Follow the Pipeline: Anticipatory Effects of Proposed Regulations*

Suzanne Chang
Tulane University

Joseph Kalmenovitz
University of Rochester

Alejandro Lopez-Lira
University of Florida

June 19, 2024

Abstract

We develop a novel measure of regulatory pipeline: the amount of federal rule proposals which are relevant to the firm. The measure is based on a new data set that tracks the timeline of each of the 43,000 rule proposals developed by all federal agencies since 1995. The average proposal spends two years in the rulemaking pipeline and only two-thirds convert into a final rule. Training a machine-learning algorithm on the regulatory text and corporate disclosure files, we identify proposals which are likely relevant to each firm, yielding a firm-specific measure of exposure to the regulatory pipeline. We find that greater exposure is associated with increased overhead costs, reduced capital investments, and lower reported profits, independent of the current regulatory burden. The effect is stronger when the proposals address novel issues or aim to increase future compliance costs, suggesting that regulatory pipeline reshapes expectations and adds uncertainty about future regulatory burden. Finally, we calibrate a latent factor model of stock returns with our new firm-specific measure, and thus identify systematically important regulatory topics such as Securities, Natural Resources, and Environment. Combined, our results highlight substantial anticipatory effects among firms, based on the entire body of potential federal regulations.

*We received excellent comments from a former Administrator of the Office of Information and Regulatory Affairs; our discussants Nathan Dong, Karim Farroukh, Matthew Faulkner, Robert James, Andrew MacKinlay, Jennifer Nou, Daniel Rettl, Adriana Robertson, Sarath Sanga, Alexander Schandlbauer, Alison Taylor, Shuai Yuan, and Ben Zhang; from Yakov Amihud, John Cochrane, Gregor Matvos, and David Yermack; and from seminar and conference participants at the Conference on Empirical Legal Studies, Law & Macroeconomics Conference, NYU-Penn Law & Finance Conference, Finance Organizations and Markets Conference, Northern Finance Association, Southern Finance Association, Southern Economic Association, Financial Management Association, FMA Napa Finance Conference, AI and Big Data in Accounting and Finance Conference, Economics of Financial Technology Conference, Boca Raton Corporate Finance & Governance Conference, Sydney Banking & Financial Stability Conference, Australasian Finance & Banking Conference, Yale University, Rochester University, George Washington University, Tulane University, Aarhus University, Southern Denmark University, Office of the Comptroller of the Currency, and the Federal Reserve Board.
Please send correspondence to: jkalmeno@simon.rochester.edu.

Introduction

The economics of regulation is a rapidly growing field of study, exploring how regulation in all its forms affects economic activity. Papers in this field focus on *effective* regulations, which are included in the Code of Federal Regulations and dictate prices and behavior (e.g., Trebbi and Zhang (2022)). Missing from this literature is the role of *proposed* regulations: proposals to create new rules or amend existing ones. This omission is unfortunate, since anecdotal evidence suggests that firms strive to stay ahead of the curve and closely monitor the pipeline of rule proposals.[1] The regulatory pipeline can materially affect firm decisions for at least two reasons: it changes the expectations for future regulatory burden, and it increases the uncertainty surrounding future burden. For instance, firms may choose to postpone projects until uncertainty is resolved. To shed light on the topic, we assemble a new dataset and construct the first firm-level measure of regulatory pipeline. It captures the amount of regulations which are still under development by federal agencies and are relevant to the firm. Using the new measure, we document substantial anticipatory effects on firms, driven by changes in expectations as well as heightened uncertainty.[2] Notably, the effects are independent of the current regulatory burden, and independent of whether and when the proposed regulations are ultimately finalized.

To carry out the analysis, we source a granular data set from the Unified Agenda. The Agenda is a semi-annual official publication, summarizing the expected and pending rule-making activities of all federal agencies at that time point.[3] Crucially, the Agenda includes detailed timetables which retroactively describe the actions taken so far for each rule proposal. Based on that, we reconstruct the precise timeline of each proposal since the day it was

[1] For example, in its annual report for fiscal year 2021, Meta (Facebook's parent company) discusses how several pending legislative proposals can "cause us to incur significant compliance costs and could potentially impose new restrictions and requirements on companies like ours." To quote a prominent law firm, "[k]nowing which type of rule is under consideration [and] whether and when to comment on a proposed rule [...] is increasingly important in a shifting regulatory landscape" (link).

[2] The new measure is available on our website.

[3] We benefited greatly from a conversation with a former administrator of the Office of Information and Regulatory Affairs, the agency overseeing the data collection and the publication of the Agenda.

first introduced until its final resolution. Thus, we observe the federal government's pipeline at any point in time. Nearly 43,000 proposals have passed through the pipeline since 1995, and the daily average is 3,500 rule proposals. The average proposal spends 29 months in the pipeline. One out of three proposals does not survive, and is officially rescinded after 41 months. Two out of three proposals survive, clearing the pipeline after 21 months to become an effective rule. Given the large number of proposals in the pipeline, the uncertainty about the ultimate outcome, and the long time it takes to resolve the uncertainty, it is reasonable to expect potential regulations to be on the collective minds of managers.

In the first part of the paper, we map the aggregate rulemaking pipeline to the cross-section of firms. The intuition is that firms have differential exposure to rules. For example, financial firms will likely be more affected by rule proposals regarding trading and disclosure, while pharmaceutical firms will be more attuned to lab safety issues. To capture this heterogeneity, we first decompose the rulemaking pipeline into 100 regulatory topics. The decomposition is conducted by an unsupervised Latent Dirichlet Allocation (LDA) algorithm, an increasingly popular tool in the financial economics literature.[4] We apply the same algorithm to conference calls and identify how important is each LDA topic to each firm at a given point in time. To obtain our final firm-level measure, we interact each topic's importance (the perspective of the firm) with the fraction of the pipeline associated with this topic (the perspective of regulators), and sum across all topics. Intuitively, our measure captures the average number of proposals which are currently in the rulemaking pipeline and are relevant to the firm. It is a weighted average, where weights vary over time and across firms, reflecting how important each topic is for the firm at a given time.

In the second part of the paper, we investigate the properties of the regulatory pipeline. First, we verify that the LDA algorithm reasonably identifies distinct areas of regulatory activities. We do so by examining unique keywords associated with each topic and showing

[4]It has been used to study, for example, asset pricing puzzles (Bao and Datta (2014), Bybee et al. (2019), Lopez-Lira (2019), Bybee, Kelly, and Su (2022)), risks in the financial sector (Hanley and Hoberg (2019)), and issues discussed in FOMC speeches (Hansen, McMahon, and Prat (2018)).

that they vary intuitively across agencies and industries. For instance, Topic 45 is prevalent in the National Oceanic and Atmospheric Administration's rules (keywords: "ocean," "fish", "species"), while Topic 31 is closely aligned with the SEC (keywords: "investment", "disclosure", "company"). From the firm's perspective, Topics 26 is a dominant topic only for the Healthcare industry (keywords: "medicare", "medicaid", "hospital"), while Topic 27 is largely confined to the Banking industry (keywords: "loan", "mortgage", "soundness"). Thus, the LDA is a useful method to categorize the firm's operations and the federal government's rulemaking activities.[5]

More broadly, we establish three facts regarding the firm-level regulatory pipeline. *First*, we conduct a simple variance decomposition: regressing our measure on a growing number of fixed effects and noting the incremental increase in R^2. We find that economy-wide factors (time FE) account for 50% of the variation, industry factors explain an additional 10%, and the remaining 40% of the variation plays out at the firm level. *Second*, related industries share a commonality in regulatory pipeline (à la Chen and Kalmenovitz (2023)). For example, the regulatory pipeline of Business Services is positively correlated with Computers and Electronic Equipment, but negatively correlated with Construction Materials and Rubber and Plastic. *Third*, firms with greater exposure to the regulatory pipeline are subject to lower regulatory burden in the present (using data from Kalmenovitz (2023)), while expressing more concerns about future political risk (using data from Hassan et al. (2019)).

Combined, these exercises sharpen the interpretation of our measure. It captures a distinct notion of regulatory burden: potential future changes in the firm's regulatory environment, rather than the firm's current regulatory environment. There is a strong time-series component, which is expected given that our measure explicitly accounts for the government's pipeline, and a weaker cross-industry component, which reflects potential commonality in the regulatory environment across industries. Importantly, our measure varies substantially across firms, as opposed to time-series measures capturing various policies (such as Baker,

[5] We also develop a formal process to label each LDA topic by its economic content (Appendix A.2).

Bloom, and Davis (2016) and Al-Ubaydli and McLaughlin (2017)). Each firm has a unique exposure to the aggregate pipeline, consistent with recent empirical findings on firm-specific regulatory burden and political risk.

In the third part of the paper, we study the economic impact of the regulatory pipeline. We formulate two hypotheses. On the one hand, greater exposure to the pipeline could be beneficial. In the short term, firms might boost production before the rules are finalized. In the longer term, the expected rise in regulatory burden could create barriers to entry and help incumbent firms. On the other hand, greater exposure could be detrimental to firms. While the proposals are pending in the pipeline, the legal uncertainty could have a chilling effect on firms. Moreover, it becomes increasingly more difficult to capture regulators, given the broad range of rule proposals. Importantly, all those effects are independent of the current regulatory burden and are determined ex-ante, even if ex-post the proposed regulations failed to turn into final rules.

Our results are broadly consistent with the second hypothesis. We find that regulatory pipeline is associated with large overhead costs and smaller profit margins. We use SGA (sales, general, and administrative) and COGS (cost of goods sold) for the former and net income for the latter. The effects are identified within-firm over time (firm fixed effects), net of macroeconomic conditions and industry trends (year or year$\times$industry fixed effects). We control for known determinants of costs and profits: size, cash flow, leverage, Tobin's Q, the burden of effective regulations (from Kalmenovitz (2023)),[6] political risk (from Hassan et al. (2019)), age, and complexity. For the latter, we use the number of unique business segments and the concentration of LDA topics in the firm's conference call (high concentration indicates less complexity). With the large set of controls and fixed effects, our findings are more consistent with the second hypothesis: greater exposure to the regulatory pipeline is associated with a detrimental impact on firms, independent of other factors.[7]

[6]Note that regulatory pipeline is associated with *lower* regulatory burden, and thus an omitted variable would likely cause an attenuation bias.

[7]We compute a parallel measure of pipeline fragmentation, similar to Kalmenovitz, Lowry, and Volkova (2023). Instead of measuring the quantity of the pipeline ("how many rule proposals"), it measures how

Additional analysis highlights the two key mechanisms at play: anticipation and uncertainty. First, we find that the effects on costs and profits increase with the expected burden of the regulatory pipeline: when the proposals are considered "significant" by agencies,[8] or when they are more likely to convert into a final rule.[9] This is consistent with an anticipation mechanism: the pipeline shapes expectations of future regulatory changes, and firms are willing to take more actions today if the expected changes are more imminent and burdensome. Second, we find that the pipeline leads to significant decline in capital investment, especially when the pipeline includes proposals with unusual and unprecedented content.[10] This is consistent with an uncertainty mechanism, where managers postpone investments until regulatory uncertainty is resolved (McDonald and Siegel (1986); Bernanke (1983); Julio and Yook (2012); Gulen and Ion (2015)).

In the last part of the paper, we investigate how the regulatory pipeline relates to systematic risk. We introduce a factor model of returns, where the firm's β is a function of its exposure to regulatory topics. Hence, the regulatory pipeline could increase β and consequently the firm's sensitivity to latent common factors. Using the bootstrapping methodology in Kelly, Pruitt, and Su (2019) and our new firm-specific measure, we estimate the model and identify systematically important topics: those that significantly increase the firm's sensitivity to macroeconomic shocks. These topics are related to Pharmaceuticals, Healthcare, Natural Resources, Securities Markets, and Environmental Protection. Firms that are highly exposed to any of these topics will be more sensitive to macroeconomic shocks, hence increasing their exposure to systematic risk.

There is a growing recognition of the need to better understand the economic impact of regulation. Against the backdrop of this important challenge, the main innovation in our

fragmented the pipeline is ("how many agencies propose rules"). The fragmentation has a strong impact on corporate outcomes, but smaller in magnitude than the pipeline quantity.

[8] Rules which raise novel legal issues or have an annual economic effect of $100 million or more.

[9] We estimate a linear probability model in the cross-section of rule proposals, and interpret the predicted values as the ex-ante probability of success (Appendix A.4).

[10] For each focal rule i, we compute the Hellinger distance to all other peer rules $j \neq i$ issued by the same agency. Higher distance means that the rule ventures into topics which are uncharacteristic of the agency.

paper is to focus on *potential* regulations. With our new firm-specific measure of regulatory pipeline, we shed light on a relatively unexplored aspect of regulatory burden: potential future regulations that are relevant to the firm. Exposure to the regulatory pipeline differs conceptually from current studies, which focus on the burden of effective regulations, and it has little correlation with that burden. From a methodological standpoint, our measure has several advantages. First, it does not require firms to explicitly discuss regulation, thus mitigating concerns of strategic disclosure.[11] Second, we account for the relative importance of different regulatory topics, both within firms and within the government's rulemaking pipeline. This helps us measure the exposure more precisely. Third, the underlying data set (Unified Agenda) is publicly available and updated continuously. It can be extended for future studies and invites nuanced analysis at various frequencies or for a subset of regulations and agencies.

In addition, we contribute to the literature on the economic impact of regulation. This topic is the subject of many policy discussions and broad theories, ranging from *public interest* to *regulatory capture*.[12] Applying those theories to the context of regulatory pipeline, we document the impact of pending regulations on key firm-level outcomes, highlight the role of anticipation and uncertainty, and identify systematically important regulatory areas. The effects we uncover are independent of the current regulatory burden and materialize even if the proposed regulations ultimately failed to turn into final rules. Combined, these results shed more light on the real effects of regulation and the underlying mechanisms. Of course, we do not argue that the process of developing new regulations is inherently flawed, since the benefits of rule proposals must be ultimately weighed against any potential costs.

Lastly, our work relates to a large body of papers that study how specific rules affect companies. In this domain, several scholars have questioned the use of exact dates of regu-

[11]For example, Calomiris, Mamaysky, and Yang (2020) report that 70% of firms avoid discussing regulation altogether in their quarterly earning calls.

[12]The two overarching theories of regulation view regulation as a welfare-increasing institution (e.g., Pigou (1938); Demsetz (1974)) or as a rent-seeking process (e.g., Tullock (1967); Stigler (1971); Krueger (1974); Posner (1974); Peltzman (1976); Becker (1983)).

latory policies, given that market participants may change their behavior in anticipation of future rules (Bird et al. (2020); Jha, Karolyi, and Muller (2020); Bessembinder et al. (2018); Trebbi and Xiao (2019); Borochin et al. (2021); Hendricks et al. (2023); Khan and Khederlarian (2021)). Using a novel methodology and comprehensive data on all federal rules, we document substantial anticipatory effects: companies adjust their operations in expectation of future regulatory developments, independent of the current regulatory environment. This is important because it shows that focusing on the passage of a final rule misses a large part of the effect, as firms make substantial adjustments years before the regulation is finalized. The anticipatory effects are surprisingly large, given that a third of the proposed rules fail to convert into a final rule, and even those who do succeed spend nearly two years in the pipeline. Future studies can dig deeper into the origins and consequences of anticipatory effects, tactics adopted by companies to mitigate those effects, and potential heterogeneity across firms and rules.

1 Institutional setting and data

1.1 Institutional setting

Rulemaking is the process by which federal agencies develop, amend, and repeal regulations. In this paper we focus on the vast majority of regulations which are developed through the "informal" notice-and-comment process. The alternative, "formal" procedure must be specifically required by Congress and is rarely used (Yackee and Yackee (2010)).

The notice-and-comment process is outlined in Section 553 of the 1946 Administrative Procedures Act (APA), and additional requirements are in Executive Order 12866 and a few other statutes and decrees. Figure 1 provides a simple illustration of the process. It starts with a triggering event, such as an act of Congress. For example, the Wall Street Reform and Consumer Protection Act (Dodd-Frank Act) authorized the Securities and Exchange Commission to develop dozens of rules to enhance transparency and efficiency in financial markets. Once an agency decides that a regulatory action is necessary, it publishes a proposed rule in the Federal Register and solicits written comments by interested parties. The public has 30 to 60 days to comment on the proposed rule. The agency must consider the relevant comments when formulating the final rule. In some cases, the agency decides to withdraw the regulation. Otherwise, the agency makes changes where appropriate and publishes a final rule in the Federal Register with a specific date upon which the rule becomes effective. That rule is then integrated into the Code of Federal Regulations.

Some exceptions apply. For example, some rules are initiated by the agency as part of its broad authority, not by a specific act of Congress. A final rule might be issued without first issuing a notice of proposed rulemaking, and some rules may be published for public comment more than once.

1.2 Data sources

The primary source for this paper is the Unified Agenda (UA), widely considered to be the most comprehensive source of federal government rulemaking.[13] It is the official, semiannual publication of all expected and pending rulemaking activities of all federal agencies at that time point.[14] Twice a year, each agency prepares a detailed report on rules under development. The entire process is overseen by the Office of Information and Regulatory Affairs (OIRA), a unit within the Office of Management and Budget. The reports cover all regulations in various stages of development, some merely an early draft and some nearing completion. The agency provides a timetable of all the actions it has taken so far with regards to each regulation: publishing a draft, convening a public meeting, or a full or partial retraction of the proposed regulation. OIRA processes the reports from the entire federal government and produces a single publication, the Unified Agenda. The Agenda is thus a comprehensive snapshot of all federal regulations that are still under development or have

[13] The information in this section is based on a conversation with a former administrator of the Office of Information and Regulatory Affairs; see also Administrative Conference of the United States (2015).

[14] Agencies use the UA to satisfy their legal requirements under Section 602 of the Regulatory Flexibility Act and Section 4(b) of Executive Order 12866.

been resolved since the previous edition.

In the Agenda, rules are uniquely identified by the Regulatory Identification Number (RIN). The RIN remains constant throughout the entire rulemaking process. For ease of notation, we will use the terms "RIN," "rule," and regulation" interchangeably to describe a single regulatory action.[15] Finally, note that we download and parse all the Agenda editions since Fall 1995 until Spring 2023.[16] However, our ultimate firm-level measure of regulatory pipeline relies on conference calls, and those transcripts are widely available only since 2008. Therefore, our empirical exercises start in 2008.

For our empirical analysis, it is important to establish the timeline of each rule (RIN). This helps us determine if the rule is actively under consideration and therefore potentially relevant the the firm. In a nutshell, a rule enters the pipeline when it is first formally announced by the agency, and exits the pipeline when a final version has been published, or if the proposed regulation has been officially retracted. Note that the Agenda is published only twice a year, but we use the detailed timetables in the Agenda, which retroactively describe the actions taken so far for each RIN. We provide a full description of the procedure in Appendix A.1.

Table 1 summarizes the sample of 42,934 regulations under development.[17] The latest edition of the Unified Agenda was published in Spring 2023, and 7.5% of the rules were still actively under consideration (total of 3,199). Out of the remaining 39,735 rules, which were resolved prior to Spring 2023, nearly one third (30%) were officially rescinded without any rule and nearly two thirds (67%) concluded with at least one published rule. The remaining 3% ended with mixed results: part of the original proposal was withdrawn and

[15] Occasionally a single regulatory action would have multiple RINs, for example if multiple agencies collaborate on a new regulation. Also, a single RIN could yield more than one regulation or rule, i.e. it could add or modify more than a single section in the Code of Federal Regulations.

[16] Federal agencies have been required to compile regulatory agendas since 1978, and comparable data is available beginning in 1983, but in electronic format only since 1995.

[17] Table A.1 lists the top 20 agencies by rule quantity. For instance, the Securities and Exchange Commission had 743 rule proposals in the pipeline during the sample period. The top four agencies are the National Oceanic and Atmospheric Administration, the Internal Revenue Service, the Fish & Wildlife Service, and the Office of Air & Radiation at the EPA. Combined, they are responsible for 21% of the government's rulemaking.

part of it was finalized and codified into the CFR. One quarter (26%) of the rules were put on hold at least once during their lifetime, meaning that the agency decided to suspend the rulemaking process for at least 12 months. The table also shows that the average rule proposal spent 869 days in the pipeline (29 months). For the subset of rule proposals that culminated in a final rule, the average proposal spent 640 days in the pipeline (22 months). In Figure 1, Panel B, we plot the distribution of anticipatory periods across proposals. For each proposal that culminated in a final rule, we compute the number of days that the proposal was in the rulemaking pipeline: from the moment it was first announced and until the final rule was published. The mean and median values are 640 and 363 (22 and 12 months), respectively. In other words, the majority of federal regulations were officially announced at least a year before being finalized. This significantly long period opens the door for substantial anticipatory effects, and firms can already start adjusting their policies in anticipation of rules clearing the pipeline and becoming final rules.

