Compliance Spillovers and the Limits of Free-Entry Neutrality

Please click here for the latest version

Vishavdeep Sharma*

Abstract

When large firms face cost shocks, does free entry preserve consumer surplus? I study a free-entry oligopoly with large incumbents and a fringe of smaller firms, in which compliance costs can depend on rivals' compliance. An incumbent cost shock changes long-run output if and only if a new entry–compliance wedge is nonzero and compliance spills over across firms. The wedge measures how industry compliance changes when fringe output replaces incumbent output. Entry also creates a compliance externality that can reinforce or reverse the classic excess-entry result. U.S. environmental data show that compliance obligations grow far less than proportionally with firm size. In a calibration, a 10 percent increase in incumbents' marginal costs raises the fringe mass by 43 percent and industry compliance by 35 percent, lowering output under a rivalrous spillover.

Keywords: Free entry; regulatory compliance; compliance spillovers; environmental regulation; market structure; oligopoly.

JEL Classification: L11, L13, L51, Q58, D62.

*Department of Agricultural, Environmental, and Development Economics (AEDE), The Ohio State University, Columbus, OH 43210, United States. Email: sharma.1094@osu.edu

1 Introduction

When a large firm faces a cost shock, or two large firms merge, these incumbents produce less. With the set of firms held fixed, market output falls, the market price rises, and consumer surplus falls. A benchmark result in industrial organization shows that these effects need not persist: under free entry, smaller firms, which I call fringe firms, enter until the reduction in incumbents' output is fully offset, leaving long-run market output, the market price, and consumer surplus unchanged (Anderson et al., 2020). I refer to this result as *free-entry neutrality*. Merger enforcement relies on the same logic, as antitrust agencies assess whether entry or expansion by rivals would be sufficient to counteract the competitive effects of a merger (U.S. Department of Justice and Federal Trade Commission, 2023).[1] When fringe firms offset the reduction in incumbents' output, does the long-run equilibrium necessarily remain unchanged?

One reason the long-run equilibrium may change is regulatory compliance. Firms in regulated industries must meet obligations such as permits, monitoring, testing, and reporting, and these obligations need not increase proportionally with output. A large firm spreads its obligations over a large output, while several smaller firms producing the same total output each bear their own. Replacing incumbents' output with fringe output can therefore increase the total compliance required to supply a given market output. Evidence on compliance costs points the same way: these costs weigh relatively more on smaller firms (Coates and Srinivasan, 2014; Trebbi et al., 2023).

Yet a greater compliance burden on smaller firms is not, by itself, enough to change the long-run equilibrium. If each firm's compliance cost depends only on its own obligations, long-run market output is still unchanged. What matters is whether the additional compliance changes other firms' compliance costs. When firms compete for scarce compliance inputs, such as auditors or accredited testing laboratories, more compliance by one firm raises other firms' compliance costs (Duguay et al., 2020; U.S. Government Accountability Office, 2025). When firms share compliance resources, as in industry compliance groups, more compliance by one firm lowers other firms' compliance costs (California State Water Resources Control Board, 2014).

[1] Courts apply this test, and they do not always accept that entry will replace the lost competition. When JetBlue sought to acquire Spirit Airlines, the two airlines argued that other low-cost carriers would enter the routes Spirit served; the court agreed that entry was likely but found that it might not be large enough to replace Spirit (*United States v. JetBlue Airways Corp.*, D. Mass. 2024). When Kroger sought to merge with Albertsons, the two grocers argued that rival chains would expand; the court found that opening a new supermarket can take years and millions of dollars, and that the evidence did not show that expansion would replace the lost competition in time (*Federal Trade Commission v. Kroger Co.*, D. Or. 2024). In both cases, the question was whether entry would happen on a large enough scale. This paper asks a different question: whether entry restores market outcomes even when it does happen.

This paper asks when free-entry neutrality survives regulatory compliance. I show that the answer depends on two conditions: replacing incumbent output with fringe output must change industry compliance, the total compliance undertaken in the industry, and the cost of meeting a firm's obligations must depend on the compliance of other firms. When both conditions hold, entry changes every firm's compliance costs and therefore the long-run equilibrium.

