUSB).

	State Voting Share for Republican Party in 2016 Presidential Election		State Voting Share for Republican Party in 2016 House Delegation Elections		State Voting Share for Republican Party in 2017-18 Senate Elections	
	(1)	(2)	(3)	(4)	(5)	(6)
State RegIndex	-0.538***	-0.539***	-0.958***	-0.957***	-1.835***	-1.817***
	(0.109)	(0.109)	(0.253)	(0.254)	(0.468)	(0.478)
State RegData	-0.561***	-0.512**	-0.547***	-0.607**	-2.109***	-2.978**
	(0.102)	(0.198)	(0.137)	(0.271)	(0.597)	(1.133)
#Establishments		-0.026		0.032		0.457
		(0.082)		(0.106)		(0.472)
Constant	1.414***	1.413***	2.089***	2.090***	3.693***	3.708***
	(0.174)	(0.175)	(0.390)	(0.395)	(0.748)	(0.776)
Observations	49	49	49	49	49	49
Adjusted R^2	0.339	0.325	0.365	0.352	0.218	0.214

Table A.2: **50 Keywords for Each Regulatory Agency**

This table lists the 50 keywords we used to identify each regulatory agency. To obtain these keywords, we first extract all relevant volumes from the Code of Federal Regulations to each of the 12 agencies in Section 5. Then, we compute the term-frequency ratio for each word as the count of the word in the agency i's relevant CFR volumes over the count of that word in all 12 agencies' relevant CFR volume. This table lists the top 50 keywords with the highest term-frequency ratio for each agency.

Rank	DOL	DOT	ED	EPA	FCC	FDIC
1	labors	transported	education	environmental	communications	deposit
2	workers	traveling	educational	pollution	telecommunication	depositor
3	unemployment	transit	school	epa	transmitting	depositors
4	employer	cargo	schools	environment	telecommunications	depository
5	workforce	freight	learners	conservation	transmit	deposits
6	worker	trains	academic	epas	broadcasting	fdic
7	wages	traffic	teachers	ecological	channels	fdics
8	jobs	taxiing	diploma	emissions	radiocommunication	bank
9	employers	buses	teacher	pollutants	broadcasts	banks
10	workplaces	cargocarrying	student	eco	telephony	banking
11	employees	intercity	colleges	contamination	broadcast	insured
12	bargaining	passengers	instructional	renewable	reception	savings
13	wage	bus	literacy	recycling	channel	investments
14	job	vehicle	graduate	pollutant	conversation	dividend
15	employee	highways	students	ecosystem	signals	fdi
16	miners	driving	parents	contaminated	multichannel	fdicsupervised
17	workmens	railroads	teachout	endangerment	telegraphy	forex
18	machinery	train	vocational	ecology	transmissions	paycheck
19	occupation	taxi	graduation	chlorinated	networks	eximbank
20	subsistence	locomotive	curricula	preventable	telephones	fsi
21	welders	cargoes	tuition	greenhouse	radiotelephony	pd
22	unemployed	railroad	postsecondary	preventative	broadcaster	unfunded
23	farmworkers	cruising	undergraduate	pesticide	radio	surcharge
24	wageloss	cars	institution	pesticides	transmitters	ssfa
25	demanding	roadside	baccalaureate	chemicals	wireless	loans
26	workweek	trips	parental	cleaner	broadband	mortgagebacked
27	workday	haul	elementary	hazardous	signalling	portfolio
28	jobrelated	roadway	extracurricular	compliance	fcc	collateral
29	workrelated	passengercarrying	faculty	ordinance	modulation	unsecured
30	workdays	taxiway	achievement	contaminants	transmitter	portfolios
31	workings	towing	semester	ecosystems	interference	institutionaffiliated
32	surplus	commuter	schoolwide	habitats	cochannel	securitization
33	cutting	baggage	mathematics	containment	transceiver	loantovalue
34	employmentrelated	car	coursework	cleaners	telecommand	lending
35	recruitment	ferry	bachelors	remediation	cable	securitizations
36	welder	passenger	campus	ozone	television	brokered
37	occupations	flight	preschool	aeration	broadcastingsatellite	fdia
38	worksites	itinerary	children	wetlands	bandwidths	qfc
39	layoff	luggage	cognitive	warming	audio	dif
40	apprenticeship	congestion	enrolled	decontaminated	rf	securitized
41	contracture	fares	talent	recycled	bandwidth	safekeeping
42	erecting	highway	geography	wastewaters	decoders	gaap
43	shafting	carriage	doctoral	permits	fccs	fiduciary
44	contractorissued	routes	racial	petroleum	voip	liquidity
45	farmworker	route	childs	wastewater	messages	creditworthiness
46	jobsite	drivertrainees	scholar	biocides	stations	institution
47	economical	riding	accrediting	landfills	handsets	statelicensed
48	men	movement	athletic	wildlife	interconnected	assets
49	sickness	aboard	peer	antidegradation	interconnection	lei
50	clothing	airline	disabilities	epadc	antenna	pledged