In addition to the Unified Agenda, we use conference call transcripts from the Capital IQ Transcripts database and financial performance data from Compustat. We also source a set of control variables from the emerging literature on the economics of regulation, which we describe below in Section 3.3.

2 Measuring regulatory pipeline

Our goal is to measure firm-specific exposure to the regulatory pipeline. Intuitively, we seek to capture how many potential regulations apply to the company. Formally, it is defined as:

$$RegPipeline_{i,t} = \sum_{o=1}^{O} \omega_{i,o,t} \cdot TopicPipeline_{o,t}, \tag{1}$$

where $TopicPipeline_{o,t}$ represents the fraction of the aggregate pipeline at time t which belongs to topic o. This is a time-series measure which varies across regulatory topics and over time, but does not vary across firms. The variation across firms is driven only by $\omega_{i,o,t}$, which captures the importance of topic o in the eyes of firm i at time t.

2.1 Regulatory topics

Our first step is to identify regulatory topics. For each RIN, we first combine all the textual descriptions across all years and editions of the UA, using the rule's title and abstract. We then follow standard practice in this literature and remove common verbs and stop words such as "also," "including," "will" and "must."[18] Next, we use Latent Dirichlet Allocation (LDA) to identify regulatory topics. LDA is a Bayesian unsupervised algorithm increasingly used by financial economists.[19] Intuitively, LDA looks at co-occurrence of words within each document (RIN), and groups together words that often appear together in the same document.[20] It then describes each rule as a distribution over topics. Consequently, LDA provides us a summary of the topics that the Unified Agenda describes, and what percentage of each rule discusses each topic.

By relying on the machine to define topics, we mitigate biases that might arrive from manually defining topics, for example through word lists or human coding (Nanda et al. (2023)). The only input is to define the number of topics. By way of comparison, Kalmenovitz, Lowry, and Volkova (2023) identify 100 topics in the Federal Register, the government's official daily publication. Lowry, Michaely, and Volkova (2020) identify 30 topics in firms' prospectuses and 8 topics in the SEC's comment letters pertaining to these prospectuses. We therefore chose 100 topics as our baseline. In a set of robustness exercises, we consolidate these 100 topics into a smaller set and conversely expand the number of topics beyond 100 ones. The results remain qualitatively similar.[21]

[18] The full list is available on request.

[19] Topics of interest include asset pricing puzzles (Israelsen (2014), Bao and Datta (2014), Bybee et al. (2019), Lopez-Lira (2019), Bybee, Kelly, and Su (2022)), emerging risks in the financial sector (Hanley and Hoberg (2019)), issues discussed in FOMC speeches (Hansen, McMahon, and Prat (2018)), and government regulations (Kalmenovitz, Lowry, and Volkova (2023)).

[20] Note that LDA does not consider semantic relationships between words or the larger contextual meaning of phrases.

[21] Increasing the number of topics could increase measurement error by picking up idiosyncratic changes in language that are not important. In general, we believe that any error stemming from the text processing

2.2 Aggregate regulatory pipeline

The next step is to decompose the aggregate pipeline into distinct regulatory topics. We retain the set of rules which are in the pipeline as of time t, denoted by R_t. We then define the pipeline share of topic o at time t as:

$$TopicPipeline_{o,t} = \sum_{r=1}^{R_t} Weight_{o,r}, \tag{2}$$

where $Weight_{o,r}$ is the fraction of rule r dedicated to topic o, as identified by the LDA algorithm. Higher values imply that potential rules touch more frequently on topic o. Note that $Weight_{o,r}$ does not vary over time (the rule's topic distribution remains constant).[22] Instead, the variation over time comes from R_t, that is, the set of rules that are currently in the government's pipeline. Another thing to note is that the sum of weights within rule ($\sum^{O} Weight_{o,r}$) must equal 1, but the sum of weights within topic ($\sum^{R_t} Weight_{o,r}$) can be greater than 1. Therefore, the topic's share of the pipeline ($TopicPipeline_{o,t}$) can exceed 1.

2.3 Firm-level regulatory pipeline

Our next step is to identify how important is each LDA topic to the firm. To that end, we use quarterly conference calls from the Capital IQ Transcripts database.[23] We keep all the text in each transcript except the parts corresponding to the operator speaking. We remove standard stop words, as described in Section 2.1. We project each conference call into the LDA model, trained on the Unified Agenda. Consequently, a conference call is represented as a 100-dimension vector, where each component corresponds to the fraction of the text allocated to each of the 100 topics discussed in the regulations. Each component can be

or algorithmic fitting is uncorrelated with firm-level economic factors.

[22] While the rule's content may be updated, its core topic distribution should remain largely invariant. If it did change dramatically, it would probably be replaced by a new RIN.

[23] As in Correia et al. (2012), Davis et al. (2015), Rennekamp, Sethuraman, and Steenhoven (2022), Cao et al. (2023), Brown, Hillegeist, and Lo (2004), Mayew et al. (2012), and Jiang et al. (2019).

defined as:

$$\omega_{i,o,t} = \frac{Words_{o,i,t}}{Words_{i,t}} = \frac{Words_{o,i,t}}{\sum_{o=1}^{O} Words_{o,i,t}}, \tag{3}$$

where $Words_{i,t}$ is the number of words in the conference call of firm i at time t, and $Words_{o,i,t}$ is words devoted specifically to topic o. Thus, $\omega_{i,o,t}$ reflects the relative importance of topic o for firm i at time t. By construction, $\omega_{i,o,t}$ ranges from 0 to 1 and its average is 1% (since there are 100 topics). Note that $\omega_{i,o,t}$ is measured at a quarterly frequency. Some of our specifications are at an annual frequency, and in those instances we use the annual average of $\omega_{i,o,t}$ across quarters. Lastly, we combine the topic's weight ($\omega_{i,o,t}$) with the topic's frequency ($TopicPipeline_{o,t}$) to generate our final firm×year measure, $RegPipeline_{i,t}$ (Equation (1)).

3 Understanding regulatory pipeline

In Section 2, we described our methodology to measure regulatory pipeline. In this section, we focus on the interpretation and validation of our new measure. We begin by studying the output of the LDA topic classifier, and then examine various aspect of the final product: firm-level exposure to the regulatory pipeline.

3.1 Patterns in regulatory topics

The LDA algorithm is unsupervised, except for the number of topics, which must be selected by the econometrician. Thus, it is important to examine whether the output reasonably captures regulatory topics.

We start by examining the economic content of each topic. As a first pass, we look at the word clouds, where a larger font corresponds to a larger weight of the word in that topic. The examples in Figure 2 reveal how each topic is centered on a distinct set of keywords, which correspond to a potential area of regulation. For example, Topic 1 concerns regulations about food labeling with the keywords "label," "dietary," and "content." More generally, we combine the word clouds with additional information, and propose a brief label for each

LDA topic that summarizes its content. This step is not necessary for the empirical analysis in the next sections, but it sheds light on the regulatory activities as they are described by the LDA algorithm. For brevity, the details of the methodology are in Appendix A.2 and a list of selected LDA topics with their labels and keywords is in Table A.2. For consistency, we will refer to the LDA topics in the main text by their ordinal number (which is arbitrarily determined and has no meaningful content in itself), rather than by their labels.

Next, we study the relative prominence of LDA topics within the government's rule-making pipeline. Each quarter we rank the topics from 1 to 100, based on their weights in the pipeline, where the topic with the highest weight is ranked 1. We then count, for each topic and ranking, the number of quarters in which the topic has achieved the given ranking. Figure 3, Panel A, displays the total counts for all topic-ranking pairs. By comparing the shading within column, we can see whether the topic's importance fluctuates over time (entire column is lightly shaded), or whether the topic tends to have a fixed degree of importance (only a handful of ranks are darkly shaded). Most topics are characterized by light shadings, indicating that their dominance in the Unified Agenda varies greatly over time. However, a handful of topics are consistently either one of the most or least dominant in the Agenda. For example, the dark shaded rectangle at the bottom of the figure conveys the fact that Topic 35 had the top ranking in 63 of the 108 quarters. Looking at the most important keywords for Topic 35 (which we label "Government operations: legal"), we can see the reason for the topic's dominance: many of the keywords apply to rules from a broad set of agencies. On the other hand, the dark rectangle near the top and middle of the figure shows that Topic 45 (which we label "Fisheries and fishing: treaties") is one of the least dominant topics in most periods, with a rank of 98 for 73 of the 108 quarters. Again, we can make sense of this result by looking at the keywords for Topic 45. Keywords such as "ocean," "fish", and "species" strongly suggests that this topic involves wildlife protection issues. Not surprisingly, this topic is frequently mentioned in rules promulgated by the National Oceanic and Atmospheric Administration (NOAA).

We can confirm the narrow applicability of Topic 45 across agencies in Figure 3, Panel B. We take the top 20 agencies in terms of the total number of rules in the publication history of the Unified Agenda. We then calculate, for each agency, the average weight of each topic across its history of rules and report the top and bottom five topics for each agency. We see that Topic 45 is one of the least dominant topics for more than half of the agencies, but one of the most dominant topics for NOAA. In contrast, Topic 35 is a top-five topic for several of the top-20 agencies. This demonstrates the uniqueness of Topic 45 versus the universality of Topic 35. Finally, we can also see that Topic 31 (which we label "Securities: investment companies") is the highest-ranked topic for the SEC, but is not in the top-five for any other of the top agencies. Looking at the keywords associated with Topic 31 (Figure 2), we can clearly see how many of those words align closely with the SEC's mission (for example, "investment", "disclosure", and "company").

We conduct another cross-sectional validation of our topic classifications by analyzing the importance of topics across industries. We start with the topic weights for our firms each quarter and then calculate the average weight of each of the 100 topics for firms in the same Fama-French 48 industry. The resulting highest-ranked and lowest-ranked topics for each industry are illustrated in Figure 3, Panel C. Two topics, numbered 40 and 83, consistently rank as the highest and lowest, respectively, across industries. Again, we can validate this pattern by looking at the keywords for each topic. Topic 40, which we label "Banking: credit and income taxes," has keywords such as "rate", "percent", and "annual", which should apply to a wide set of industries. In contrast, Topic 83 seems to be associated with immigration issues ("immigration", "alien", and "visa"), and we indeed label it "Immigration." This topic should apply only to firms that deal intensively with immigration policies. We can also see that Topics 26 and 27 are one of the most important topics only for the Healthcare and Banking industries, respectively. The keywords for the two topics are predominantly associated specifically with each of these two industries: "medicare", "medicaid", and "hospital" for Topic 26 (which we label "Health: medicare"), and "loan", "mortgage", and "soundness" for topic 27 (which we label "Rural credit").

Finally, we identify conference calls with abnormally high weight on a single topic (the average maximum weight is 0.1). We then read the transcript to understand why the call relates to the topic. The analysis is summarized in Table A.3. One example is the call of GSI Technology on July 31, 2014, with 41% weight for Topic 40, which we label as "Banking: credit and income taxes." Indeed, the content of this conference call mostly pertains to the firm's financials.[24] Another example is the call of Minim (formerly Zoom) on February 24, 2011, with 29% weight for Topic 34, which we label as "Telecommunication and aviation." Indeed, the firm used the call to discuss its products in the space of internet software, hardware, and security. Our final example is the call of Warren Resources, an oil and natural gas company, on August 7, 2013. The highest weight for this call is on Topic 74, which we label as "Government contracts: natural resources." Much of the call contains an extensive overview of the current state of the firm's business, including drilling activity and exploration.

In sum, the LDA algorithm appears to reasonably identify distinct areas of regulatory activities. Close examination of specific topics reveals typical keywords associated with those areas, such as banking, fish and wildlife, and healthcare. The topics vary intuitively over time and across agencies and industries, suggesting that the LDA is a useful method to parse the federal government's rulemaking activities.

3.2 Patterns in regulatory pipeline

The previous section examined the building block of LDA topics. We now turn to the final output: *RegPipeline*, a firm-level measure of exposure to the regulatory pipeline. We begin with a simple variance decomposition, regressing *RegPipeline* on a growing number of fixed effects and reporting the incremental increase in R^2. Our results are in Table 2, Panel A. Economy-wide factors (time FE) account for 49.5% of the variation in regulatory pipeline.

[24] Topic 40 carries the greatest weight in about 51% of the calls, which is consistent with the fact that firms typically discuss their financials during their regular calls.

Using the Fama-French 48 industry classification, time-invariant industry factors (industry FE) account for an additional 9.3%, while time-varying factors (industry×time FE) add 1.2%. Thus, 40% of the variation in *RegPipeline* plays out at the firm level, with 11% time invariant (firm FE) and 29% varying over time.[25] For comparison, Kalmenovitz (2023) finds that one-third of the variation in compliance costs is at the firm level, and Hassan et al. (2019) find that nearly 90% of the variation in political risk is at the firm level. Recall that *RegPipeline* explicitly includes a time-series component (Equation (1)), and it is therefore not surprising that the time trend explains a larger fraction of the variation relative to those other measures.

Motivated by these findings, we study the properties of the aggregate regulatory pipeline. We decompose each rule proposal into 100 topics, and then sum the rule fractions within topic×quarter. The resultant panel includes 100 topics over 112 quarters, from 1995 to 2022, totaling 11,200 observations. We describe this sample in Table 2, Panel B. The average topic×quarter includes 36.6 rules. The average topic includes 3.6 new rule proposals, which have been announced for the first time during this quarter, and 17.2 "empty" rule proposals, for which the agency has not published any written draft so far.

Next, we study the commonality in regulatory pipeline across industries (à la Chen and Kalmenovitz (2023)). We first residualize out the time variation of *RegPipeline* and then, for each quarter, calculate the average regulatory pipeline for firms in the same Fama-French 48 industry. Finally, we calculate the correlation of *RegPipeline* between each pair of industries and illustrate the correlations in Figure 4. The figure is a 48×48 square, where the shading of each square indicates the correlation of an industry pair. The industries have been ordered by their average correlation with the other industries, such that the first column on the left is the industry with the highest average correlation with the other industries, in this case Fama-French industry 34 (Business Services). There appears to be a wide variation in regulatory pipeline across industries. In addition, the correlation between industries' regula-

[25] From a different perspective, regressing *RegPipeline* on its lag values we find autocorrelation of 0.59.

tory pipeline generally reflects how closely related the industries are. As an example, the top left square is the correlation of one for the regulatory pipeline of Fama-French industry 34 ($RegPipeline^{34}$) with itself. Moving from the top, the three squares directly below show the correlation between $RegPipeline^{34}$ and $RegPipeline^{36}$ (Computers), $RegPipeline^{35}$ (Communication), and $RegPipeline^{32}$ (Electronic Equipment), with correlations of 0.62, 0.78, and 0.69, respectively. Starting from the bottom left square and moving directly upward, the three squares show the correlation between $RegPipeline^{34}$ and $RegPipeline^{38}$ (Business Supplies), $RegPipeline^{15}$ (Construction Materials), and $RegPipeline^{17}$ (Rubber and Plastic Products), with correlations of -0.56, -0.47, and -0.63, respectively.

Finally, Panel C in Table 2 reports summary statistics. Each quarter we split firms into five groups based on their exposure to the rulemaking pipeline, such that the level of exposure increases across quintiles. We calculate the average characteristic within each group and the difference between the top and bottom quintiles. Companies with higher exposure to the regulatory pipeline tend to be larger but younger. They have better investment opportunities (Tobin's Q) and yet spend less on capital investment as a fraction of their size. They have higher overhead costs, lower profits, and worse cash flows, all relative to their size. Interestingly, firms with high exposure to the regulatory pipeline tend to be subject to lower regulatory burden but more political risk. We elaborate on this point in the next section. Taken together, the initial analysis suggests that exposure to the regulatory pipeline is associated with significant changes in firm operations. We will explore those in greater detail in Section 4.

3.3 Stock versus flow of regulation

Our measure of regulatory pipeline captures the flow of potential regulations. This flow could be related to the stock of effective regulations, those which are currently in place. To investigate this possibility, we correlate our measure with a host of measures from the literature that intend to capture the burden of effective regulations.[26] Our findings are summarized in Table 3.

We start with the costs of compliance with federal paperwork regulations, as introduced by Kalmenovitz (2023). Based on an administrative dataset and machine-learning classification model, the author develops four measures of regulatory burden. They represent the number of paperwork regulations that apply to the company ($RegIn$) and the estimated compliance costs in terms of hours ($RegIn^{time}$), forms ($RegIn^{forms}$), and dollars ($RegIn^{dollars}$). Additionally, we use data from Kalmenovitz, Lowry, and Volkova (2023), who measure the exposure of firms to the broad scope of government activities as reported in the Federal Register. We label this variable $FedRegister$. Finally, we compute the exposure of firms to sections in the Code of Federal Regulations. We label this variable CFR, and the methodology is described in Appendix A.3. Overall, we find that regulatory pipeline is positively associated with the quantity of regulations: the number of paperwork regulations ($RegIn$), government activity ($FedRegister$), and restrictions in the Code of Federal Regulations (CFR).[27] However, regulatory pipeline is negatively associated with the economic burden of regulations, represented by more refined measures of regulatory burden that take into account the costs of compliance ($RegIn^{time}$, $RegIn^{forms}$, and $RegIn^{dollars}$).

We also consider the measure of political risk from Hassan et al. (2019). The authors rely on transcripts of conference calls, and identify keywords that are associated with political risk. Scaling the number of those keywords by the total length of the transcript, they obtain a firm-specific measure of political risk. Interestingly, regulatory pipeline is positively associated with political risk. Note that both measures use conference calls, and conceptually, political risk should have some overlap with the risk of regulations being added or modified in the future. The correlation is again statistically significant but economically small (0.03). Finally, we follow Calomiris, Mamaysky, and Yang (2020) and count the number of words

[26] A related measure, the Economic Policy Uncertainty (EPU) from Baker, Bloom, and Davis (2016), is a time-series property which does not vary across firms, unlike our measure.

[27] We obtain similar correlations between current levels of burden and lagged values of $RegPipeline$.

associated with regulation in the 10-K.[28] We find that the count of regulatory terms in the firm's 10-K is negatively associated with our measure of regulatory pipeline. In other words, firms with high exposure to the pipeline are reluctant to explicitly discuss regulation in their disclosure filings.

In sum, firms that face higher regulatory pipeline are in general subject to lower regulatory burden in the present while being concerned about future political risk. This could happen, for instance, if the current pipeline acts as a mean-reversal mechanism: regulators plan more activity in the future for topics with currently low burden, and less activity for "saturated" topics (those with high burden). More broadly, those facts help sharpen the interpretation of our measure. It captures a distinct notion of regulatory burden, one that is focused on potential future changes in the regulatory environment, rather than the current regulatory environment the company is facing. We will rely on these insights to guide the empirical strategy in Section 4.

4 Economic consequences of regulatory pipeline

4.1 Main results: costs and profits

How does exposure to the regulatory pipeline affect firms? We develop two broad hypotheses. On the one hand, greater exposure could be beneficial for firms. In the short term, firms might boost production before the rules are finalized. Moreover, the expected rise in regulatory burden could create barriers to entry, which would improve the position of incumbent firms. Regulation could also generate positive externalities which would improve firm performance.[29] On the other hand, greater exposure could be detrimental to firms. While the proposals are pending in the pipeline, the legal uncertainty could have a chilling effect

[28] The words are: regulation, regulatory, legislation, legislative, compliance, restriction, restrictive, supervision, and supervisory.

[29] This is rooted in *public interest* theories, which view regulation as a welfare-increasing institution designed to correct market failures (Pigou (1938); Demsetz (1974); Melody (2016)).

on firms. Moreover, it is more difficult to capture regulators given the broad range of rule proposals.[30] If the proposals become effective rules and increase the burden, then firms will incur more costs and perhaps lose investment opportunities.