The paper makes three contributions. First, I characterize when free-entry neutrality survives regulatory compliance. Building on Anderson et al. (2020), I show that, under regularity conditions, an incumbent cost shock changes long-run output if and only if two objects are both nonzero: the compliance spillover and a new object, the *entry–compliance wedge*, which compares a fringe firm's average compliance per unit of output with an incumbent's marginal compliance. With a positive wedge, long-run output falls under a rivalrous spillover and rises under a cooperative one, and mergers among incumbents have analogous effects.

Second, entry creates a compliance externality alongside the business-stealing distortion of Mankiw and Whinston (1986), which can reinforce or reverse the classic excess-entry result.

Third, I quantify the mechanism. I calibrate the model to a fact that I document by linking records of the U.S. Environmental Protection Agency (EPA) to Compustat: compliance obligations grow far less than proportionally with firm size, with an elasticity with respect to sales of about 0.1 to 0.15 (Figures 1 and 2). A 10 percent increase in incumbents' marginal costs raises the fringe mass by about 43 percent and industry compliance by about 35 percent, and long-run output falls by about 0.2 percent at the baseline and by up to about 1.2 percent when the supply of compliance inputs is inelastic (Figure 5).

Related literature. The paper contributes to three literatures. The first studies free entry and market structure. Seade (1980) examines how entry affects Cournot equilibrium, and Mankiw and Whinston (1986) identify business stealing as a source of excessive entry. Amir et al. (2014) generalize the comparison between free and socially optimal entry under business-stealing and business-enhancing competition. Most closely related, Anderson et al. (2020) show that, in aggregative oligopoly games with free entry, shocks to incumbents and mergers among them leave long-run consumer surplus unchanged. Related work applies these methods to platform markets (Anderson and Peitz, 2020; Peitz and Sato, 2026), where an additional aggregate determines whether neutrality holds. I study a different second aggregate: compliance obligations that do not scale proportionally with output make industry compliance a determinant of firms' costs. The entry–compliance wedge

characterizes when the adjustment of the fringe ceases to neutralize shocks to incumbents. The mechanism also relates to models of research and development spillovers with free entry (Erkal and Piccinin, 2010; Khazabi and Quyen, 2012). Unlike research and development, however, compliance is imposed by regulation rather than chosen as a cost-reducing investment, and its spillover can be either cooperative or rivalrous. The welfare result also relates to Basak and Petrakis (2021), where a network externality can reverse the excess-entry result.

The second literature examines the relationship between regulation and market structure. Regulation may serve incumbent interests (Stigler, 1971; Laffont and Tirole, 1991) and can raise rivals' costs (Salop and Scheffman, 1983). Environmental and workplace regulations can impose relatively greater burdens on smaller plants (Pashigian, 1984) and generate competitive advantages for some firms (Bartel and Thomas, 1987). Environmental regulation can raise entry costs and reduce competition (Ryan, 2012) and affect industrial concentration (Liu et al., 2026). With subadditive costs, a regulated price floor can determine which market structures are sustainable (Sharma, 2026). Size-dependent regulatory thresholds can distort the firm-size distribution (Garicano et al., 2016), while regulatory costs are associated with market power and absorb substantial firm resources (Singla, 2023; Trebbi et al., 2023). I show that the scaling of compliance obligations also determines how an industry adjusts to shocks to its incumbent firms. Even when regulation does not prevent entry, the shift of production toward fringe firms can change industry compliance and, through spillovers, long-run market outcomes.

The third literature studies environmental compliance and enforcement. Building on the firm-level approach to enforcement of Becker (1968), existing work examines environmental inspections in oligopolistic markets (Häckner and Herzing, 2017), equilibrium compliance decisions (Häckner and Herzing, 2020), and dynamic enforcement policies (Blundell et al., 2020). A growing empirical literature documents spillovers in compliance behavior: enforcement against one facility can affect the pollution or compliance decisions of other facilities and firms (Shimshack and Ward, 2005; Gray and Shimshack, 2011; Evans et al., 2018; Dasgupta et al., 2023; Li and Lyon, 2026), and firms' compliance also responds to that of their peers (Liu et al., 2018). Collective reputations provide one explanation for such interdependence (Tirole, 1996). I study a different margin: how one firm's compliance affects the *cost* of compliance for others. Evidence from auditing shows that regulation-induced demand for shared compliance inputs can generate such cost spillovers (Duguay et al., 2020). The mechanism relates to cost spillovers in research and development (d'Aspremont and Jacquemin, 1988), but here the mechanism operates through compliance obligations and affects

product-market outcomes in an industry with free entry.