Table A.1: **50 Keywords for Each Regulatory Agency**—*Continued*

This table lists the 50 keywords we used to identify each regulatory agency. To obtain these keywords, we first extract all relevant volumes from the Code of Federal Regulations to each of the 12 agencies in Section 5. Then, we compute the term-frequency ratio for each word as the count of the word in the agency i's relevant CFR volumes over the count of that word in all 12 agencies' relevant CFR volume. This table lists the top 50 keywords with the highest term-frequency ratio for each agency.

Rank	FTC	HHS	HUD	NRC	SEC	USDA
1	seller	health	housing	nuclear	securities	livestock
2	sellers	hospitals	dwellings	reactorrelated	brokers	grazing
3	buyers	medicine	residential	reactor	brokerdealers	tomatoes
4	marketer	healthrelated	redevelopment	reactors	securitiesxexxd	growers
5	advertised	doctors	neighborhoods	radioactive	broker	seedlings
6	gp	physician	homes	fission	investor	potatoes
7	franchised	hospital	cities	uranium	brokerdealer	seeds
8	marketers	hospitalspecific	apartment	neutron	investors	organically
9	valued	physicians	rents	plutonium	trader	grower
10	merchandise	ambulance	neighborhood	isotope	shareholders	potato
11	solicitations	inpatients	tenancy	irradiation	shareholder	berries
12	acquisitions	patients	renting	atomic	currencies	germination
13	clothes	clinicians	dwelling	tritium	accountant	varieties
14	opt	manpower	homeowners	radiation	prospectus	variety
15	franchise	hospitalization	rent	radioactivity	accountants	weed
16	pearl	nurse	condominium	irradiator	prospectuses	peanuts
17	wholesalers	hipaa	bedrooms	radionuclides	stockholder	apples
18	camera	diseases	reside	deuterium	offerings	seedling
19	apparel	clinics	households	isotopes	syndicate	pear
20	warrantor	hospices	homebuyers	radionuclide	advisers	grapes
21	advertiser	professions	tenants	irradiated	depositor	pears
22	diamond	inpatient	homelessness	strontium	prudential	seed
23	paypercall	doctor	mortgages	irradiators	adviser	almonds
24	advertisement	medically	residents	nrc	securityholder	cotton
25	octane	nursing	homeownership	radiological	edgar	leaf
26	deception	clinical	mortgagees	fissile	underwriter	upland
27	fur	clinic	amenities	securityrelated	offering	usda
28	rvalue	hospitalbased	homeowner	strategic	interdealer	pork
29	optout	profession	homeless	thorium	dividends	cottonseed
30	franchisee	aides	poverty	nrcs	intermediarys	rot
31	furs	telemedicine	homebuyer	technetium	futures	goat
32	advertisers	medicare	rental	unrestricted	securitybased	ripe
33	franchisees	hhs	shelter	doenrc	depositary	flesh
34	abc	clinician	shelters	licenses	registrants	clover
35	ftc	fdas	mortgaged	license	dealers	seedless
36	telemarketing	hmos	developments	fsar	promoter	insects
37	unfair	medicaid	landlord	commissionapproved	nms	onions
38	wool	practitioners	mortgage	gamma	counterpartys	roots
39	biomassbased	patient	architect	snm	counterparties	tobacco
40	recyclable	stewardship	household	rulemakings	reliance	cherries
41	consumers	care	incomes	safeguards	fasb	raisins
42	imitation	biomedical	restructuring	physicist	dealer	stems
43	conveys	drugs	builder	licensee	repurchase	apple
44	telemarketer	hospice	modernization	licensees	penny	insect
45	textile	caregivers	occupancy	byproduct	diversified	olives
46	furnisher	shortage	tenant	engineered	soliciting	aggregating
47	franchisor	servings	buildingcomplex	enrichment	sx	kernel
48	freezer	disease	vacant	coc	crs	spready
49	rayon	interventions	occupy	repository	intercompany	lamb
50	conditioners	surgeons	vacancies	nb	ob	dirty