To test the alternative hypotheses, we estimate the following specification:

$$y_{i,j,t+l} = \alpha + \beta \cdot RegPipeline_{i,t} + \gamma \overrightarrow{X}_{i,t} + \lambda_i + \lambda_{j,t} + \epsilon_{i,j,t}, \tag{4}$$

where $y_{i,j,t}$ is the outcome for firm i at time t, and j denotes the company's Fama-French 48 industry. The two primary outcomes are SGA (sales, general, and administrative) and COGS (cost of goods sold). Those two item lines are likely to capture costs associated with regulatory compliance (Kalmenovitz, Lowry, and Volkova (2023)). Both COGS and SGA are scaled by beginning-of-period total assets. The main variable of interest represents the quantity of the regulatory pipeline ($RegPipeline_{i,t}$). All independent variables are lagged one period to limit concerns of reverse causality, and divided by their cross-sectional standard deviation to facilitate the comparison of their economic magnitudes. The baseline regressions are at a quarterly frequency and limited to the years 2008-2021 to match the availability of regulatory pipeline variables. We winsorize all variables at the 2.5% and 97.5% level to reduce the impact of extreme outliers and cluster standard errors at the firm level.

Regulation is not randomly assigned across companies. In particular, an adverse productivity shock could drive up both costs and planned regulatory changes. Struggling companies may face surging costs, and at the same time initiate more contact with regulators (for example, apply for financial support) or attract more attention from regulators (for example, if they wish to protect the industry).[31] In the baseline analysis, we mitigate the concerns with a rich set of controls and fixed effects; we discuss additional concrete concerns below

[30] This is related to public choice and regulatory capture theories, which view regulation as a rent-seeking process (Tullock (1967); Stigler (1971); Krueger (1974); Posner (1974); Peltzman (1976); Becker (1983); Benmelech and Moskowitz (2010)).

[31] A related concern is strategic disclosure: firms anticipate future low profits and/or higher costs, and use the conference calls to lay the blame on regulators. However, note that our LDA-based methodology does not require firms to explicitly discuss any regulatory developments, which should assuage such concerns.

in Section 4.6. We control for a bevy of factors which are likely correlated with both regulatory burden and with company policies: size, Tobin's q, operating cash flows, leverage, and market-to-book ratio (see definitions in Appendix A.3). We add firm fixed effects to exploit variation within firm over time, and either time or time×industry fixed effects. The former removes the time trend, an important factor in regulation (see Table 2, Panel A). The latter further removes industry trends, ensuring that our results are not driven by the overall regulatory environment in the industry at a given time.

Another concern is that regulatory pipeline is correlated with the burden of effective regulations. As shown in Section 3.3, our measure is largely associated with lower regulatory burden. Thus, omitted variables would likely cause an attenuation bias. To further address this concern, we control for compliance costs using the $RegIn$ measure from Kalmenovitz (2023).[32] Additionally, we control for political risk using the $PRisk$ variable from Hassan et al. (2019). Finally, it is possible that firms with high exposure to the pipeline are more complex, operating across multiple markets and thus facing more potential regulations. To account for that, we use the distribution of topics in the conference call to define a variable which we label $Complexity$. Intuitively, if the firm discusses multiple topics, it operates in many markets and is thus more complex. Conversely, if the firm discusses predominantly a single topic, it operates in fewer areas and is thus less complex.

Table 4 summarizes the results. Across all specifications, regulatory pipeline is associated with material increases in costs. The effect is identified within firm over time, conditional on other determinants of costs, and is identified in the cross-section of industries (time FE) and net of industry trends (time×industry FE).[33] Note that, when adding explicit controls for the level of regulations, the coefficient increases in absolute value. This is consistent with Section 3.3, where we show that our measure is largely associated with lower regulatory

[32] In the main text we use the number of active regulations, but our conclusions remain similar when we control instead for the estimated compliance costs in terms of hours, forms, and dollars.

[33] The coefficient remains largely intact with the addition of year×industry fixed effects. This might be due to the weak cross-industry component of our measure, which we discussed above in Section 3.2. Alternatively, it could reflect the cross-industry nature of regulation (Chen and Kalmenovitz (2023)).

burden. Indeed, failing to control for the current burden causes an attenuation bias. In all specifications, the effect is statistically significant and economically meaningful. Since all independent variables are divided by their standard deviation, we can directly compare their magnitudes. Looking at the tightest specification with time×industry fixed effects, a one-standard-deviation increase in *RegPipeline* increases COGS and SGA by 0.31% and 0.10%, respectively (as a percentage of total assets). In comparison, a one-standard-deviation increase in assets leads to 2.2% and 8.1% declines in COGS and SGA, respectively, due to economies of scale. Thus, the regulatory pipeline effect is 15-17% of the size effect, in absolute value. The burden of effective regulations (*RegIn*) is also positive and significant, and is 20-30% larger than the anticipatory effects (*RegPipeline*). Interestingly, political risk is associated with lower overhead costs and turns insignificant in some specifications.[34]

Since costs are rising in the face of regulatory pipeline, a natural follow-on question is how it affects firm profits. On one hand, holding all else equal, higher costs should lead to lower profits. However, this adverse effect could be moderated or even reversed for several reasons. Generally, future regulations could generate positive externalities, especially according to *public interest* theories. Moreover, the anticipated regulatory burden can raise barriers to entry and let incumbents enjoy higher profit margins. Finally, firms may choose to boost production before the rule proposals are finalized, leading to at least a short-term rise in profits. To examine the net effect, we compute net income over lagged total assets and re-estimate our workhorse specification (Equation (4)). The results are in the last three columns of Table 4, showing that regulatory pipeline has a significant negative impact on net income. In the tightest specification with time×industry fixed effects, a one-standard-deviation increase in *RegPipeline* reduces net income by 0.1%. This is equal to the effect of current regulatory burden and more than double the effect of political risk.

In sum, regulatory pipeline leads to a significant rise in spending with a knock-on effect on profits. This is broadly consistent with our second hypothesis, which predicts that greater

[34]When estimating our baseline specification without *RegPipeline*, the coefficient on political risk is positive and significant.

exposure has detrimental effects on firms.

4.2 Anticipation

According to our interpretation, the results in Section 4.1 are due to anticipation of future regulatory actions. While we cannot directly observe and quantify the anticipation, we note that if firms respond because of anticipation, then their response should increase with the expected burden of the rule proposals. This expected burden is a function of two factors: the probability that the proposals will succeed, and the significance of the proposals. We design two tests to support this interpretation.

Our first test exploits the fact that the Agenda differentiates between two tiers of rules.[35] The top tier includes rules which are *significant* (raise novel legal issues) or *economically significant* (annual economic effect of $100 million or more). The lowest tier includes rules that are *substantive* (important but not significant) or *administrative* (how federal agencies govern themselves). We group the four categories into two tiers based on our conversations with former OIRA officials, from whom we learned that the differences between the categories from the same tier are not strictly maintained in practice.[36] We then construct two measures of *RegPipeline*, one for each tier, and estimate the baseline specification (Equation (4)) separately using each measure. The results are in Table 5, Panel A. Clearly, the effect on firm-level outcomes is driven by high-impact rules from the top tier. For SGA and net income, low-tier rules have no statistically significant effect, as opposed to top-tier rules. For COGS, the effect of low-tier rules is statistically significant but 50% lower than top-tier rules. For all three outcomes, the effect of top-tier rules is more than twice the baseline effect reported in Table 4, where we pool all rules together.

Our second test is focused on the probability of conversion. We first estimate a linear

[35]The system was established in 1993, following Executive Order 12866 which subjected significant regulations to a more rigorous vetting process.

[36]The top tier includes 29% of the rules (see Table 2 and Table 1, Panel C).

probability model in the cross-section of rule proposals:

$$\mathbb{1}(Rule)_r = Length_r + \lambda_{a,t,i}, \tag{5}$$

where the outcome equals one for proposals that yielded a final rule, and zero otherwise. We control for the length of the rule's language as a possible predictor. We then add multiplicative fixed effects for agency, cohort (year in which the rule entered the pipeline), and tier (whether the rule is significant or not). In other words, we compare rule proposals that were initiated by the same agency at the same time and with similar significance. We define the predicted values from Equation (5) as the ex-ante probability that the proposal will successfully navigate the pipeline and become an effective rule.[37] We split the sample of proposals by the median value, and construct two measures of $RegPipeline$ for rules with high (above the median) and low (at or below the median) ex-ante probability. Finally, we estimate the baseline specification (Equation (4)) twice, each time using only one of the two firm-level measures. The results are presented in Table 5, Panel B. For low-likelihood rules, the effects are close to zero and are all statistically insignificant, except for SGA in a specification which does not control for $RegIn$ and $PRisk$. In contrast, the effect of high-likelihood rules is significant and more than twice the baseline effect reported in Table 4, where we pool all rules together.

4.3 Uncertainty

The regulatory pipeline is associated with at least three aspects of uncertainty. First, how long would it take the rule proposal to clear the pipeline (timing uncertainty). Second, when clearing the pipeline, will the proposal convert into a final rule or be rescinded (outcome uncertainty). Third, if converting into a final rule, how would it differ from the initial draft proposal (content uncertainty). Thus, firms with high exposure to the pipeline operate

[37] We discuss the prediction model and its output in Appendix A.4.

under significant uncertainty. While waiting for the uncertainty to resolve, managers have an incentive to postpone investments that can be delayed (McDonald and Siegel (1986); Bernanke (1983); Julio and Yook (2012); Gulen and Ion (2015)). Consequently, if uncertainty is an important channel, we expect regulatory pipeline to have a significant impact on capital investment. We conduct two tests to investigate this possibility.

Our first test looks into the average impact on capital investment. We estimate a specification similar to Equation (4), except that the outcome is capital expenditures scaled by beginning-of-period total assets ($CAPX$). The results are in the first three columns of Table 6. Consistent with our expectation, we find that regulatory pipeline leads to a substantial decline in capital investment. The effect is significant across all specifications, including the one with the tightest set of fixed effects and controls. This suggests that regulatory pipeline acts as a source of uncertainty, motivating companies to sit on the sidelines and reduce their capital investments.

Our second test exploits the heterogeneity across rules. We hypothesize that rule proposals with unique content create more uncertainty, and therefore have a larger adverse effect on capital investment. Intuitively, if the new rule proposal is significantly different from other rule proposals, firms will struggle to interpret the proposal and to predict the probability it will convert into a final rule. Therefore, the proposal will create more legal uncertainty. In contrast, if the new proposal is similar to others, it will only modestly increase uncertainty. To test this possibility, for each focal rule i, we compute the Hellinger distance to all other peer rules $j \neq i$ issued by the same agency.[38] We then take the average distance across all peer rules. A higher score means that the focal rule is uniquely focused on regulatory topics that the agency does not normally get involved in. We split the sample of proposals by the median score and construct a measure of $RegPipeline$ for rules with high (above median) and low (at or below median) score. Finally, we estimate the baseline specification (Equation (4)) twice, each time using only one of the two firm-level measures. The results

[38]The Hellinger distance helps compare two distributions, which in our case are the distributions of LDA topics in rule i and rule j.

are presented in the last six columns of Table 6. Clearly, the effect on investment is driven almost entirely by unique rule proposals. Those have a high distance score from other rules and create more uncertainty, which has a dampening effect on investment. The effect of typical rules, which are similar to the agency's core regulatory areas, are close to zero and are all statistically insignificant.

In sum, the evidence suggests that *RegPipeline* affects firm decisions also via uncertainty. The legal uncertainty motivates firms to reduce their capital investment, especially when the pipeline includes proposals with unusual content (high distance score from the agency's typical LDA topic allocation).[39]

4.4 Economies of scale

Regulatory pipeline could have a differential impact on large and small firms. On the one hand, if compliance requires mainly a large fixed cost and fewer variable costs, then large firms are in an advantageous position due to economies of scale. For example, a company with an existing compliance division can absorb several more potential regulations, while a small company might struggle. Moreover, large firms could be more successful in their efforts to subvert regulatory requirements. This suggests that larger firms should be less sensitive to the pipeline. On the other hand, large firms are more likely to receive attention from regulatory agencies. For example, regulators could choose to tighten regulatory requirements for larger firms to advance their own career prospects (deHaan et al. (2015)).

To test the competing predictions, we estimate a version of Equation (4):

$$y_{i,t+l} = \alpha + \beta_1 \cdot RegPipeline_{i,t} + \beta_2 \cdot Large_{i,t} \cdot RegPipeline_{i,t} + \overrightarrow{X} + \epsilon_{i,t+l}, \tag{6}$$

where *Large* is an indicator for large companies. Each quarter, we assign firms from the

[39]Analogously, we find that *RegPipeline* increases cash holdings, especially when the pipeline includes proposals with unusual content. This is also consistent with the uncertainty channel, whereby firms are motivated to pull out of investment projects and simultaneously build up cash reserves to handle future regulatory developments.

same Fama-French 48 industry into two groups based on whether the firm's total assets lie above or below the industry's median. If it is higher than the median, then $Large_{i,t} = 1$; otherwise, $Large_{i,t} = 0$. The results are reported in Table 7, columns 1-3 (for SGA) and columns 5-7 (for COGS). Across all specifications, the effect of regulatory pipeline is 30-78% lower for large companies. This conclusion applies both for SGA and for COGS, especially for the former, and remains similar in the tightest specification with year×industry fixed effects and the full set of controls. We further test whether the decline in costs is linear. Instead of an indicator variable, we use a category variable for the five quintiles within the quarter×industry. The results are in columns 4 and 8 of Table 7. We find that the effect of regulatory pipeline declines monotonically across size quintiles. For companies in the top quintile, the effect on COGS is not distinguishable from zero while the effect on SGA is negative.

4.5 Political investment

In Section 4, we show that firms increase their overhead costs when they have greater exposure to the regulatory pipeline. This is consistent with a reactive mode, where firms take the pipeline as given and repurpose their resources as best as possible. Another possible reaction is to try and shape the oncoming regulation via lobbying (Bertrand, Bombardini, and Trebbi (2014); Blanes i Vidal, Draca, and Fons-Rosen (2012); Borisov, Goldman, and Gupta (2015)). On one hand, a firm with higher exposure to the regulatory pipeline could spend more on lobbying for at least two reasons: to reduce future regulatory burden, or to limit the impact of any pending rule proposal. On the other hand, if the firm is exposed to more potential regulations, it could limit the benefit from capturing regulators via lobbying. In other words, the return per dollar spent on lobbying declines if the firm must handle multiple regulatory actions.

To test the competing predictions, we match the Compustat sample to lobbying data from Kim (2018). The information is collected from mandatory filings, published regularly by the Secretary of the U.S. Senate. We first look at how *RegPipeline* affects the probability of any lobbying in a given year (while Compustat data is available at a quarterly frequency, lobbying data is at an annual frequency). Conditional on any lobbying during the year, we look at the dollar expenses and at the number of federal agencies the firm has been lobbying. Note that we retain only firms that have any lobbying activity at least once during the sample period. This leaves a sample of 1,548 unique firms and 12,877 firm×year observations. When we condition on any lobbying during the year, and add controls and fixed effects, the sample is reduced to 831 unique firms and 5,170 firm×year observations.

The results are summarized in Table 8. In the first specification we use year and firm FE and standard controls. In the second specification we replace year with year×industry fixed effects. In the third specification we control for regulation quantity ($RegIn$) and political risk ($PRisk$), in addition to the baseline vector of controls which are included in all specifications. In all three specifications, we find that pipeline exposure does not change the probability of lobbying: the coefficient on *RegPipeline* is positive but statistically insignificant. While the extensive margin is not sensitive to the pipeline, we observe a significant decline in the scope of lobbying activities on the intensive margin (conditional on having any lobbying). The effect is statistically significant and economically large: a one-standard-deviation increase in *RegPipeline* reduces lobbying expenses by $101,000-$98,000 and the number of lobbied federal agencies by 0.3. These effects are quite similar across industries (industry FE) and within industries (year×industry FE), and conditional on the full set of controls.

In sum, exposure to the regulatory pipeline seems to limit the benefit from lobbying. Firms appear to scale back their lobbying activities while still maintaining some lobbying footprint. This result is similar to Kalmenovitz, Lowry, and Volkova (2023), who find that the fragmentation of regulation across multiple agencies leads to fewer lobbying activities. Overall, this suggests that regulatory pipeline reduces the benefits from regulatory capture via lobbying.

4.6 Alternative explanations

Our main results show that regulatory pipeline increases costs and reduces profits. This is consistent with a causal interpretation and theoretical predictions. However, we recognize that causal inference is challenging since regulatory pipeline is not randomly assigned across companies. In this section we discuss specific concerns and how we aim to address them. The results are summarized in Table 9.

One possibility is that *RegPipeline* is correlated with current regulatory burden. As explained, our baseline specification controls for the number of paperwork regulations ($RegIn$) and political risk ($PRisk$). Moreover, regulatory pipeline is generally associated with lower regulatory burden across a broad range of measures (see Section 3.3). Thus, if current burden is an omitted variable associated with higher costs, then the reported coefficients are attenuated. To further address this concern, we assign firms from the same Fama-French 48 industry into two groups based on whether a firm's measure of burden lies above or below the industry's median. We then interact the sorting variable with *RegPipeline* and estimate the baseline specification (Equation (4)) with SGA, COGS, and net income. We use the tightest specification with time×industry fixed effects and the full set of controls. The results are in columns (1), (4), and (7) in Table 9. The coefficient on the interaction term is statistically insignificant and effectively zero. This suggests that regulatory pipeline increases costs and reduces profits across the board, regardless of how burdensome the current regulatory environment is for the company. In another test, we augment the vector of controls with two additional proxies for regulatory burden: *CFR* and *10K*. These represent the firm's exposure to the Code of Federal Regulations and how intensively the firm discusses regulatory affairs in its annual 10-K form (see Appendix A.3). Due to data availability, the sample shrinks by more than half. Nevertheless, all the results hold.

Another possibility is that the results are driven by the fragmentation of the regulatory pipeline. This is motivated by Kalmenovitz, Lowry, and Volkova (2023), who introduce the concept of regulatory fragmentation: when multiple federal agencies regulate a single topic.

We construct a measure which represents the fragmentation of the regulatory pipeline, denoted with $RegPipeline^{frag}$.[40] We then re-estimate our regressions with the addition of $RegPipeline^{frag}$ and present the results in columns (2), (5), and (8) in Table 9. The coefficient on our primary measure, $RegPipeline$, remains large and statistically significant. Compared to specifications without controlling for $RegPipeline^{frag}$, the coefficients are slightly larger. This suggests that pipeline size ($RegPipeline$) is a dominant source of burden, while its fragmentation ($RegPipeline^{frag}$) might impose additional burden.

Finally, $RegPipeline$ could be correlated with the maturity of the firm. Mature firms likely operate across multiple areas, which leads to higher costs and smaller profit margins, while simultaneously increasing their exposure to a broad collection of proposed federal regulations. To test this we control for the firm's age and for the number of business segments, both derived from Compustat. The results are in columns (3), (6), and (9) in Table 9. The coefficients on $RegPipeline$ remain statistically significant and comparable to those from the baseline specification.

In sum, the adverse impact of $RegPipeline$ is not explained by firm complexity and maturity, the current level of regulatory burden, or by the fragmentation of regulation across agencies. Instead, the results are consistent with anticipatory effects on firms.

5 Systematic risk

In this section, we investigate how the regulatory pipeline relates to systematic risk. Concretely, we identify regulatory topics which are systematically important. Similar to other studies,[41] we consider topic o to be systematically important if it significantly increases the sensitivity of firms to macroeconomic shocks. As a stylized example, suppose we find that the topic Pharmaceuticals is systematically important. This would mean that firms with

[40] See details of our methodology in Appendix A.3. $RegPipeline^{frag}$ is highly correlated with overall fragmentation from Kalmenovitz, Lowry, and Volkova (2023), but negatively correlated with our primary measure, $RegPipeline$. In other words, when the regulatory pipeline increases, it tends to be concentrated among a small number of federal agencies.

[41] E.g., Kelly, Pruitt, and Su (2019), Lopez-Lira (2019), and Giglio et al. (2023).

high exposure to pharmaceuticals regulations are more sensitive to macroeconomic conditions, such as pandemic risk and healthcare reforms. Therefore, these firms will be more exposed to systematic risk.