The remainder of the paper is organized as follows. Section 2 describes the institutional setting. Section 3 presents the data and descriptive facts. Section 4 develops the model of compliance spillovers and free entry. Section 5 characterizes the equilibrium and comparative statics. Section 6 studies welfare, Section 7 presents the quantitative illustration, and Section 8 concludes.

2 Institutional Setting

This paper studies the obligations that environmental regulation places on firms and how the cost of meeting them can depend on other firms. This section describes the obligations I measure in Section 3, who monitors and enforces them, and the institutional channels through which one firm's compliance can change the cost of compliance for others.

2.1 Compliance Obligations

Under the Clean Water Act, a facility that discharges pollutants from a point source into U.S. waters must hold a permit under the National Pollutant Discharge Elimination System (NPDES). A permit sets limits on the pollutants the facility may discharge and specifies how the facility must monitor its discharges and report the results, in Discharge Monitoring Reports submitted at the frequency the permit requires. Permits are issued for fixed terms of no more than five years, after which the facility must apply to renew them (Code of Federal Regulations, 2024a).

Under the Clean Air Act, stationary sources are subject to emission standards that apply by source category, such as the New Source Performance Standards and the National Emission Standards for Hazardous Air Pollutants. The EPA records the programs and subparts that apply to each facility, and a facility in several source categories is subject to several sets of requirements.

In both programs, obligations attach to facilities, permits, and source categories. A facility's monitoring and reporting schedule is set by its discharge points, its pollutants, and the standards that apply to it, not directly by how much it produces. A larger firm may operate more facilities, or facilities subject to more standards, but its obligations need not rise in proportion to its output. Fact 1 in Section 3.2 documents that they rise much less than proportionally. The counts I use measure these requirements, not the dollar cost of meeting them.

2.2 Monitoring and Enforcement

The EPA consolidates facility-level compliance and enforcement records in the Enforcement and Compliance History Online (ECHO) system. ECHO combines records from program databases, including the Integrated Compliance Information System for Clean Air Act sources (ICIS-Air) and for NPDES permits (ICIS-NPDES), and the Facility Registry Service (FRS) assigns each regulated facility an identifier that links its records across programs (U.S. Environmental Protection Agency, 2024b). ECHO covers more than one million regulated facilities (U.S. Environmental Protection Agency, 2024a). Appendix A describes the files I use.

The EPA authorizes states to administer these programs, and state agencies carry out much of the monitoring and enforcement. In the ICIS-Air violation records, state agencies determined 77,735 of the 101,550 recorded violations. The EPA sets national enforcement priorities for particular sectors and pollutants (U.S. Environmental Protection Agency, 2024d), and within them agencies allocate limited inspection resources across facilities, taking into account risk and past violations (Helland, 1998; Stafford, 2002; Gray and Shimshack, 2011). Noncompliance is common: about a quarter of NPDES permit-quarters between 2009 and 2025 are recorded as out of compliance (Section 3.4).

2.3 How Compliance Can Spill Over Across Firms

Because obligations are organized around shared rules, shared regulators, and shared inputs, the cost to one firm of meeting its obligations can depend on how much compliance other firms undertake. These institutions suggest several potential channels, which can work in either direction. I do not claim that the data identify any particular one.