Table A.2: **Robustness—Enforcement vs. Regulatory Requirement: An Instrumental Variable Approach**

This table reports the robustness check of Table 11 by reconstructing the instrumental variable for enforcement shocks using only enforcement-related regulatory occupations in each agency. See Online Appendix A and Table 11 for details.

	Firm-Level			Establishment-Level		
	(1)	(2)	(3)	(4)	(5)	(6)
$iv(\Delta\log(p_{it}))$	0.223**		0.115	0.237***		0.125*
	(0.070)		(0.078)	(0.061)		(0.067)
$iv(\Delta\log(\tilde{R}_{it}))$		0.257***	0.224***		0.269***	0.231***
		(0.038)	(0.056)		(0.036)	(0.052)
Δ log(Wage)	-0.051***	-0.049***	-0.049***	-0.083***	-0.081***	-0.080***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)	(0.008)
Δ log(w^r)	0.289***	0.282***	0.280***	0.277***	0.270***	0.268***
	(0.010)	(0.012)	(0.012)	(0.011)	(0.013)	(0.013)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	608,500	608,500	608,500	628,733	628,733	628,733
Adjusted R^2	0.323	0.340	0.344	0.323	0.340	0.344

Table A.3: **Robustness—Enforcement vs. Regulatory Requirements in Subsamples by Size**

This table reports the robustness check of Table 12 by reconstructing the instrumental variable for enforcement shocks using only enforcement-related regulatory occupations in each agency. See Online Appendix A and Table 12 for details.

	Firm-Level				Establishment-Level			
	1-19	20-399	400-749	$\geq$ 750	1-19	20-399	400-749	$\geq$ 750
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$iv(\Delta\log(p_{it}))$	0.103	0.124	0.105	0.086	0.112	0.136*	0.169*	0.090
	(0.079)	(0.081)	(0.077)	(0.063)	(0.065)	(0.069)	(0.077)	(0.073)
$iv(\Delta\log(\tilde{R}_{it}))$	0.166**	0.253***	0.287***	0.309***	0.170**	0.271***	0.313***	0.365***
	(0.059)	(0.056)	(0.055)	(0.046)	(0.054)	(0.051)	(0.053)	(0.057)
Δ log(Wage)	-0.139***	0.002	0.047***	0.028**	-0.164***	-0.024**	0.029**	0.001
	(0.008)	(0.008)	(0.010)	(0.012)	(0.007)	(0.008)	(0.013)	(0.015)
Δ log(w^r)	0.353***	0.237***	0.182***	0.230***	0.356***	0.215***	0.107***	0.122***
	(0.011)	(0.013)	(0.017)	(0.020)	(0.011)	(0.013)	(0.021)	(0.026)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	189,404	352,779	29,482	36,835	220,464	375,485	19,622	13,162
Adjusted R^2	0.400	0.324	0.243	0.214	0.397	0.320	0.265	0.233

Table A.4: **Robustness—Enforcement vs. Regulatory Requirements for Firms in Each Sector**

This table reports the robustness check of Table 13 by reconstructing the instrumental variable for enforcement shocks using only enforcement-related regulatory occupations in each agency. See Online Appendix A and Table 13 for details.