To identify systematically important topics, we consider the following model for returns (similar to Kelly, Pruitt, and Su (2019)):

$$r_{i,t+1} = \alpha_{i,t} + \beta'_{i,t} f_{t+1} + \epsilon_{i,t+1},\ i = 1, ..N_t,\ t = 1, ..T \tag{7}$$

where $r_{i,t+1}$ is the return for stock i between t and $t+1$. The vector f_{t+1} contains L unobserved common factors, $\beta_{i,t}$ is the firm-specific sensitivity to those factors, and $\alpha_{i,t}$ is unrelated to the systematic risk but affects the cross-sectional variation of returns. In our model, β is affected by the regulatory pipeline in the following manner:

$$\beta_{i,t} = \Gamma'_b x_{i,t},$$

where $x_{i,t}$ represents the exposure of firm i to each regulatory topic, and Γ_b represents the exposure of each topic to each latent macroeconomic factor. The firm-topic vector ($x_{i,t}$) is of size K, based on $K = 100$ topics, and we can measure it directly.[42] The object of interest is the topic-factor vector (Γ_b). It is of dimension $K \times L$, based on K topics and L latent factors, and it does not vary over time or across firms. Our goal is to estimate Γ_b, and thus determine which topics have a high sensitivity to latent macroeconomic factors. These topics will yield higher β and hence increase the exposure to systematic risk.[43]

To estimate Γ_b, we follow the bootstrapping procedure in Kelly, Pruitt, and Su (2019).

[42] A typical component is $x_{i,o,t} = \omega_{i,o,t} \cdot TopicPipeline_{o,t} > 0$, where $\omega_{i,o}$ is the proportion of the text assigned by firm i to topic o (Equation (3)), and $TopicPipeline_{o,t}$ is the fraction of the regulatory pipeline devoted to this topic (Equation (2)). Here, we use the log-transformation of $x_{i,o,t}$ and then standardize cross-sectionally each period, to make the procedure numerically stable.

[43] Equivalently, we have $\alpha_{i,t} = \Gamma'_\alpha x_{i,t}$, where Γ_a determines how each topic affects stock returns unrelated to the systematic risk factors. Γ_a is of dimension $K \times 1$, based on K topics, and also does not vary over time or across firms. Our baseline test uses five factors and we do not impose $\Gamma_\alpha = 0$. In words, we allow regulation to affect stock returns independently of common factor exposures.

For brevity, the full derivations and technical explanations are in Appendix A.5. Intuitively, we seek to test the impact of each topic o in terms of statistical significance and economic magnitude. We estimate a model of stock returns with and without topic o, that is, setting the corresponding row in Γ_b to zero. We then compute the total R^2 in both scenarios, which represents the model's ability to explain the time-series variation in stock returns. We repeat the process 1,000 times for each topic to quantify the statistical significance, and adjust for multiple testing using the Holm (1979) method. Finally, we categorize topic o as systematically important if it satisfies two conditions: (1) the corresponding row in Γ_b is significantly different from zero, and (2) its contribution to R^2 is amongst the top.

The results are summarized in Table 10. Starting with the incremental contribution to the R^2, we identify the 20 topics with the largest impact. Imposing the additional restriction of statistical significance, we highlight a narrower set of six regulatory topics that satisfy both criteria. Each of these topics is statistically significant and can explain more than 10 basis points of the R^2. These topics are related to pharmaceuticals (topic 93), healthcare (topic 24), natural resources (topic 74), securities markets (topic 14), and environmental protection (topic 59 and topic 97). Firms that are highly exposed to either topic will be more sensitive to macroeconomic shocks, hence increasing their exposure to systematic risk.

6 Conclusions

We develop the first firm-level measure of regulatory pipeline. It counts the number of rule proposals, which are currently pending in the federal government's rulemaking pipeline, and will apply to the individual firm if converted into a final rule. The measure is based on a novel administrative data set that tracks all 43,000 rule proposals since 1995 and allows us to reconstruct the precise timeline of each proposal. On an average day, the pipeline consists of 3,500 rule proposals. One-third of the proposals are rescinded, and the rest successfully convert into a final rule after spending nearly two years (on average) in the pipeline.

To develop the firm-specific measure of regulatory pipeline, we assign rules from the pipeline to firms based on linguistic similarities (a machine-learning algorithm known as LDA). We find that exposure to the regulatory pipeline has significant anticipatory effects. Firms with higher exposure express more concerns about future political risk (based on data from Hassan et al. (2019)), increase their overhead costs (SGA and COGS), and suffer from lower profits. The effects are identified within industries and are independent of the firm's current regulatory burden and political risk, which we add as control variables to all our regressions. Moreover, anticipating future regulatory changes, firms build up cash reserves and reduce capital investment. This is strongly suggestive of regulatory pipeline acting as an independent source of uncertainty, motivating companies to invest in political connections and prepare for potential future shocks. Small firms are especially responsive to the regulatory pipeline, highlighting the role of economies of scale.

Overall, our results are the first to consistently document substantial anticipatory effects across all federal regulations. The main innovation is to focus on potential regulations and shed light on a relatively unknown aspect of regulatory burden: the amount of potential future regulations that are relevant to the firm. Future studies can dig deeper into the origins and consequences of anticipatory effects, tactics adopted by companies to mitigate these effects, and potential heterogeneity across firms and rules.

References

Administrative Conference of the United States. 2015. The Unified Agenda: Proposals for Reform. *Available online* .

Al-Ubaydli, O., and P. A. McLaughlin. 2017. Regdata: A numerical database on industry-specific regulations for all united states industries and federal regulations, 1997–2012. *Regulation & Governance* 11:109–23.

Baker, S. R., N. Bloom, and S. J. Davis. 2016. Measuring economic policy uncertainty. *Quarterly Journal of Economics* 131:1593–636.

Bao, Y., and A. Datta. 2014. Simultaneously Discovering and Quantifying Risk Types from Textual Risk Disclosures. *Management Science* 60:1371–91.

Becker, G. S. 1983. A theory of competition among pressure groups for political influence. *Quarterly Journal of Economics* 98:371–400.

Benmelech, E., and T. J. Moskowitz. 2010. The political economy of financial regulation: Evidence from US state usury laws in the 19th century. *Journal of Finance* 65:1029–73.

Bernanke, B. S. 1983. Irreversibility, uncertainty, and cyclical investment. *Quarterly Journal of Economics* 98:85–106.

Bertrand, M., M. Bombardini, and F. Trebbi. 2014. Is it whom you know or what you know? an empirical assessment of the lobbying process. *American Economic Review* 104:3885–920.

Bessembinder, H., S. Jacobsen, W. Maxwell, and K. Venkataraman. 2018. Capital commitment and illiquidity in corporate bonds. *The Journal of Finance* 73:1615–61.

Bird, A., A. Ertan, S. A. Karolyi, and T. Ruchti. 2020. Information, financing, and investment. *Unpublished working paper* .

Blanes i Vidal, J., M. Draca, and C. Fons-Rosen. 2012. Revolving door lobbyists. *American Economic Review* 102:3731–48.

Borisov, A., E. Goldman, and N. Gupta. 2015. The corporate value of (corrupt) lobbying. *The Review of Financial Studies* 29:1039–71.

Borochin, P., M. A. Celik, X. Tian, and T. M. Whited. 2021. Identifying the heterogeneous impact of highly anticipated events: Evidence from the tax cuts and jobs act. *Unpublished working paper* .

Brown, S., S. A. Hillegeist, and K. Lo. 2004. Conference calls and information asymmetry. *Journal of Accounting and Economics* 37:343–66. ISSN 0165-4101. doi:10.1016/J.JACCECO.2004.02.001.

Bybee, L., B. T. Kelly, A. Manela, and D. Xiu. 2019. The Structure of Economic News. *Unpublished working paper* .

Bybee, L., B. T. Kelly, and Y. Su. 2022. Narrative Asset Pricing: Interpretable Systematic Risk Factors from News Text. *Unpublished working paper* .

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

Cao, S., W. Jiang, B. Yang, and A. L. Zhang. 2023. How to talk when a machine is listening: Corporate disclosure in the age of ai. *The Review of Financial Studies* 36:3603–42. ISSN 0893-9454. doi:10.1093/RFS/HHAD021.

Chen, J., and J. Kalmenovitz. 2023. Regulatory similarity. *Journal of Law and Economics* .

Chen, Q., N. L. Roussanov, and X. Wang. 2021. Semiparametric Conditional Factor Models: Estimation and Inference. *Unpublished working paper* .

Correia, M., J. Friedman, M. Gutman, D. Jurafsky, S. Larcker, A. Leone, S. Lobanov, M. Angel, M. Meza, M. Ogneva, B. Tayan, D. F. Larcker, and A. A. Zakolyukina. 2012. Detecting deceptive discussions in conference calls. *Journal of Accounting Research* 50:495–540. ISSN 1475-679X. doi:10.1111/J.1475-679X.2012.00450.X.

Davis, A. K., W. Ge, D. Matsumoto, and J. L. Zhang. 2015. The effect of manager-specific optimism on the tone of earnings conference calls. *Review of Accounting Studies* 20:639–73. ISSN 13806653. doi:10.1007/S11142-014-9309-4/TABLES/7.

deHaan, E., S. Kedia, K. Koh, and S. Rajgopal. 2015. The revolving door and the SEC's enforcement outcomes: Initial evidence from civil litigation. *Journal of Accounting and Economics* 60:65–96.

Demsetz, H. 1974. Toward a theory of property rights. In *Classic papers in natural resource economics*, 163–77. Springer.

Fama, E. F., and J. D. MacBeth. 1973. Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy* 81:607–36. ISSN 00223808, 1537534X.

Giglio, S., T. Kuchler, J. Stroebel, and X. Zeng. 2023. Biodiversity risk. *Unpublished working paper* .

Gulen, H., and M. Ion. 2015. Policy uncertainty and corporate investment. *The Review of Financial Studies* 29:523–64.

Hanley, K. W., and G. Hoberg. 2019. Dynamic Interpretation of Emerging Risks in the Financial Sector. *The Review of Financial Studies* .

Hansen, S., M. McMahon, and A. Prat. 2018. Transparency and Deliberation Within the FOMC: A Computational Linguistics Approach. *The Quarterly Journal of Economics* 133:801–70.

Hassan, T. A., S. Hollander, L. van Lent, and A. Tahoun. 2019. Firm-level political risk: Measurement and effects. *Quarterly Journal of Economics* 134:2135–202.

Hendricks, B. E., J. J. Neilson, C. Shakespeare, and C. D. Williams. 2023. Anticipatory effects around proposed regulation: Evidence from basel iii. *The Accounting Review* 98:285–315.

Holm, S. 1979. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 6:65–70. ISSN 03036898, 14679469.

Israelsen, R. D. 2014. Tell It Like It Is: Disclosed Risks and Factor Portfolios. *Unpublished working paper* .

Jha, A., S. A. Karolyi, and N. Z. Muller. 2020. Polluting public funds: The effect of environmental regulation on municipal bonds. *Unpublished working paper* .

Jiang, F., J. Lee, X. Martin, and G. Zhou. 2019. Manager sentiment and stock returns. *Journal of Financial Economics* 132:126–49. ISSN 0304-405X. doi:10.1016/J.JFINECO.2018.10.001.

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

Kalmenovitz, J. 2023. Regulatory intensity and firm-specific exposure. *The Review of Financial Studies* .

Kalmenovitz, J., M. Lowry, and E. Volkova. 2023. Regulatory fragmentation. *Journal of Finance* .

Kelly, B. T., S. Pruitt, and Y. Su. 2019. Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics* 134:501–24.

Khan, S. Y., and A. Khederlarian. 2021. How does trade respond to anticipated tariff changes? evidence from nafta. *Journal of International Economics* 133:103538–.

Kim, I. S. 2018. Lobbyview: Firm-level lobbying & congressional bills database. *Working paper* .

Krueger, A. O. 1974. The political economy of the rent-seeking society. *American Economic Review* 64:291–303.

Lopez-Lira, A. 2019. Risk Factors That Matter: Textual Analysis of Risk Disclosures for the Cross-Section of Returns. *Unpublished working paper* .

Lowry, M., R. Michaely, and E. Volkova. 2020. Information revealed through the regulatory process: Interactions between the sec and companies ahead of their IPO. *The Review of Financial Studies* 33:5510–54.

Mayew, W. J., M. Venkatachalam, D. Ariely, J. Bettman, L. Cohen, P. Dechow, L. Koonce,

F. Li, M. F. Luce, G. Miller, C. Moorman, C. Parsons, E. Riedl, K. Schipper, S. Sunder, P. Tetlock, T. Wong, D. Ames, E. Ames, J. Ames, P. Badolato, Z. Chen, A. Gupta, S. Li, M. Uh, and Y. Zhou. 2012. The power of voice: Managerial affective states and future firm performance. *The Journal of Finance* 67:1–43. ISSN 1540-6261. doi:10.1111/J.1540-6261.2011.01705.X.

McDonald, R., and D. Siegel. 1986. The value of waiting to invest. *Quarterly Journal of Economics* 101:707–27.

Melody, W. H. 2016. Institutionalizing "the public interest" in public utility regulation: Harry M. Trebing and the second wave of reform. *Journal of Economic Issues* 50:519–26.

Nanda, G., A. Jaiswal, H. Castellanos, Y. Zhou, A. Choi, and A. J. Magana. 2023. Evaluating the Coverage and Depth of Latent Dirichlet Allocation Topic Model in Comparison with Human Coding of Qualitative Data: The Case of Education Research. *Machine Learning and Knowledge Extraction* 5:473–90.

Nou, J., and E. H. Stiglitz. 2015. Strategic rulemaking disclosure. *S. Cal. L. Rev.* 89:733–.

Peltzman, S. 1976. Toward a more general theory of regulation. *The Journal of Law and Economics* 19:211–40.

Pigou, A. 1938. *The economics of welfare*. Routledge.

Posner, R. A. 1974. *Theories of economic regulation*. National Bureau of Economic Research Cambridge, Mass., USA.

Rennekamp, K. M., M. Sethuraman, and B. A. Steenhoven. 2022. Engagement in earnings conference calls. *Journal of Accounting and Economics* 74:101498–. ISSN 0165-4101. doi:10.1016/J.JACCECO.2022.101498.

Stigler, G. J. 1971. The theory of economic regulation. *The Bell Journal of Economics and Management Science* 3–21.

Trebbi, F., and K. Xiao. 2019. Regulation and market liquidity. *Management Science* 65:1949–68.

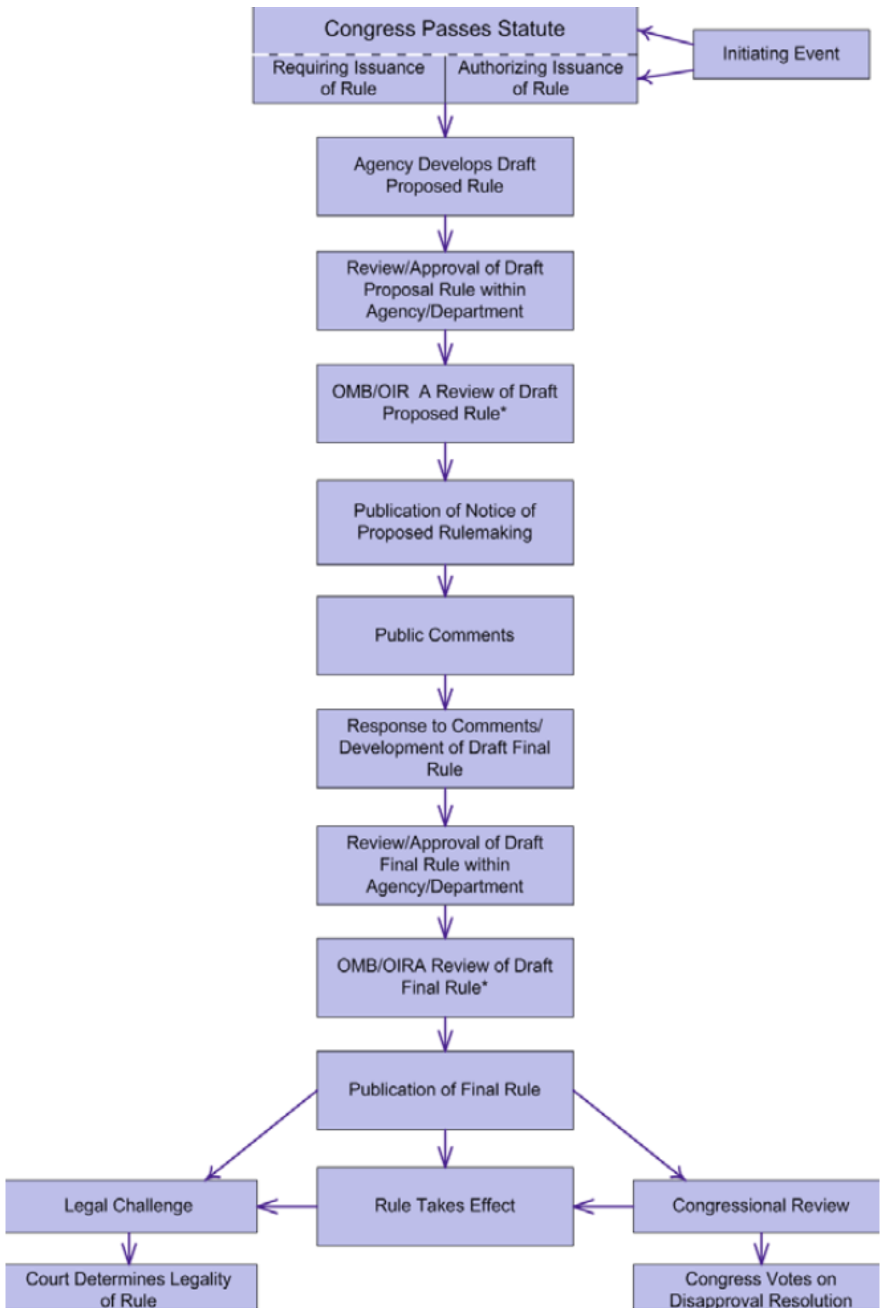
Trebbi, F., and M. B. Zhang. 2022. The cost of regulatory compliance in the United States. Working Paper, National Bureau of Economic Research.

Tullock, G. 1967. The welfare costs of tariffs, monopolies, and theft. *Economic Inquiry* 5:224–32.

Yackee, J. W., and S. W. Yackee. 2010. Administrative procedures and bureaucratic performance: Is federal rule-making "ossified"? *Journal of Public Administration Research and Theory* 20:261–82.

Figure 1: **Rulemaking process**

Panel A. The figure describes the federal rulemaking process. Reproduced from "The Federal Rulemaking Process: An Overview" by the Congressional Research Service.



Panel B. The figure plots the distribution of time in the pipeline. We focus on subset of 27,094 proposals that eventually converted into a final rule. For each proposal, we compute the number of days that the proposal was in the rulemaking pipeline, from the moment it was first announced and until the final rule was published. The mean and median values are 668 and 365 (22 and 12 months), respectively.

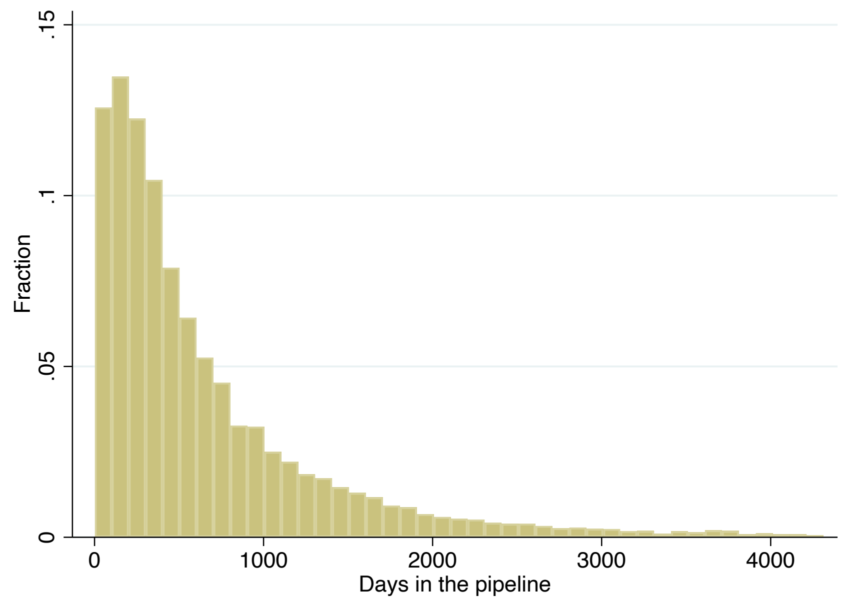


Figure 2: **Examples of regulatory topics**

Word clouds depicting selected topics. Font size reflects the relative frequency of the word within the topic. The topics, from left to right, are 1, 26 and 27 (top row); and 31, 45, and 83 (bottom row).