	All Sizes			Subsample by Firm Size			
				1-19	20-399	400-749	$\geq$ 750
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Agriculture, Mining, and Construction							
$iv(\Delta\log(p_{it}))$	0.195**		0.096	0.093	0.095	0.056	0.206**
	(0.069)		(0.080)	(0.084)	(0.077)	(0.107)	(0.089)
$iv(\Delta\log(\tilde{R}_{it}))$		0.269***	0.241***	0.195**	0.273***	0.278**	0.198*
		(0.044)	(0.061)	(0.064)	(0.061)	(0.099)	(0.106)
Manufacturing							
$iv(\Delta\log(p_{it}))$	0.360***		0.195**	0.122*	0.200**	0.260**	0.251**
	(0.074)		(0.079)	(0.064)	(0.085)	(0.094)	(0.100)
$iv(\Delta\log(\tilde{R}_{it}))$		0.331***	0.281***	0.239***	0.303***	0.320***	0.412***
		(0.057)	(0.043)	(0.042)	(0.048)	(0.044)	(0.041)
Retail							
$iv(\Delta\log(p_{it}))$	0.127***		0.096**	0.094**	0.125**	0.063	0.028
	(0.038)		(0.035)	(0.041)	(0.046)	(0.051)	(0.040)
$iv(\Delta\log(\tilde{R}_{it}))$		0.134***	0.122***	0.122***	0.130***	0.102**	0.105***
		(0.028)	(0.025)	(0.020)	(0.032)	(0.033)	(0.022)
Wholesale							
$iv(\Delta\log(p_{it}))$	0.215***		0.125**	0.073*	0.165**	0.094	0.202*
	(0.052)		(0.040)	(0.034)	(0.058)	(0.097)	(0.109)
$iv(\Delta\log(\tilde{R}_{it}))$		0.194***	0.168***	0.125***	0.193***	0.288***	0.275***
		(0.038)	(0.032)	(0.035)	(0.040)	(0.073)	(0.070)
Utilities							
$iv(\Delta\log(p_{it}))$	0.538***		0.307**	0.204*	0.336*	-0.028	0.159
	(0.119)		(0.100)	(0.100)	(0.153)	(0.264)	(0.135)
$iv(\Delta\log(\tilde{R}_{it}))$		0.443***	0.335***	0.501***	0.211	0.529***	0.376***
		(0.130)	(0.090)	(0.081)	(0.167)	(0.054)	(0.045)
Transportation							
$iv(\Delta\log(p_{it}))$	0.192*		0.107	0.068	0.111	0.142**	0.198**
	(0.086)		(0.065)	(0.060)	(0.069)	(0.058)	(0.073)
$iv(\Delta\log(\tilde{R}_{it}))$		0.374***	0.354***	0.305***	0.362***	0.441***	0.503***
		(0.037)	(0.031)	(0.038)	(0.028)	(0.065)	(0.059)
Finance							
$iv(\Delta\log(p_{it}))$	0.371***		0.167**	0.164	0.169*	0.152	0.063
	(0.058)		(0.073)	(0.090)	(0.079)	(0.141)	(0.110)
$iv(\Delta\log(\tilde{R}_{it}))$		0.304***	0.248**	0.215**	0.284***	0.246***	0.347**
		(0.067)	(0.079)	(0.094)	(0.067)	(0.060)	(0.107)
Service							
$iv(\Delta\log(p_{it}))$	0.215**		0.107	0.106	0.110	0.069	0.044
	(0.070)		(0.085)	(0.084)	(0.085)	(0.068)	(0.065)
$iv(\Delta\log(\tilde{R}_{it}))$		0.251***	0.217***	0.164**	0.251***	0.265***	0.276***
		(0.040)	(0.060)	(0.061)	(0.060)	(0.047)	(0.047)

Table A.5: **Robustness—Enforcement vs. Regulatory Requirements for Establishments in Each Sector**

This table reports the robustness check of Table 14 by reconstructing the instrumental variable for enforcement shocks using only enforcement-related regulatory occupations in each agency. See Online Appendix A and Table 14 for details.

	All Sizes			Subsample by Firm Size			
				1-19	20-399	400-749	$\geq$ 750
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Agriculture, Mining, and Construction							
$iv(\Delta\log(p_{it}))$	0.194**		0.097	0.105	0.091	0.146	0.033
	(0.065)		(0.069)	(0.071)	(0.067)	(0.096)	(0.128)
$iv(\Delta\log(\tilde{R}_{it}))$		0.275***	0.247***	0.193***	0.286***	0.241*	0.161
		(0.041)	(0.054)	(0.056)	(0.052)	(0.119)	(0.115)
Manufacturing							
$iv(\Delta\log(p_{it}))$	0.389***		0.233**	0.154**	0.246***	0.311**	0.359***
	(0.072)		(0.073)	(0.066)	(0.071)	(0.100)	(0.108)
$iv(\Delta\log(\tilde{R}_{it}))$		0.342***	0.288***	0.239***	0.313***	0.330***	0.459***
		(0.054)	(0.027)	(0.027)	(0.029)	(0.037)	(0.041)
Retail							
$iv(\Delta\log(p_{it}))$	0.165*		0.149***	0.111**	0.185***	0.179*	0.023
	(0.073)		(0.045)	(0.046)	(0.048)	(0.080)	(0.138)
$iv(\Delta\log(\tilde{R}_{it}))$		0.189***	0.183***	0.132***	0.215***	0.296***	0.411***
		(0.049)	(0.044)	(0.030)	(0.057)	(0.076)	(0.102)
Wholesale							
$iv(\Delta\log(p_{it}))$	0.237***		0.156***	0.107**	0.192***	0.509***	0.435*
	(0.062)		(0.040)	(0.039)	(0.052)	(0.065)	(0.202)
$iv(\Delta\log(\tilde{R}_{it}))$		0.191***	0.161***	0.133***	0.189***	0.223**	0.093***
		(0.035)	(0.025)	(0.029)	(0.029)	(0.082)	(0.028)
Utilities							
$iv(\Delta\log(p_{it}))$	0.420***		0.219**	0.293*	0.101	0.269	0.067
	(0.090)		(0.096)	(0.133)	(0.109)	(0.259)	(0.312)
$iv(\Delta\log(\tilde{R}_{it}))$		0.373***	0.297***	0.291*	0.276***	0.009	0.298
		(0.081)	(0.056)	(0.139)	(0.052)	(0.222)	(0.232)
Transportation							
$iv(\Delta\log(p_{it}))$	0.231**		0.146**	0.100*	0.154*	0.202**	0.294
	(0.074)		(0.064)	(0.047)	(0.070)	(0.084)	(0.164)
$iv(\Delta\log(\tilde{R}_{it}))$		0.366***	0.341***	0.286***	0.363***	0.384***	0.508***
		(0.033)	(0.023)	(0.034)	(0.023)	(0.060)	(0.060)
Finance							
$iv(\Delta\log(p_{it}))$	0.352***		0.162**	0.163**	0.167**	0.101	-0.095
	(0.049)		(0.052)	(0.070)	(0.052)	(0.067)	(0.147)
$iv(\Delta\log(\tilde{R}_{it}))$		0.294***	0.234**	0.232**	0.252***	0.327***	0.290***
		(0.063)	(0.075)	(0.086)	(0.071)	(0.081)	(0.076)
Service							
$iv(\Delta\log(p_{it}))$	0.239***		0.124	0.116	0.133*	0.134**	0.075
	(0.058)		(0.074)	(0.072)	(0.072)	(0.054)	(0.066)
$iv(\Delta\log(\tilde{R}_{it}))$		0.260***	0.217***	0.167**	0.257***	0.263***	0.282***
		(0.040)	(0.058)	(0.057)	(0.056)	(0.065)	(0.075)

Figure A.1: **RegData Across States**

This figure plots the state-level RegData measure from QuantGov.org.