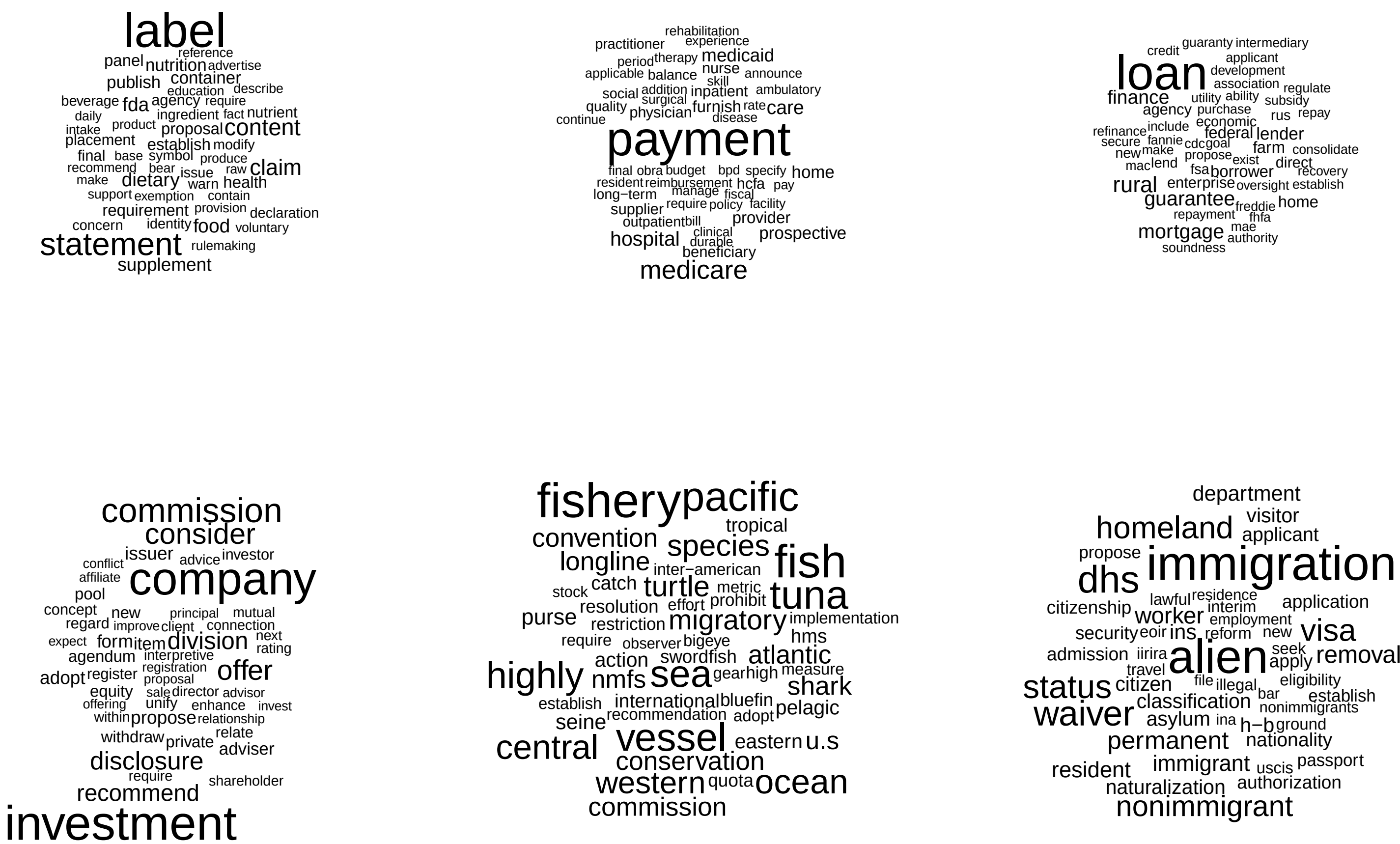
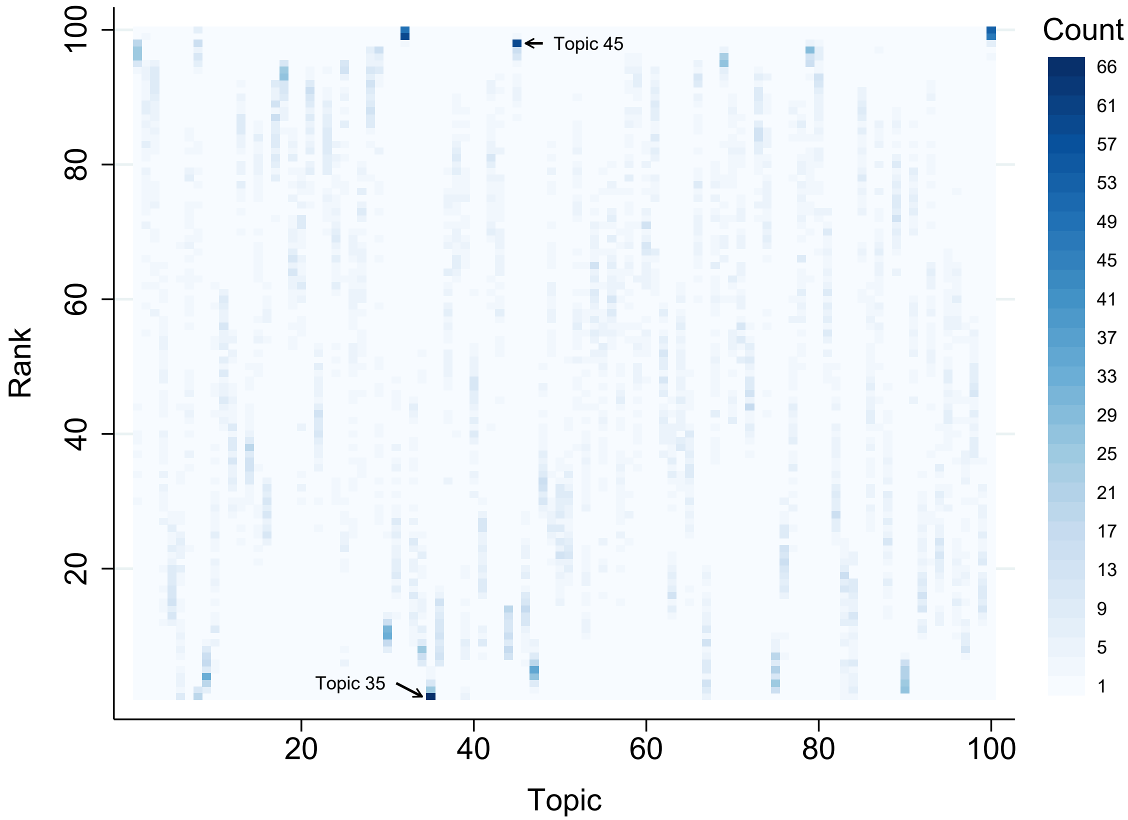
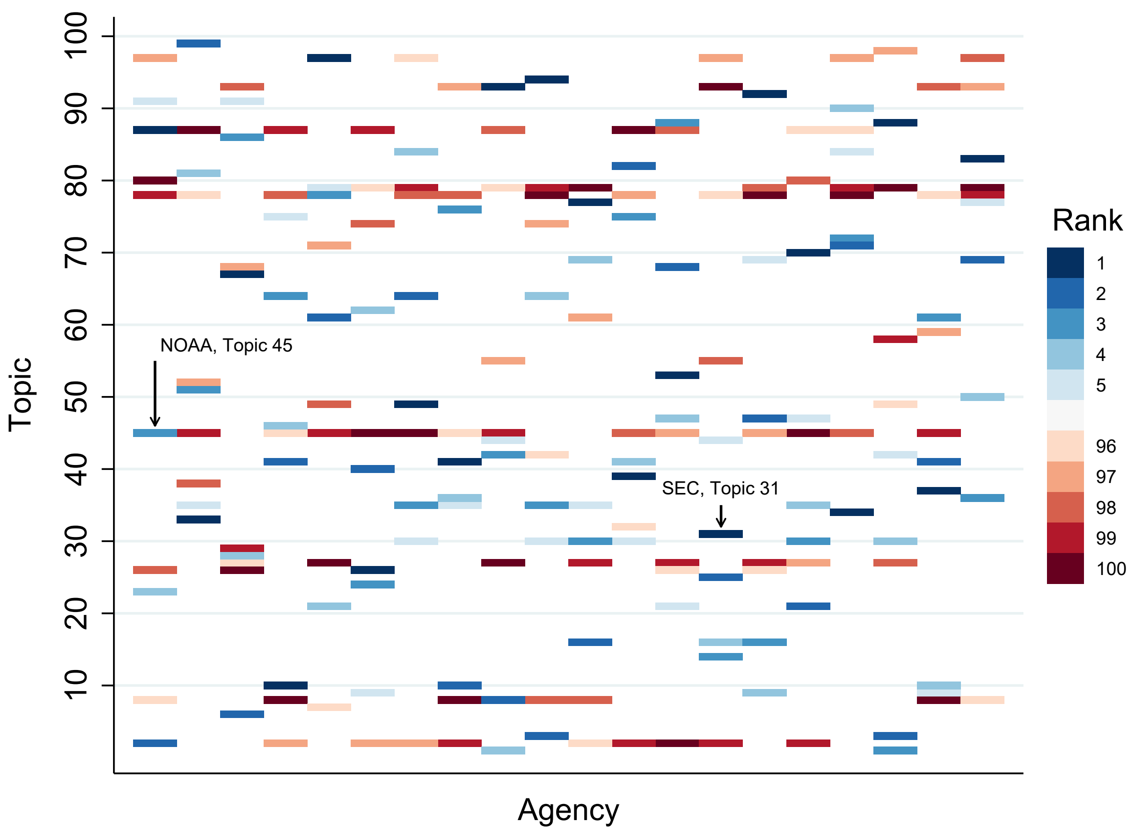


Figure 3: **Topic distribution**

Panel A. This figure plots the frequency of each topic's ranking throughout the history of the federal rulemkaing pipeline. Each quarter we rank the topics from one to 100 based on their weights in the rulemaking pipeline, where the topic with the highest weight is ranked one. We then count, for each topic and ranking, the number of quarters in which the topic has achieved the given ranking. See a printer-friendly black and white replication of this figure in Figure A.2, Panel A.



Panel B. This figure plots the most and least dominant topics for the top 20 agencies in terms of the total number of rules in the Unified Agenda from 1995-2022. For each of the top 20 agencies, we calculate the average weight of each topic across the agency's history of rules and report the top and bottom five topics for each agency. A rank of one indicates the most dominant topic for the agency. See a printer-friendly black and white replication of this figure in Figure A.2, Panel B.



Panel C. This figure plots the most and least dominant topics for each of the Fama-French 48 industries. We start with the topic weights for our firms each quarter and then calculate the average weight of each of the 100 topics for firms in the same Fama-French 48 industry. A rank of one indicates the most dominant topic for the industry. See a printer-friendly black and white replication of this figure in Figure A.2, Panel C.

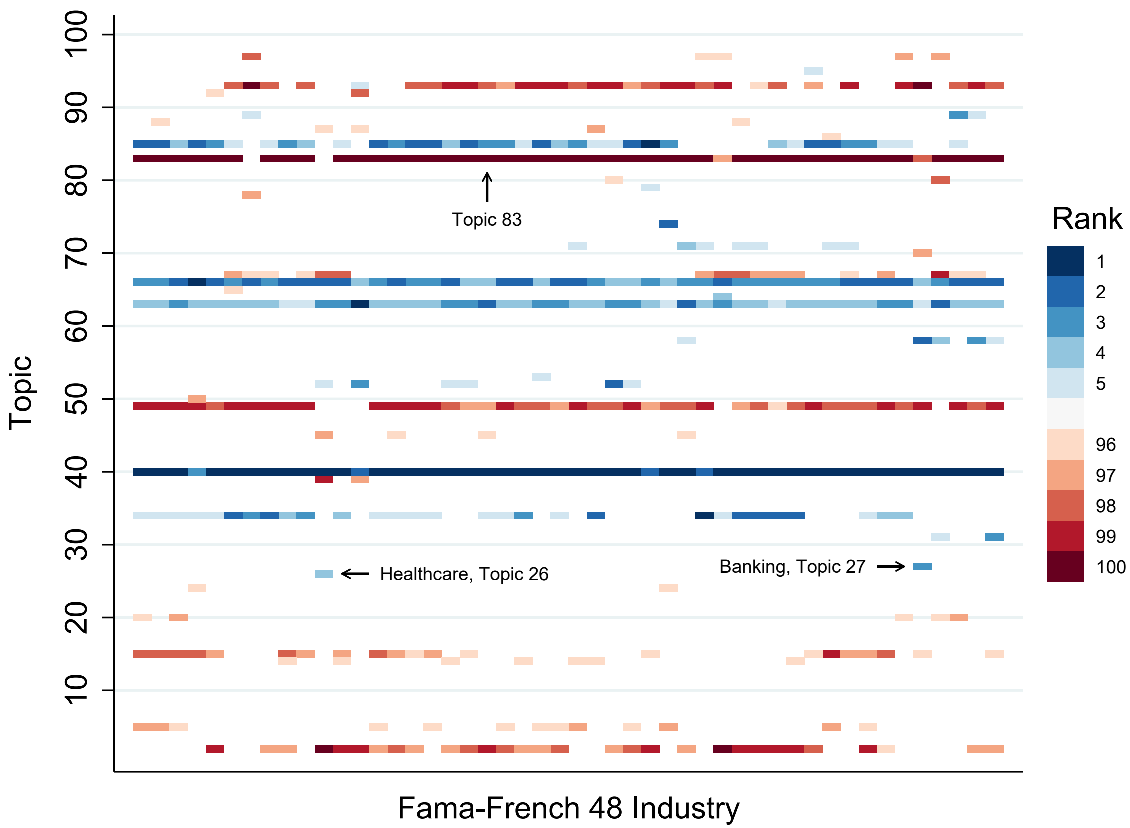


Figure 4: **Commonalities in regulatory pipeline**

This figure illustrates the the pairwise correlations of $RegPipeline$ for the Fama-French 48 industries. We first residualize out the time variation of $RegPipeline$ and then, for each quarter, calculate the average regulatory pipeline for firms in the same Fama-French 48 industry. We then calculate the correlation of $RegPipeline$ between each pair of industries. The figure is a 48×48 square, where the shading of each square indicates the correlation of an industry pair. The industries have been ordered by their average correlation with the other industries, such that the first column on the left is the industry with the highest average correlation with the other industries. See a printer-friendly black and white replication of this figure in Figure A.3.

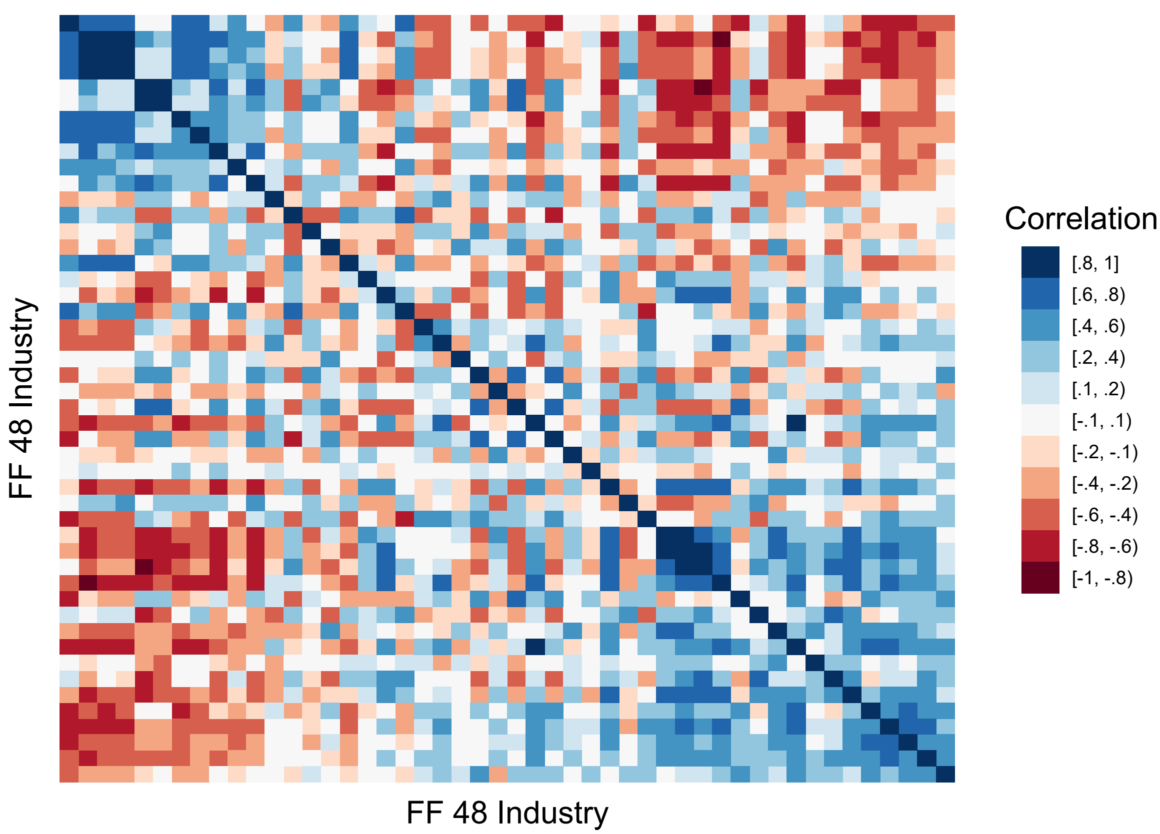


Table 1: **Summary statistics: rulemaking pipeline**

The sample includes all federal regulations being developed between 1995 and Spring 2023. A regulation is identified by its unique RIN. *Alive* = 1 if the RIN was still under development by Spring 2023. Conditional on *Alive* = 0, *Rule* = 1 if the RIN successfully converted into a rule; *Repeal* = 1 if the RIN was officially withdrawn before any rule was published; and *Mixed* = 1 if the RIN was partially successful (part of it was codified into a rule while part of it of withdrawn). *Frozen* = 1 if the RIN was put on hold ("long-term action") at least once. *PendingTime* is the number of days the rule has been in the pipeline (or still is, if *Alive* = 1). Tier 1 includes significant rules, economically or otherwise, and Tier 2 includes substantive and administrative rules.

	Mean	**SD**	**Min**	**Max**	**Obs.**
Outcome:					
Alive	7.5	26.3	0.0	100.0	42,934
Rule	67.1	47.0	0.0	100.0	39,735
Repeal	30.2	45.9	0.0	100.0	39,735
Mixed	2.7	16.2	0.0	100.0	39,735
Frozen	25.6	43.6	0.0	100.0	39,735
Pending Time	869.2	1,184.8	1.0	15,237.0	42,934
If Rule=0:	1,222.1	1,463.5	1.0	12,827.0	13,053
If Rule=1:	640.4	878.1	1.0	11,085.0	26,682
Importance:					
Major	5.2	22.1	0.0	100.0	42,934
Tier1	28.9	45.3	0.0	100.0	42,934
Tier2	71.1	45.3	0.0	100.0	42,934
Substantive:	63.3	48.2	0.0	100.0	42,934
Administrative:	7.8	26.9	0.0	100.0	42,934

Table 2: **Decomposing regulatory pipeline**

Panel A. Variance decomposition. We regress our primary measure of firm-level regulatory pipeline on a growing number of fixed effects, and report the resulting R^2.

	(1)	(2)	(3)	(4)
Industry Classification	FF48	2-digit SIC	3-digit SIC	4-digit SIC
Time FE	49.5%	49.5%	49.5%	49.5%
Industry FE	9.3%	9.6%	12.3%	13.4%
Industry×time FE	1.2%	1.2%	1.8%	2.1%
Subtotal	60.0%	60.2%	63.6%	65.0%
Firm-specific (one minus subtotal)	40.0%	39.8%	36.4%	35.0%
Time-invariant (firm FE)	11.2%	11.0%	8.1%	7.0%
Firm-specific variation (residual)	28.8%	28.7%	28.2%	28.0%
Number of industries	48	68	258	409

Panel B. Distribution of aggregate pipeline. We decompose each rule proposal into 100 topics, and then sum the rule fractions within topic×quarter. The resultant panel includes 100 topics over 112 quarters, from 1995 till 2022. For *Baseline* we use all active rule proposals, and in the remaining lines we use various subsets: rule proposals that are in the top tier (*significant*) and those in the lowest tier; proposals considered to be *major* (a category that could partially overlap with top-tier rules); proposals for which the government has released at least one draft of the planned rule (*with draft*); and proposals which were at least partially codified (*with rule*).

	Mean	**SD**	**Min**	**Max**	**Obs.**
Baseline	36.6	16.5	4.6	119.5	11,200
By importance:					
Top tier	12.9	7.1	0.9	72.9	11,200
Lower tier	23.7	13.0	2.8	108.1	11,200
Major	2.2	2.0	0.0	24.3	11,200
With draft	19.4	9.4	1.0	71.8	11,200
Without draft	17.2	9.7	2.0	98.1	11,200
With rule	2.5	1.7	0.1	23.3	11,200
Without rule	34.2	15.6	4.5	110.7	11,200

Panel C. Sample companies. Each quarter we split companies into five groups, based on their total exposure to the federal rulemaking pipeline (our primary measure of *RegPipeline*), such that the exposure increases across quintiles. We calculate the average characteristic within each group and the difference between the top and bottom quintiles. All differences are significant at the 1% level.

	Q1	Q2	Q3	Q4	Q5	Q5-Q1
Regulation:						
$RegPipeline$	30.7	32.0	33.0	33.6	34.8	4.1
$LobbyEver$	33.6	37.0	38.9	39.9	38.5	4.9
$10K$	4.9	4.8	4.8	4.7	4.3	-0.7
CFR	58,250.6	71,953.5	76,309.2	75,923.0	75,240.4	16,989.8
$RegIn$	108.6	109.2	110.2	111.3	111.8	3.2
$RegIn^{forms}$	120.2	119.5	119.7	120.3	118.4	-1.8
$RegIn^{hours}$	116.1	115.4	115.4	115.8	113.9	-2.2
$RegIn^{dollar}$	114.4	113.4	112.2	111.3	108.1	-6.3
$PRisk$	108.9	119.3	124.4	125.3	121.4	12.5
Financials:						
$Assets$	10,062.8	11,358.4	11,154.3	10,424.5	10,511.9	449.1
SGA	3.9	5.2	6.4	7.4	7.9	3.9
$COGS$	13.0	13.2	13.2	12.7	11.4	-1.6
$NetIncome$	0.1	-0.1	-0.3	-0.5	-0.8	-0.9
$CAPX$	3.1	2.4	2.1	2.1	2.0	-1.1
$TobinQ$	1.7	1.9	2.0	2.1	2.1	0.3
MTB	2.8	3.2	3.3	3.5	3.5	0.7
CF	3.5	2.8	2.4	2.5	2.4	-1.2
$Leverage$	29.7	27.4	25.5	23.7	23.0	-6.7
$Complexity$	96.3	96.4	96.5	96.5	96.2	-0.1
Age	199.2	198.7	154.6	126.9	111.8	-87.4
Observations	39,323	36,907	38,473	38,386	35,666	

Table 3: **Stock and flow of regulations**

We report univariate correlations between our primary measure, $RegPipeline$, and a host of measures from the literature related to regulation: costs of compliance with paperwork regulations (four versions of $RegIn$ from Kalmenovitz (2023)); share of restrictions in the Code of Federal Regulations (CFR, based on data from Al-Ubaydli and McLaughlin (2017)); share of regulatory keywords in the firm's 10-K ($10K$); and discussion of political risk in the firm's conference calls ($PRisk$ from Hassan et al. (2019)).

	$RegPipeline$	CFR	$FedRegister$	$RegIn$	$RegIn^{forms}$	$RegIn^{hours}$	$RegIn^{dollar}$	$10K$	$PRisk$
$RegPipeline$	1.00								
CFR	0.03***	1.00							
$FedRegister$	0.26***	0.00	1.00						
$RegIn$	0.06***	0.07***	0.19***	1.00					
$RegIn^{forms}$	-0.17***	-0.08***	0.06***	0.66***	1.00				
$RegIn^{hours}$	-0.16***	-0.03***	0.10***	0.70***	0.92***	1.00			
$RegIn^{dollar}$	-0.29***	0.04***	0.04***	0.37***	0.57***	0.60***	1.00		
$10K$	-0.17***	-0.07***	0.00	-0.05***	0.10***	0.06***	0.09***	1.00	
$PRisk$	0.03***	0.03***	-0.01	-0.00	-0.02***	-0.03***	-0.01	0.06***	1.00

Table 4: **Costs of regulatory pipeline**

Results from estimating Equation (4). Outcomes represent SG&A costs, cost of goods sold, and net income. $RegPipeline$ is our primary measure of potential regulations relevant to the firm. $Assets$ is total assets, CF are operating cash flows, MTB is market-to-book ratio, $TobinQ$ is Tobin's Q, $Leverage$ is book leverage, $Complexity$ is the dispersion of LDA topics, $RegIn$ is regulatory intensity from Kalmenovitz (2023), and $PRisk$ is political risk from Hassan et al. (2019). See variable definitions in Appendix A.3. Independent variables are lagged and divided by their standard deviation. Standard errors, clustered by firm, are in parentheses.

Outcome:	SGA			$COGS$			$NetIncome$		
$RegPipeline$	0.123***	0.088***	0.103***	0.214***	0.203***	0.312***	-0.067**	-0.047	-0.112**
	(0.032)	(0.032)	(0.039)	(0.068)	(0.067)	(0.085)	(0.033)	(0.033)	(0.047)
$Assets$	-0.789***	-0.676***	-0.805***	-1.361***	-1.354***	-2.171***	-0.196***	-0.234***	-0.143
	(0.107)	(0.108)	(0.177)	(0.255)	(0.246)	(0.522)	(0.049)	(0.051)	(0.104)
$TobinQ$	0.368***	0.419***	0.374***	0.319***	0.402***	0.349***	0.672***	0.615***	0.783***
	(0.058)	(0.059)	(0.068)	(0.096)	(0.097)	(0.116)	(0.049)	(0.050)	(0.066)
MTB	0.519***	0.511***	0.574***	1.105***	1.058***	1.138***	0.183***	0.166***	0.171***
	(0.044)	(0.043)	(0.051)	(0.086)	(0.086)	(0.096)	(0.036)	(0.036)	(0.046)
CF	-0.385***	-0.386***	-0.369***	-0.408***	-0.428***	-0.534***	0.757***	0.792***	0.705***
	(0.031)	(0.031)	(0.037)	(0.055)	(0.057)	(0.075)	(0.038)	(0.039)	(0.052)
$Leverage$	-0.728***	-0.674***	-0.751***	-1.479***	-1.283***	-1.555***	-0.364***	-0.322***	-0.302***
	(0.051)	(0.050)	(0.062)	(0.110)	(0.106)	(0.138)	(0.042)	(0.042)	(0.055)
$Complexity$	-0.026	-0.031*	-0.045**	-0.054	-0.028	-0.082	-0.036**	-0.030*	-0.044**
	(0.016)	(0.017)	(0.021)	(0.041)	(0.041)	(0.055)	(0.017)	(0.017)	(0.022)
$RegIn$			0.122			0.410**			-0.114
			(0.085)			(0.181)			(0.083)
$PRisk$			-0.008			-0.074***			-0.023*
			(0.011)			(0.026)			(0.014)
Obs.	104,369	104,358	64,938	123,795	123,784	67,182	124,394	124,383	67,539
R^2	.914	.918	.923	.892	.901	.916	.632	.653	.61
Firm, Time FE	YES	-	-	YES	-	-	YES	-	-
Firm, Time×FF48 FE	-	YES	YES	-	YES	YES	-	YES	YES

Table 5: **Anticipating the costs of regulatory pipeline**

Results from estimating Equation (4) using multiple versions of *RegPipeline*. In Panel A, we differentiate between significant (columns 1, 3, 5) and non-significant (columns 2, 4, 6) rule proposals. In Panel B, we differentiate between proposals with high (columns 1, 3, 5) and low (columns 2, 4, 6) ex-ante probability of converting into a final rule. We report only the coefficient on *RegPipeline*. Firm controls are *Assets*, *TobinQ*, *CF*, *MTB*, *Leverage*, and *Complexity*. Regulation controls are *RegIn* from Kalmenovitz (2023) and *PRisk* (from Hassan et al. (2019)). See variable definitions in Appendix A.3. Independent variables are lagged and divided by their standard deviation. Standard errors, clustered by firm, are in parentheses. See Section 4.2.

Outcome:	*SGA*		*COGS*		*NetIncome*	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Rule significance						
$RegPipeline^{high}$	0.236***		0.355***		-0.452***	
	(0.058)		(0.124)		(0.065)	
$RegPipeline^{low}$		0.039		0.222***		-0.011
		(0.033)		(0.076)		(0.041)
Obs.	64,938	64,938	67,182	67,182	67,539	67,539
R^2	.923	.923	.916	.916	.61	.61
Panel B. Conversion probability						
$RegPipeline^{high}$	0.248***		0.696***		-0.216**	
	(0.076)		(0.181)		(0.086)	
$RegPipeline^{low}$		0.084		0.065		-0.127
		(0.091)		(0.214)		(0.109)
Obs.	64,938	64,938	67,182	67,182	67,539	67,539
R^2	.923	.923	.916	.916	.61	.61
Firm controls, FE	YES	YES	YES	YES	YES	YES
Time×FF48 FE	YES	YES	YES	YES	YES	YES
Regulation controls	YES	YES	YES	YES	YES	YES

Table 6: **Uncertainty**

Results from estimating Equation (4). $CAPX$ is capital expenditures scaled by beginning-of-period total assets. $RegPipeline$ is our primary measure of firm-level regulatory pipeline. $RegPipeline^{unique}$ ($RegPipeline^{standard}$) includes only rule proposals that are significantly different from (similar to) the agency's other rule proposals. Firm controls are $Assets$, $TobinQ$, CF, MTB, $Leverage$, and $Complexity$. Regulation controls are $RegIn$ from Kalmenovitz (2023) and $PRisk$ (from Hassan et al. (2019)). See variable definitions in Appendix A.3. Independent variables are divided by their standard deviation and lagged. Standard errors, clustered by firm, are in parentheses. See Section 4.3.

Outcome:					$CAPX$				
$RegPipeline$	-0.105***	-0.075***	-0.085**						
	(0.027)	(0.026)	(0.034)						
$RegPipeline^{unique}$				-0.307***	-0.290***	-0.288***			
				(0.039)	(0.037)	(0.049)			
$RegPipeline^{standard}$							0.001	0.021	0.008
							(0.020)	(0.020)	(0.025)
Obs.	105,496	105,485	63,047	105,496	105,485	63,047	105,496	105,485	63,047
R^2	.656	.704	.721	.657	.705	.721	.656	.704	.721
Firm controls, FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Time FE	YES	-	-	YES	-	-	YES	-	-
Time×FF48 FE	-	YES	YES	-	YES	YES	-	YES	YES
Regulation controls	-	-	YES	-	-	YES	-	-	YES

Table 7: **Economies of scale**

We compute the median firm size within the industry-quarter, based on total assets,and $Large = 1$ ($Large = 0$) if the firm is above (below) that median. Alternatively, we assign firms within the industry-quarter to quintiles, where Bin1 (Bin5) includes the smallest (largest) firms. Firm controls are $Assets$, $TobinQ$, CF, MTB, $Leverage$, and $Complexity$. Regulation controls are $RegIn$ from Kalmenovitz (2023) and $PRisk$ (from Hassan et al. (2019)). See variable definitions in Appendix A.3. Independent variables are divided by their standard deviation and lagged. Standard errors, clustered by firm, are in parentheses. See Section 4.4.

Outcome:	*SGA*				*COGS*			
$RegPipeline$	0.150***	0.112***	0.123***		0.247***	0.236***	0.338***	
	(0.032)	(0.032)	(0.039)		(0.067)	(0.067)	(0.085)	
$Large \times RegPipeline$	-0.089***	-0.088***	-0.074***		-0.118***	-0.122***	-0.102***	
	(0.007)	(0.007)	(0.007)		(0.014)	(0.013)	(0.014)	
$Bin1 \times RegPipeline$				0.209***				0.441***
				(0.039)				(0.085)
$Bin2 \times RegPipeline$				0.095**				0.287***
				(0.038)				(0.085)
$Bin3 \times RegPipeline$				0.015				0.184**
				(0.037)				(0.085)
$Bin4 \times RegPipeline$				-0.049				0.121
				(0.038)				(0.086)
$Bin5 \times RegPipeline$				-0.094**				0.058
				(0.038)				(0.087)
Obs.	104,369	104,358	64,938	64,938	123,795	123,784	67,182	67,182
R^2	.916	.92	.924	.927	.893	.901	.917	.918
Firm FE, controls	YES	YES	YES	YES	YES	YES	YES	YES
Time FE	YES	-	-	-	YES	-	-	-
Time×FF48 FE	-	YES	YES	YES	-	YES	YES	YES
Regulation controls	-	-	YES	YES	-	-	YES	YES

Table 8: **Political investment**

Results from estimating Equation (4). $\mathbb{1}(\mathit{Lobby}) = 1$ if the firm conducts lobbying activity in a given year. Conditional on $\mathbb{1}(\mathit{Lobby}) = 1$, $\mathit{Lobby}^{\$}$ is dollar spending on lobbying and $\mathit{Lobby}^{agencies}$ is the number of federal agencies the firm has been lobbying. $RegPipeline$ is our primary measure of firm-level regulatory pipeline. Firm controls are $Assets$, $TobinQ$, CF, MTB, $Leverage$, $Complexity$, $RegIn$ from Kalmenovitz (2023), and $PRisk$ (from Hassan et al. (2019)). See variable definitions in Appendix A.3. Independent variables are divided by their standard deviation and lagged. Standard errors, clustered by firm, are in parentheses. See Section 4.5.

Outcome:	$\mathbb{1}(\mathit{Lobby})$			$\mathit{Lobby}^{\$}$			$\mathit{Lobby}^{agencies}$		
$RegPipeline$	0.195	1.292	1.054	-101.074***	-83.924***	-97.562**	-0.282***	-0.343***	-0.306**
	(1.518)	(1.617)	(2.203)	(32.321)	(30.915)	(43.345)	(0.106)	(0.112)	(0.155)
Obs.	12,877	12,873	8,399	8,480	8,438	5,170	8,480	8,438	5,170
R^2	.606	.628	.635	.93	.935	.939	.782	.799	.815
Firm controls, FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Time FE	YES	-	-	YES	-	-	YES	-	-
Time×FF48 FE	-	YES	YES	-	YES	YES	-	YES	YES
Regulation controls	-	-	YES	-	-	YES	-	-	YES

Table 9: **Alternative explanations**

We consider alternative explanations to the baseline results from Table 4. $RegPipeline$ is our primary measure of potential regulations relevant to the firm. $Burden = 1$ if the firm's quarterly regulatory burden, $RegIn$, is above the median in the Fama-French industry. $RegPipeline^f rag$ is the fragmentation of the regulatory pipeline across agencies. Age is the firm's age and $Segments$ is the number of business segments. Firm controls are $Assets$, $TobinQ$, CF, MTB, $Leverage$, and $Complexity$. Regulation controls are $RegIn$ from Kalmenovitz (2023) and $PRisk$ (from Hassan et al. (2019)). See variable definitions in Appendix A.3. Independent variables are divided by their standard deviation and lagged. Standard errors, clustered by firm, are in parentheses. See Section 4.6.

Outcome:	SGA			$COGS$			$NetIncome$		
$RegPipeline$	0.102***	0.179***	0.128**	0.313***	0.275***	0.395***	-0.112**	-0.142***	-0.118**
	(0.039)	(0.044)	(0.053)	(0.085)	(0.094)	(0.102)	(0.047)	(0.054)	(0.059)
$Burden \times RegPipeline$	0.003			-0.004			-0.000		
	(0.003)			(0.007)			(0.003)		
$RegPipeline^{frag}$		0.127***			-0.062			-0.050	
		(0.039)			(0.081)			(0.045)	
Age			-60.918			-202.029**			33.046
			(45.019)			(87.514)			(41.956)
$Segments$			0.132*			0.169			0.074
			(0.075)			(0.201)			(0.069)
Obs.	64,938	64,938	42,304	67,182	67,182	43,694	67,539	67,539	43,882
R^2	.923	.923	.92	.916	.916	.924	.61	.61	.62
Time×FF48 FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Firm controls FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Regulation controls	YES	YES	YES	YES	YES	YES	YES	YES	YES

Table 10: **Systematically important regulations**

Results from estimating a factor model of stock returns with regulatory pipeline (Equation (7)). The model tests whether a given LDA regulatory topic is systematically important, meaning that it increases the firm's sensitivity to macroeconomic shocks. We compute the topic's incremental contribution to the model (ΔR^2) and its statistical significance, and report the 20 topics with the highest ΔR^2. See Section 5.

Topic	Label	$\Delta\ R^2$
Topic 93	Pharmaceuticals	0.40***
Topic 24	Health: insurance	0.16***
Topic 74	Government contracts: natural resources	0.11***
Topic 14	Securities: taxes and penalties	0.11***
Topic 97	Environmental protection	0.10***
Topic 59	Environmental protection: hazardous substances	0.10**
Topic 42	Product manufacturing	0.10
Topic 47	Aviation: aircraft safety	0.10
Topic 27	Rural credit	0.09
Topic 80	Infants and children: food assistance and nutrition	0.08
Topic 98	Housing programs	0.08
Topic 44	Income tax: penalties	0.07***
Topic 81	Income taxes: disability benefits	0.07
Topic 64	Grants: education	0.06**
Topic 60	Civil rights: individuals with disabilities	0.06
Topic 56	Income taxes: foreign trade	0.06
Topic 95	Intergovernmental relations	0.05
Topic 85	Agricultural subsidies	0.05
Topic 76	Government procurement: small business	0.05
Topic 50	Government employees: COI	0.05***

Internet Appendix

A.1 Rulemaking timeline

The first step in the paper is to identify rules that are inside the pipeline, that is, rules still under development. To determine which rules are currently in the index we apply the following methodology.

Entry - A rule enters the pipeline when it is first officially discussed by the Federal government. This happens when the RIN is mentioned for the first time in the Unified Agenda or in the Federal Register, the earliest. Note that the UA is published twice a year and provides updates retroactively (Nou and Stiglitz (2015)), a feature which we take into account. Consider the following stylized example. Suppose a rule was first announced in the FR on July 2015. The UA was published in April 2015 and October 2015, and the October edition of the UA would mention the FR publication from July. For our purposes, the rule's entry date would be July, not October. Continuing with this example, the latest edition of the UA (at the time of writing) was published in April 2023. It is possible that some rules were announced in the FR between April and October, and we will only learn about that when the October edition of the UA is released. To avoid this problem, we simply remove from the sample period the years after 2021.

Exit - A rule exits the pipeline based on the last activity reported in the Agenda. This is typically when the final draft is published or when the agency announces in the Agenda that the proposed rule has been rescinded. Consider the following stylized example, similar to the one mentioned above. Suppose the final version of the rule was published in the FR on July 2018. The UA was published in April 2018 and October 2018, and only the October edition of the UA would mention the FR publication from July. For our purposes, the rule's exit date would be July, not October.

A.2 Labeling LDA topics

In this section we explain how we label each LDA topic. We rely on the fact that some RINs are indexed by subjects. We obtain this information from the Federal Register, the official daily publication of the Federal Government. When a draft rule is published in the FR, it is typically (although not always) accompanied by a list of subjects that summarize the rule's content. We utilize this official subject list to label each LDA topic, in the following manner:

1. For each LDA topic, we identify the top 100 RINs that are most closely associated with the topic. The selection is based on the topic weights, that is, the fraction of words in the RIN dedicated to the topic.