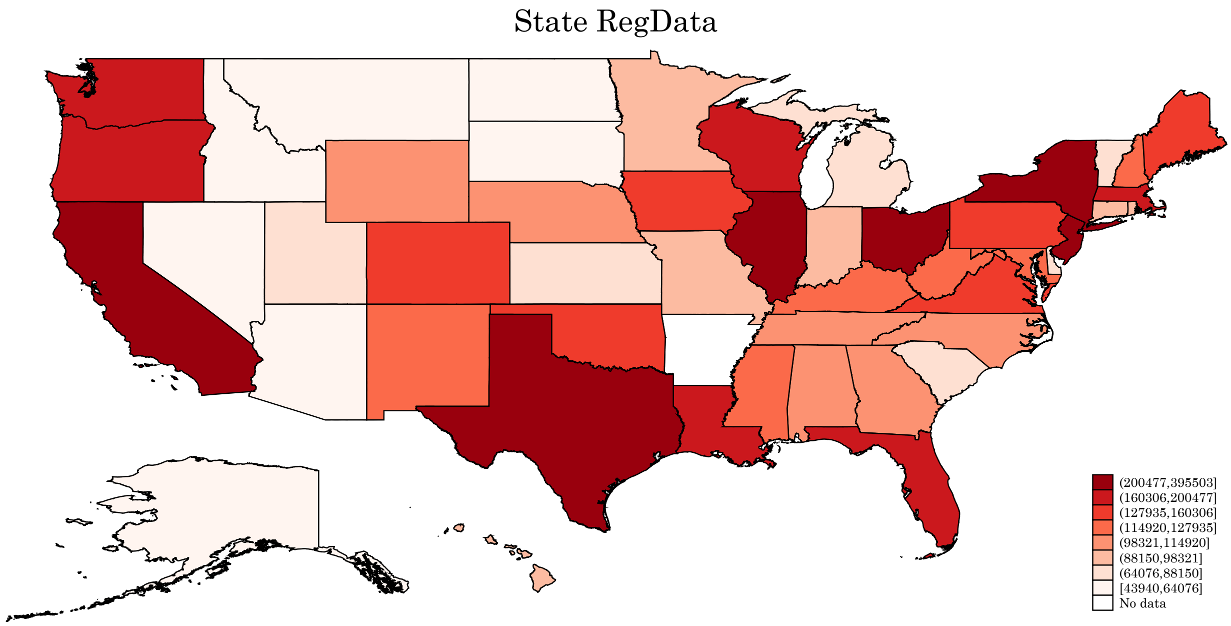


Figure A.2: **Case Studies of Industry Regulatory Shocks Using RegData**

This figure plots the response of industries' RegData measure to five industry-level regulatory shocks. RegData is from RegData Version 3.2. from QuantGov.com and is the natural logarithm of the count of restrictive words in the Code of Federal Regulations governing an industry in the year. Section 3 provides details of the industry shocks and discusses the classification of treated and control groups. To ease the comparison, we shift the lines vertically so that they have the same value in the year before the treatment. The value in the year before the treatment is the average of the regulation measures across the treated and control industries in that year. The difference between the two lines after the treatment, minus the difference between the two lines before the treatment reflects the difference-in-differences estimation.

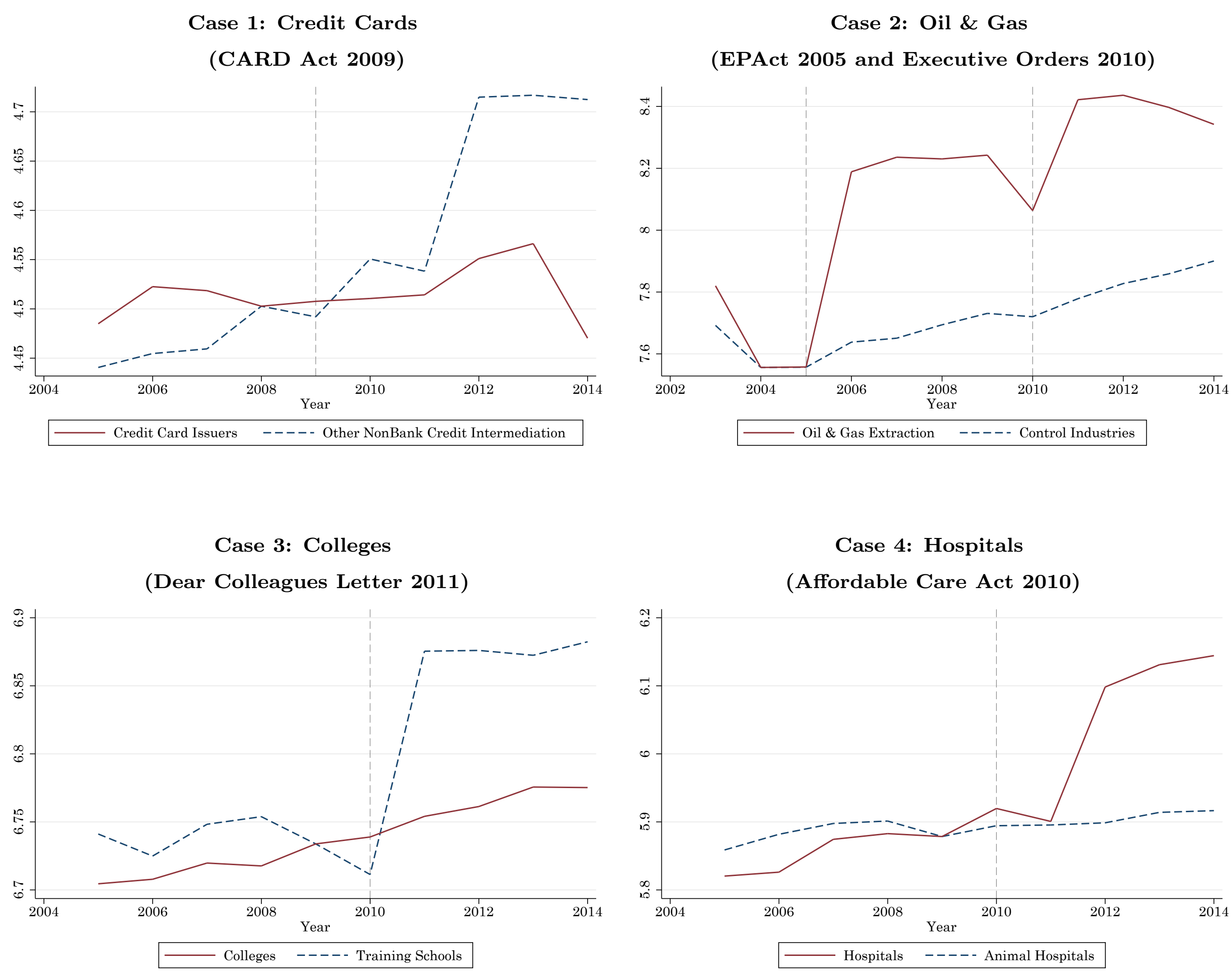


Figure A.3: **Levels of Enforcement and Regulatory Requirements Measures by Agency**

This figure plots each major agency's regulation-related employment, which is used to construct enforcement shocks, and the estimated compliance hours of the agency's regulations excluding adjustments, which are used to construct regulation-requirement shocks. See Section 5 for more details.