2. We identify the most frequent subjects among the top 100 RINs for the topic. We exclude the generic subjects "reporting and recordkeeping requirements" and "administrative practice and procedure," which are listed in a majority of rules and do not describe meaningful differences across them.

3. Focusing on the top five subjects per LDA topic, we distinguish between three cases:

   (a) For 69 of the topics, the top five subjects are unique and clearly highlight the appropriate topic label.

   (b) For 20 topics, the top five subjects are unique but some further clarification is required. We obtain this clarity by manually reading the top five RINs for the topic and use their content to help determine the topic label.

   (c) For the last 11 topics, we look at the top keywords for the topic, in addition to the top subjects and manual reading of the top RINs.

Table A.2 reports the output of the procedure for selected LDA topics.

A.3 Variable definitions

In this section we explain how several variables were constructed.

Pipeline fragmentation **-** the fragmentation of the regulatory pipeline, defined as:

$$RegPipeline_{i,t}^{frag} = \sum_{o=1}^{O} \omega_{i,o,t} \cdot TopicPipeline_{o,t}^{frag}, \tag{A.1}$$

where $\omega_{i,o,t}$ are the same firm×topic weights from Equation (3). We interact those weights with $TopicPipeline_{o,t}^{frag}$, which represents the fragmentation of topic o rather than its quantity. To compute the fragmentation of the topic, we first identify the agency responsible for each rule.[1] We then compute the fraction of topic o captured by agency a at time t, denoted as $Weight_{o,a,t}$. Finally, we compute the fragmentation of topic o at time t as the inverse of the HHI score:

$$TopicPipeline_{o,t}^{frag} = 1 - HHI_{o,t} = 1 - \sum_{a=1}^{A} (Weight_{a,o,t})^2 \tag{A.2}$$

Note that the fragmentation of the pipeline, on aggregate and at the firm level, is bound between 0 and 1 (similar to a standard HHI score). Higher values imply that topic o is fragmented, meaning that it is handled by a large number of federal agencies.

Complexity **-** the complexity of the firm's operations, defined as:

$$Complexity_{i,t} = 1 - HHI_{i,t} = 1 - \sum_{o=1}^{100} (\omega_{i,o,t})^2, \tag{A.3}$$

where $\omega_{i,o,t}$ is the weight of topic o in firm i's conference call during time t (Equation (3)). Essentially, we compute the HHI score for the firm, and complexity is one minus the HHI. Intuitively, if the firm discusses multiple topics in equal amounts (low HHI), it means that the firm operates in many markets and is thus more complex. Conversely, if the firm discusses

[1] Throughout the paper, we follow the Unified Agenda to define distinct agencies. For instance, the Department of Treasury has several sub-agencies including the Internal Revenue Service and the Office of the Comptroller of the Currency.

predominantly a single topic (high HHI), it means that the firm operates in fewer areas and is thus less complex.

Firm controls - Our baseline specification include the following variables from Compustat. *Assets* is assets expressed in inflation-adjusted USD. MTB is market-to-book ratio: market value of equity divided by the sum of book value of equity, deferred taxes, and preferred stock. $TobinQ$ is the market value of equity plus the book value of assets minus book value of equity plus deferred taxes, scaled by beginning-of-period total assets. CF is operating cash flow scaled by beginning-of-period total assets. *Leverage* is long-term debt (dltt) plus short-term debt (dlc), scaled by beginning-of-period total assets.

***PRisk* -** the firm's political risk from Hassan et al. (2019).

***RegIn* -** the burden of effective regulations from Kalmenovitz (2023). We use the version with the number of regulations but obtain similar results when using *RegIn* based on the costs of compliance in terms of time, forms, and dollars.

***CFR* -** exposure to the Code of Federal Regulations, defined as:

$$CFR_{i,t} = \sum_{j \in J_t} \omega_{i,j,t} \cdot CFR_{j,t}, \tag{A.4}$$

where J_t is the set of 2-digit NAICS industries in which firm i operates during time t, and $\omega_{i,j,t}$ is the share of firm i's revenues derived from industry j at time t (based on Compustat's business segment data). The left-hand side term, $CFR_{j,t}$, is based on industry-level data from Al-Ubaydli and McLaughlin (2017). For each CFR part, the authors report the probability that it applies to industry j and the number of words in that part (2-digit NAICS codes). By multiplying the probability with the number of words and aggregating across all CFR parts, we obtain the CFR burden on industry j at time t ($CFR_{j,t}$). Our final firm-level measure takes into account the firm's activities across different business segments, and the CFR exposure in each of those segments.

A.4 Anticipating rule conversion

In Section 4.2, we present a test based on ex-ante probability of the rule's success. To that end, we estimate a linear probability model in the cross-section of rule proposals:

$$\mathbb{1}(Rule)_r = Length_r + \lambda_{a,t,i},$$

where the outcome equals one for proposals that yielded a final rule, and zero otherwise. We control for the length of the rule's language as a possible predictor. We then add multiplicative fixed effects for agency, cohort (year in which the rule entered the pipeline), and tier (whether the rule is significant or not). In other words, we compare rule proposals which were initiated by the same agency at the same time and with similar significance. We define the predicted values from Equation (5) as the ex-ante probability that the proposal will successfully navigate the pipeline and become an effective rule.

Figure A.1, Panel A, presents the distribution of predicted values over time. We select the rule proposals which were announced in a given month, and compute the average ex-ante probability. There is no clear time trend, which is expected given that we include cohort fixed effects in the predictive model. In Panel B, we report the weighted average of the ex-ante probability, by topic. The weights represent the fraction of rule r associated with topic o.[2] The average (average) probability is 0.57, with values ranging from 0.34 to 0.78. The highest values are for fishing-related topics such as "Fisheries and fishing: Atlantic" (topic 87) and "Fisheries and fishing: treaties" (topic 45), while the lowest values are for "Drug labeling" (topic 8) and "Telecommunication and aviation" (topic 34).

[2] Suppose topic o captures 40% of rule 1 and 20% of rule 2. The weights would then be 66.6% and 33.3% for rule 1 and rule 2, respectively.

A.5 Derivations

This section details the derivations in Section 5. We follow closely the methodology in Chen, Roussanov, and Wang (2021).

Let $X_t \equiv (x_{i,t}, ...x_{N_t,t})'$ be the collection of exposure of firm i to each regulatory topic across all firms within a given time period. The vector is of size $N_t \times K$. Equation (7) in matrix form is then:

$$R_{t+1} = X_t\Gamma_\alpha + X_t\Gamma_b f_{t+1} + \epsilon_{t+1}. \tag{A.5}$$

To solve for Γ, consider projecting R_{t+1} on X_t as in ordinary least square by premultiplying it by $(X_t'X_t)^{-1}X_t'$. We then solve:

$$\begin{aligned} \hat{R}_{t+1} &\equiv (X_t'X_t)^{-1}X_t'R_{t+1} \\ &= (X_t'X_t)^{-1}X_t'X_t\Gamma_\alpha + (X_t'X_t)^{-1}X_t'X_t\Gamma_b f_{t+1} + (X_t'X_t)^{-1}X_t'\epsilon_{t+1} \\ &= \Gamma_\alpha + \Gamma_b f_{t+1} + \hat{\epsilon}_{t+1}, \end{aligned} \tag{A.6}$$

where $\hat{R}_{t+1}$ contains managed portfolios, corresponding to regulation-exposure-mimicking portfolios. Each portfolio corresponds to the coefficient to a specific topic in a Fama and MacBeth (1973) cross-sectional regression of returns in topics. In practice, each of the managed portfolios corresponds to a portfolio that has non-zero exposure to only one regulatory pipeline topic. The above equation means that because each of the managed portfolios is only exposed to one regulatory pipeline topic, the betas of those portfolios with respect to the latent systematic factors are contained in Γ_b. Furthermore, subtracting the time-series mean of each portfolio

$$\bar{\hat{R}}_{t+1} = \frac{1}{T}\sum_{t=1}^{T}\hat{R}_{t+1} = \tag{A.7}$$

$$= \frac{1}{T}\sum_{t=1}^{T}(\Gamma_{\alpha} + \Gamma_{b}f_{t+1} + \hat{\epsilon}_{t+1}) = \tag{A.8}$$

$$= \Gamma_{\alpha} + \Gamma_{b}\frac{1}{T}\sum_{t=1}^{T}f_{t+1} + \frac{1}{T}\sum_{t=1}^{T}\hat{\epsilon}_{t+1} = \tag{A.9}$$

$$= \Gamma_{\alpha} + \Gamma_{b}\bar{f}_{t+1} + 0, \tag{A.10}$$

since by construction $\frac{1}{T}\sum_{t=1}^{T}\hat{\epsilon}_{t+1} = 0$, we can express the former as:

$$\begin{aligned}\hat{R}_{t+1} - \bar{\hat{R}}_{t+1} &= \Gamma_{\alpha} + \Gamma_{b}f_{t+1} + \hat{\epsilon}_{t+1} - (\Gamma_{\alpha} + \Gamma_{b}\bar{f}_{t+1}) = \\ &= \Gamma_{b}(f_{t+1} - \bar{f}_{t+1}) + \hat{\epsilon}_{t+1}.\end{aligned} \tag{A.11}$$

Once we have Equation (A.11), the problem does not depend on the regulatory exposures, and we have a standard latent factor model representation. The standard method to solve for these latent factors and exposures is to use a principal component decomposition of the covariance of the managed portfolios since this method solves the minimum reconstruction error problem:

$$\Gamma_{b}, f_{t+1} - \bar{f}_{t+1} = argmin_{B,F}\left\|\hat{R}_{t+1} - \bar{\hat{R}}_{t+1} - BF\right\|_{2}^{2} \tag{A.12}$$

Moreover, the latent factors can be recovered by projecting the $f_{t+1} = \Gamma_{b}\hat{R}_{t+1}$. We want to find an orthogonal set of L linear basis vectors $\gamma_{b} \in \mathbb{R}^{K}$, and the corresponding factor realizations $f_{t+1} \in \mathbb{R}^{L}$, such that we minimize the average reconstruction error

$$J(\Gamma_{b}, \mathbf{Z}) = \frac{1}{T}\sum_{t=1}^{T}||\tilde{R}_{t+1} - \Gamma_{b}f_{t+1}||^{2}, \tag{A.13}$$

subject to the constraint that Γ_{b} is orthonormal. Orthonormality is imposed to obtain uniqueness (up to sign).

The optimal solution (Eckart–Young–Mirsky theorem) is obtained by setting $\hat{\Gamma_b} = \mathbf{V}_L$, where $\mathbf{V}_L$ contains the L eigenvectors with largest eigenvalues of the empirical covariance matrix, $\hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T} \tilde{R}_{t+1}\tilde{R}_{t+1}^{T}$. Furthermore, the latent factors can be recovered by projecting the returns on Γ_b, $f_{t+1} = (\Gamma_b'\Gamma_b)^{-1}\Gamma_b\hat{R}_{t+1} = \Gamma_b\hat{R}_{t+1}$ (since Γ_b is orthonormal). That is, we recover Γ_b by performing an eigenvalue decomposition of the covariance matrix of the managed portfolios.

Kelly, Pruitt, and Su (2019) develop a bootstrapping procedure to estimate the distribution of Γ_b. The procedure is designed to evaluate the significance of a specific topic exposure. It involves several steps:

1. **Residuals Computation**: Compute the residuals $\hat{\epsilon}$ as the difference between the actual returns and the returns predicted by the factor model .

2. **Initial Wald Statistic**: Calculate a Wald statistic for the test topic, $\gamma_k'\gamma_k$ where γ_k is the k-th row of Γ_b.

3. **Null Hypothesis**: Under the null hypothesis, which assumes that the test topic exposure has no significant effect, modify the Γ_b matrix to Γ_{H_0} by setting $\gamma_k = 0$.

4. **Model Fitting under Null Hypothesis**: Fit the model under the null hypothesis using the returns matrix and the modified Γ_b matrix (Γ_{H_0}).

5. **Bootstrap Analysis**:

   - Conduct a series of bootstrap iterations, typically 1000, denoted as n_{boot}.
   - In each iteration, generate a vector of Student t-distributed variable q_t with 5 degrees of freedom to ensure unit variance.
   - Create bootstrapped residuals by resampling from the original residuals and scaling them with q_t.

- Calculate the bootstrapped returns by adding these bootstrapped residuals to the model fit under the null hypothesis.
- Compute a new Γ_b matrix ($\tilde{\Gamma}_b$) for each set of bootstrapped returns.
- Recalculate the Wald statistic for the test topic exposure using $\tilde{\gamma}_k{}'\tilde{\gamma}_k$ in each bootstrap iteration and store it.

6. **Results Compilation**: Compile all the Wald statistics from each bootstrap iteration alongside the original Wald statistic. This collection of statistics is used to assess the robustness and significance of the test topic exposure under different sampling conditions.

7. **Statistical Test**: Calculate the proportion of the Wald statistics in the bootstrap samples $\tilde{\gamma}_k{}'\tilde{\gamma}_k$ that are greater than the original Wald statistic ($\gamma_k'\gamma_k$). This proportion serves as an empirical p-value, indicating the likelihood of observing a Wald statistic as extreme as the original one under the null hypothesis (i.e., the topic exposure is not significant).

Figure A.1: **From rule proposal to effective rule**

We compute the ex-ante probability that the rule proposal will clear the pipeline and become an effective rule. In Panel A, we plot the average probability by month of origination (the month in which the rule proposal was released) against the realized conversion rate. In Panel B, we plot the weighted average of the probability by topic. See Appendix A.4.

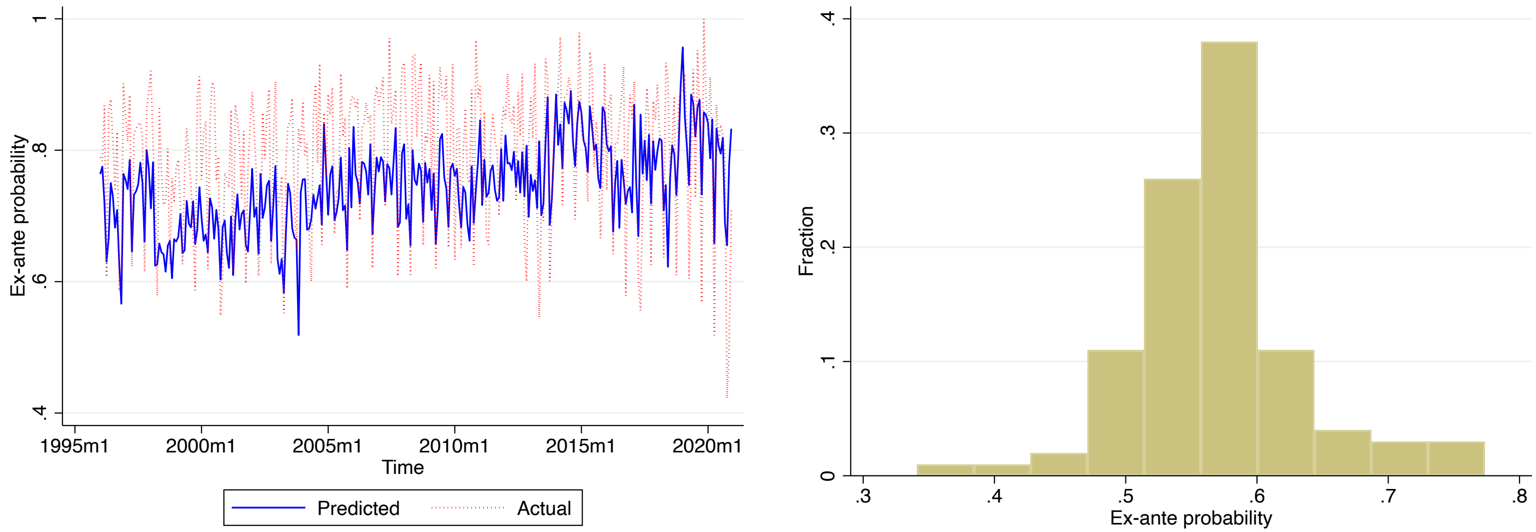
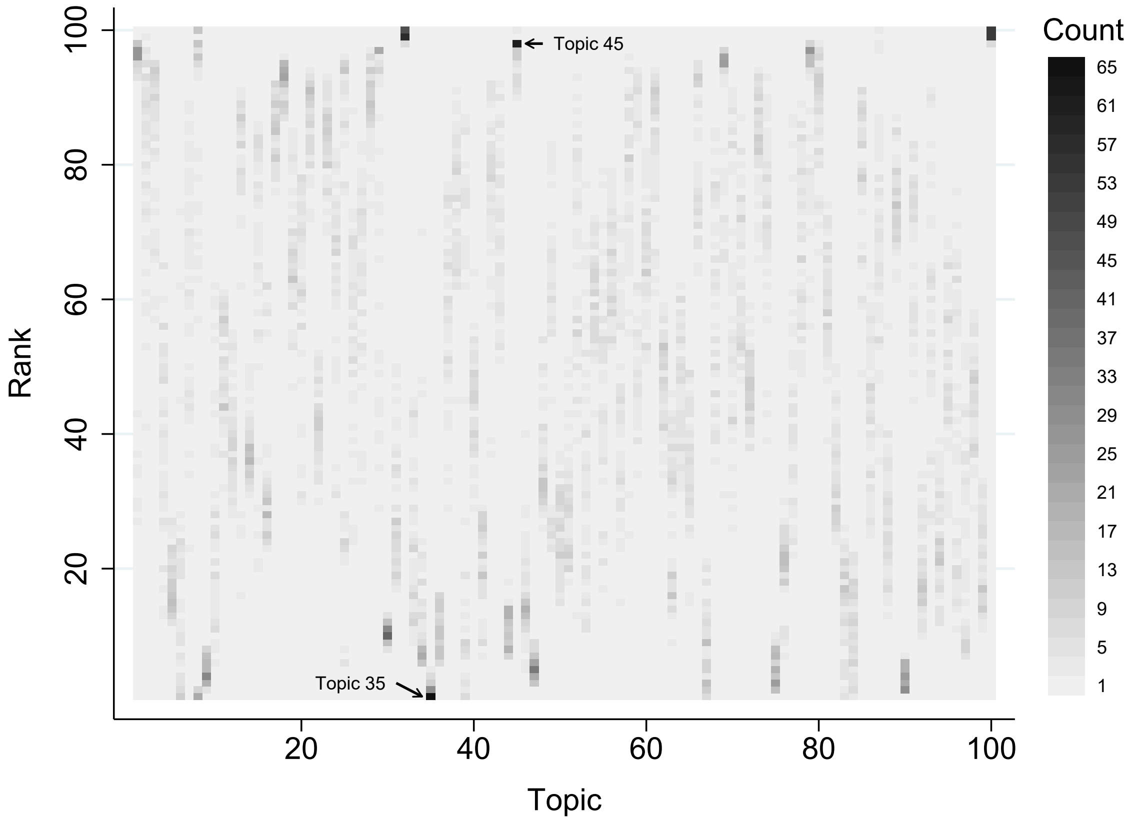
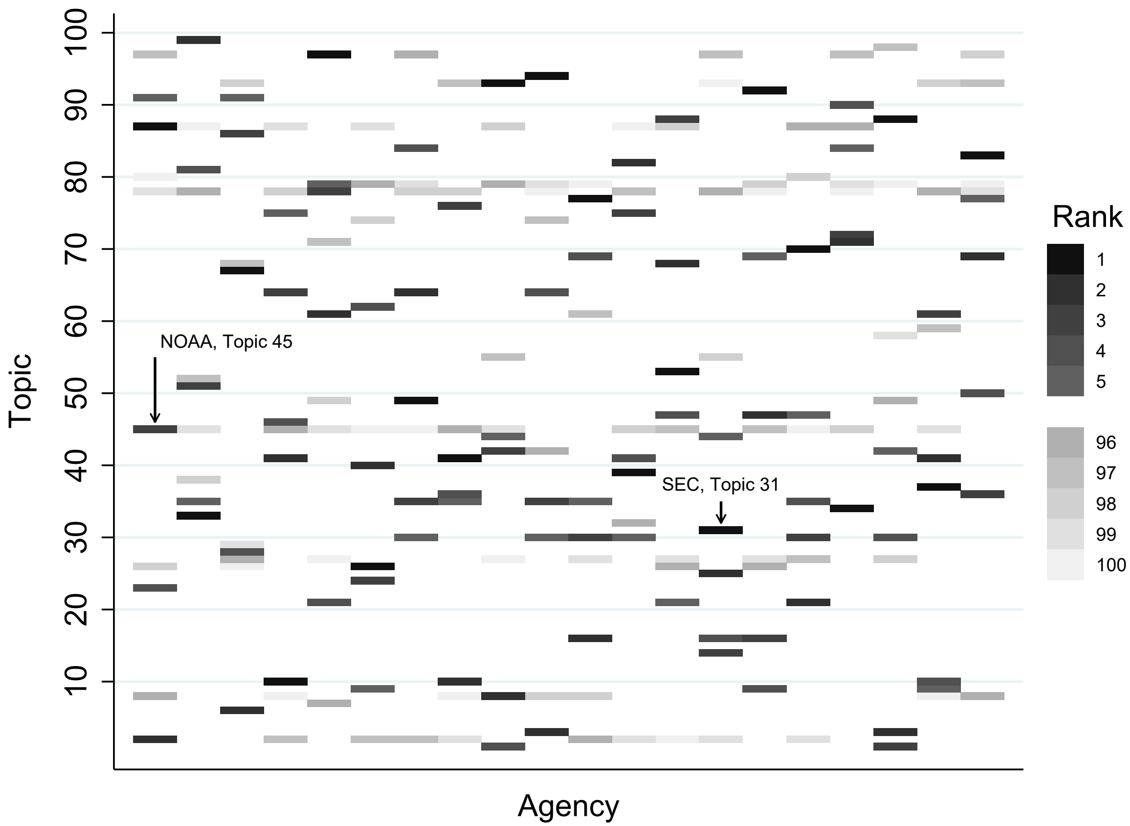


Figure A.2: **Topic distribution**

Panel A. This figure is identical to Figure 3, Panel A, except that we use black and white color palate.



Panel B. This figure is identical to Figure 3, Panel B, except that we use black and white color palate.



Panel C. This figure is identical to Figure 3, Panel C, except that we use black and white color palate.

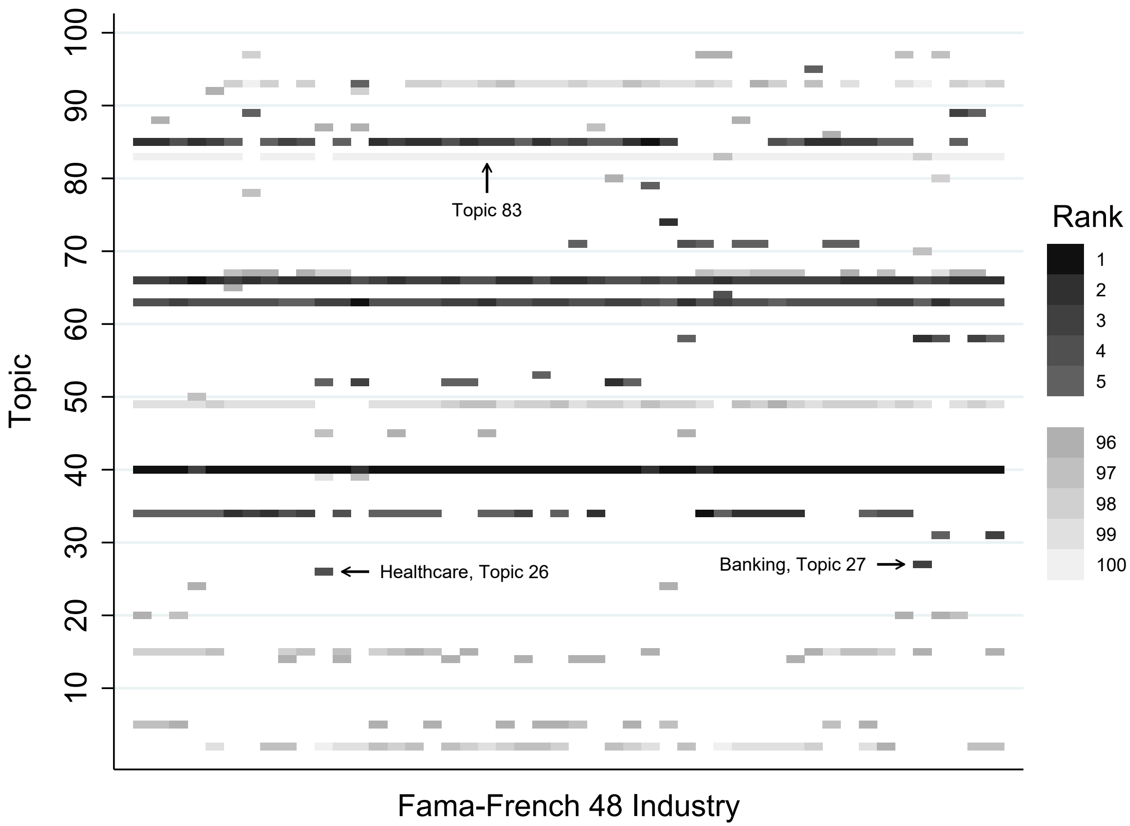


Figure A.3: **Commonalities in regulatory pipeline**

This figure is identical to Figure 4, except that we use black and white color palate.

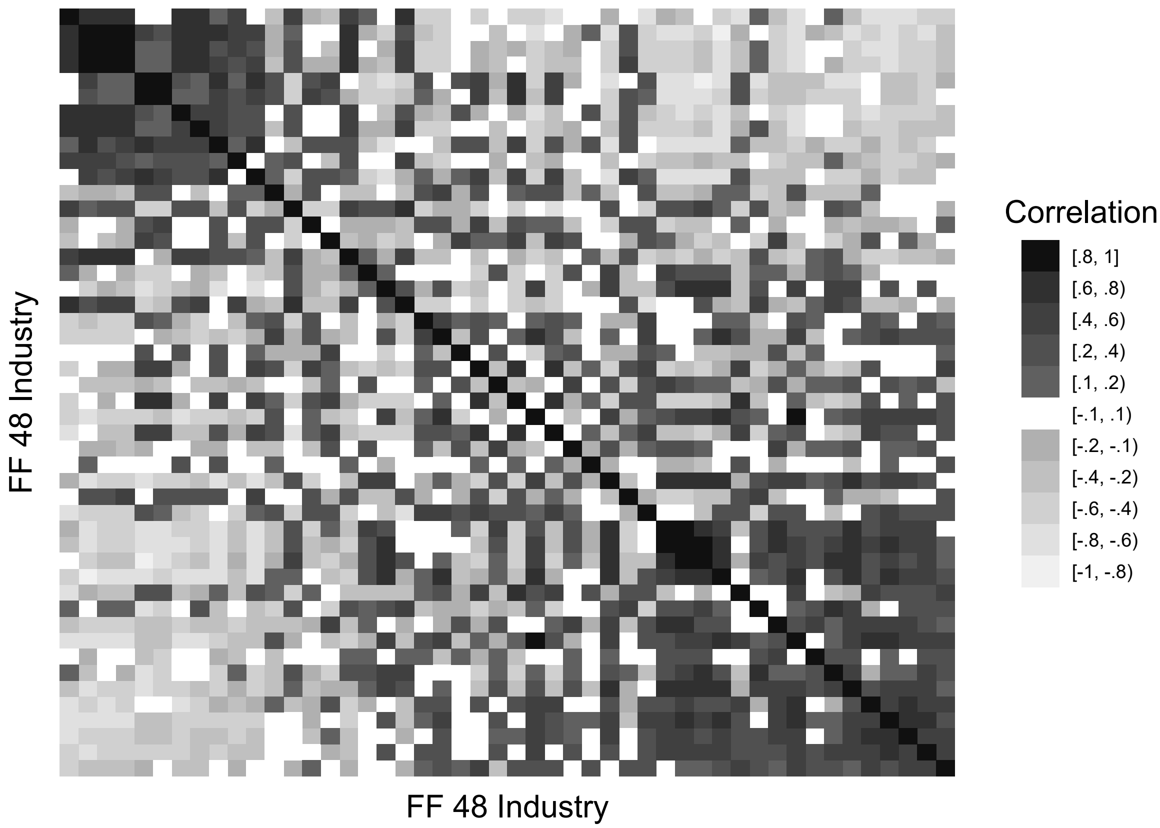
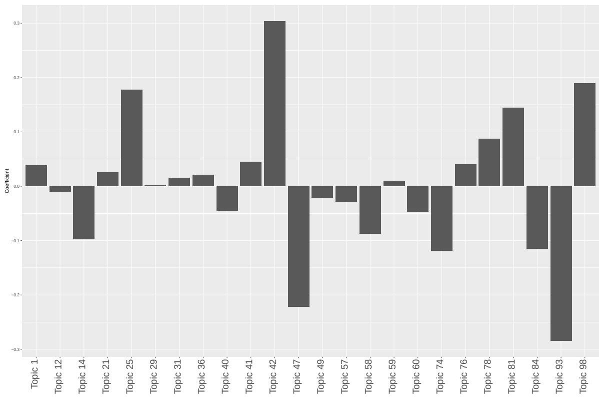
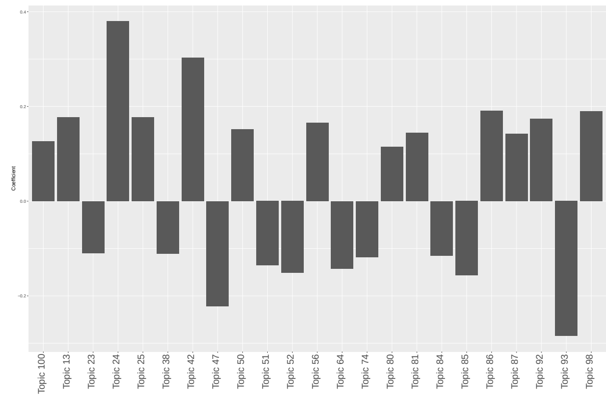


Figure A.4: Exposure of Regulatory Topics to Latent Factors

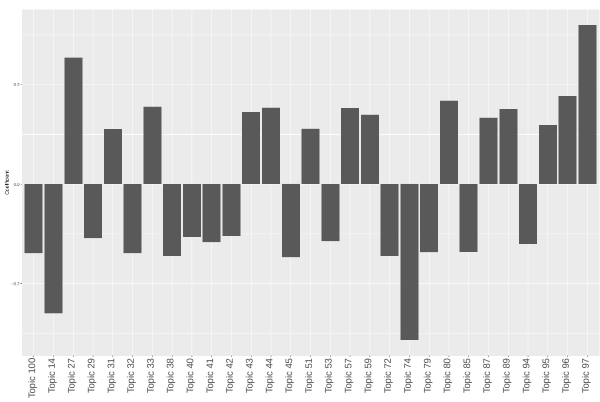
This figure shows in the first five panels the individual coefficients of the exposures to each regulatory topic in the factor-loading matrix Γ_b. The last panel shows the individual coefficients of the intercept vector Γ_a (impact on returns not through the systematic factors).



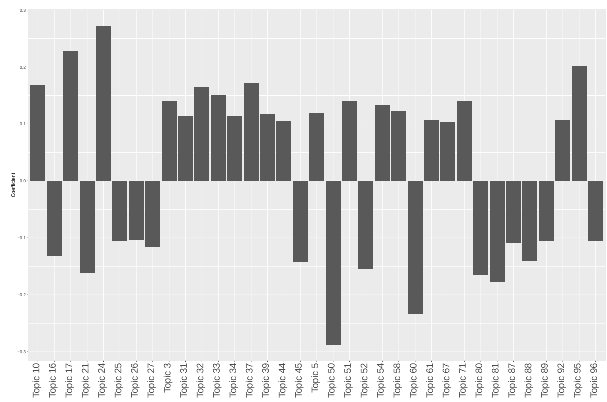
(A) Factor 1



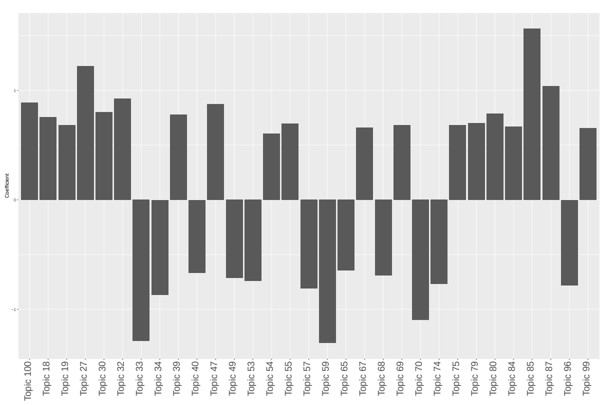
(B) Factor 2



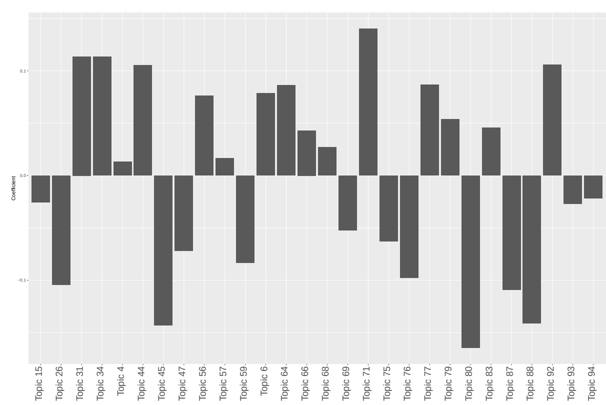
(C) Factor 3



(D) Factor 4



(E) Factor 5



(F) Expected Returns Orthogonal to Factors

Table A.1: **Summary statistics: additional**

Top 20 agencies, by number of rules under development.

Rank	**Agency**	**Rules**	**(%)**
1	National Oceanic and Atmospheric Administration	3,003	7.0
2	Internal Revenue Service	2,377	5.5
3	U.S. Fish and Wildlife Service	2,067	4.8
4	Office of Air & Radiation	1,643	3.8
5	General Services Administration	1,403	3.3
6	Centers for Medicare and Medicaid Services	1,332	3.1
7	Defense Acquisition Regulatory Council	1,154	2.7
8	Department of Veterans Affairs	1,122	2.6
9	Food and Drug Administration	838	1.9
10	Bureau of Industry and Security	797	1.9
11	Office of Personnel Management	787	1.8
12	National Highway Traffic Safety Administration	756	1.8
13	Securities and Exchange Commission	743	1.7
14	Federal Aviation Administration	714	1.6
15	Nuclear Regulatory Commission	629	1.5
16	Federal Communications Commission	588	1.4
17	Alcohol and Tobacco Tax and Trade Bureau	522	1.2
18	U.S. Coast Guard	496	1.1
19	Department of State	495	1.1
20	Department of Defense	490	1.1

Table A.2: **Labeling of LDA topics**

We label each LDA topic using the procedure described in Appendix A.2. For selected topics we report the topic's number (which is arbitrarily determined but used consistently throughout the paper), the label we assign, the most common CFR subjects, and the most common keywords.

Label	**Subjects**	**Keywords**
(1) Food labeling	labeling (40), imports (28), packaging and containers (25), food labeling (22), advertising (21)	label, statement, content, dietary, container
(2) Fisheries and fishing: Alaska and Pacific	fisheries (99), alaska (79), fishing (22), american samoa (11), northern mariana islands (10)	halibut, island, bering, alaska, crab
(3) Government employee wages	wages (72), government employees (72), freedom of information (68), wine (27), law enforcement officers (4)	wage, area, boundary, prevail, mile
(14) Securities: taxes and penalties	income taxes (44), securities (21), penalties (18), excise taxes (16), estate taxes (13)	amendment, exempt, exemption, settlement, transaction
(24) Health: insurance	health care (65), health insurance (64), disclosure (44), excise taxes (44), medical child support (44)	mine, msha, health, miner, patient
(26) Health: medicare	health facilities (96), medicare (95), health professions (62), puerto rico (56), kidney diseases (51)	medicare, payment, medicaid, nurse, prospective
(27) Rural credit	mortgages (29), rural areas (28), credit (23), agriculture (15), housing (15)	loan, borrower, guarantee, lender, rural
(31) Securities: investment companies	securities (90), investment companies (50), brokers (16), accounting (14), investment advisers (11)	investment, division, company, offer, adviser
(34) Telecommunication and aviation	radio (16), telecommunications (13), aircraft (13), aviation safety (13), incorporation by reference (13)	band, spectrum, satellite, broadband, communication

Label	Subjects	Keywords
(35) Government operations: legal	inventions and patents (44), courts (34), small businesses (33), freedom of information (31), lawyers (24)	change, clarify, make, definition, conform
(40) Banking: credit and income taxes	banking (21), banks (21), savings associations (19), credit (16), income taxes (16)	value, amount, rate, index, calculate
(45) Fisheries and fishing: treaties	fisheries (96), fishing (93), treaties (81), fish (56), marine resources (50)	tuna, highly, ocean, central, turtle
(59) Environmental protection: hazardous substances	environmental protection (68), hazardous substances (43), chemicals (42), community right-to-know (27), superfund (16)	chemical, agent, toxic, substance, inventory
(74) Government contracts: natural resources	government contracts (43), incorporation by reference (42), oil and gas exploration (42), continental shelf (41), public lands - mineral resources (39)	gas, oil, pipeline, petroleum, outer
(83) Immigration	aliens (95), immigration (85), passports and visas (41), employment (31), organization and functions (government agencies) (23)	immigration, alien, dhs, nonimmigrant, immigrant
(93) Pharmaceuticals	drugs (47), labeling (37), confidential business information (27), medical devices (26), imports (20)	drug, prescription, fda, cosmetic, device
(97) Environmental protection	environmental protection (100), air pollution control (100), hazardous substances (98), intergovernmental relations (75), incorporation by reference (54)	neshap, source, emission, mact, pollutant

Table A.3: **Conference calls and dominant topics**

We identify conference calls with abnormal weight on a single topic. For selected calls we report the dominant topic's number and label, the call's identifying details, and a representative excerpt from the call.

Topic	**Call**	**Transcript**
(40) Banking: credit and income taxes	GSI Technology (7/31/2014)	At June 30, 2014, we had: $84 million in cash, cash equivalents and short-term investments; $23 million in long-term investments; $95.5 million in working capital; no debt; and stockholders' equity of $126.9 million. Accounts payable at June 30, 2014, was $3.6 million, compared to $4.9 million at March 31, 2014. Net inventory was $8.4 million at June 30, 2014, up slightly from $8.2 million at March 31, 2014. Inventory turns at June 30, 2014, were 3.3x, compared to 3.4x at March 31, 2014. Looking forward to the second quarter, we currently expect net revenues to be in the range of $12.5 million to $13.5 million, with gross margin of approximately 43% to 45%.
(34) Telecommunication and aviation	Minim, formerly Zoom Telephonics (2/24/2011)	Looking forward, we expect to dramatically expand retail shelf space for our mobile broadband wireless routers during the second quarter of this year. We are adding exciting new capabilities to these routers including the ability to work with some tethered phones including the iPhone. This means that you can plug the tethered phone into the router and then share the phone's high-speed Internet access among all your WiFi compatible devices including computers, mobile phones, iPads and other tablets and e-readers. We also plan to introduce the DOCSIS 3.0 cable modem with wireless-N capability. This assumes we will resolve our FCC complaint against Comcast, which is likely. Zoom has successfully transitioned its DSL modem line so that all our popular DSL modems and routers use Broadcom-integrated circuits.

Topic	Call	Transcript
(74) Government contracts: natural resources	Warren Resources (8/7/2013)	We anticipate drilling 19 producing wells targeting the Monterey oil formation and 2 disposal wells. Our estimated net acquisition drilling and development cost for the project are approximately $16 million. Warren expects to commence drilling operations in October of '13, and production operations in January of '14. Warren will be the operator of the Leroy Pine project. The project's a reactivation of an older oil field with substantial well control which substantially reduces our exploration risk. In addition, according to the US Energy Department estimates, the Monterey Shale formation in California accounts for approximately 2/3 of the oil shale reserves in the United States. The Leroy Pine project allows Warren to leverage its operational and regulatory skill sets in the development of California oil projects, which was evidenced by our much larger Wilmington oil field development. Warren's drill and learned approach in Leroy Pine will be a stepping stone to the company's development of additional opportunities in the Monterey oil formation and other opportunities within California.