Shared compliance infrastructure. Common standards, industry organizations, and certification programs such as ISO 14001, the environmental management standard of the International Organization for Standardization (Potoski and Prakash, 2005), can facilitate the diffusion of reporting practices, monitoring methods, and compliance know-how across firms. Such diffusion can lower the cost of meeting similar obligations, so that compliance by other firms lowers one's own compliance cost. The channel resembles the spillovers studied in research and development (d'Aspremont and Jacquemin, 1988), applied to compliance. In the model, it corresponds to a cooperative spillover, $\phi > 0$. Regulators sometimes build such sharing into compliance itself. California's statewide permit for industrial stormwater discharges lets facilities of the same industry

type form compliance groups that share sampling, expertise, and consolidated reports, which the State Water Board describes as a way to comply through shared resources at lower cost (California State Water Resources Control Board, 2014). Similarly, EPA recognizes industry-wide task forces through which pesticide registrants share the cost of developing the exposure data that regulation requires (U.S. Environmental Protection Agency, 1994).[2]

Constrained enforcement and shared inputs. If regulatory capacity or specialized compliance inputs are constrained, greater compliance activity by other firms can increase the time or price required to meet one's own obligations. Regulators review permits and inspect facilities with limited staff and budgets, and inputs such as specialized consultants, environmental engineers, testing laboratories, and attorneys may be in limited supply in the short run. Greater demand from other firms can then raise their price, in the spirit of Salop and Scheffman (1983). Targeting rules that compare a facility with its peers could also make one firm's position depend on others' compliance. In the model, these channels correspond to a rivalrous spillover, $\phi < 0$. Evidence from auditing shows that this channel can matter. When the Sarbanes–Oxley Act raised public companies' demand for audits, the short-run supply of auditors could not keep pace, and audit fee increases more than doubled for nonprofit organizations that used the same auditors (Duguay et al., 2020). Environmental testing offers a further example: as regulation of per- and polyfluoroalkyl substances (PFAS) expanded, the limited number of accredited laboratories led to long wait times for test results (U.S. Government Accountability Office, 2025).

Existing evidence on spillovers in environmental compliance concerns behavior: a sanction on one facility changes the compliance of others (Shimshack and Ward, 2005; Evans et al., 2018). Evidence on costs is scarcer. In auditing, regulation-driven compliance demand by some firms raised compliance costs for others that used the same inputs (Duguay et al., 2020), but I am not aware of comparable estimates for environmental compliance, and I do not assume which channel dominates. The model allows either sign, and Section 7 treats the size and sign of the spillover as sensitivity parameters.

[2] Under the European Union's regulation on the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), firms registering the same substance share data and submit jointly, which avoids duplicate testing (European Parliament and Council of the European Union, 2006).

3 Data and Descriptive Facts

This section describes the data and documents two facts about how environmental compliance obligations are organized. The first disciplines a modeling choice in Section 4: that compliance obligations grow less than proportionally with firm size. The second describes who operates regulated facilities, and is consistent with the market structure in the model.

3.1 Data

I merge two sources. From the EPA's ECHO system I take facility-level records under the Clean Air Act and the Clean Water Act; from Compustat Fundamentals Annual, accessed through Wharton Research Data Services (WRDS), I take firm financials. The ECHO system covers more than one million regulated facilities (U.S. Environmental Protection Agency, 2024a). The Facility Registry Service assigns each facility a unique identifier, which lets me link a facility's records across the air and water programs. For each facility I observe its industry, its Clean Water Act permits with their effective and expiration dates, the Clean Air Act regulations that apply to it, and its operating status.

I match EPA facilities to Compustat firms by fuzzy name comparison, blocking on three-digit North American Industry Classification System (NAICS) industry and the first letter of the name, and keeping pairs with a Jaro–Winkler similarity of at least 0.90. The match links 6,109 facilities to 1,121 publicly listed firms, which I refer to as listed firms. I prefer the conservative side of this trade-off: assigning a facility to the wrong firm would contaminate that firm's obligations directly, whereas missing a true match only reduces the sample. The cost is that the match misses facilities that listed firms own under different names, such as some subsidiaries, and facilities whose industry differs from their parent's principal industry. Earlier work also links EPA records to parent firms (Akey and Appel, 2021; Dasgupta et al., 2023); I use the link to measure how compliance obligations vary with firm size. Appendix A documents each source and reports the sample-selection counts at every step.

The two facts use different samples. Fact 1 uses the matched listed firms, for which I observe both compliance obligations and sales. Fact 2 uses the population of active, privately owned facilities that hold Clean Water Act permits.

3.2 Fact 1: Compliance Obligations Scale Much Less Than Proportionally with Firm Size

By compliance obligations I mean the monitoring, reporting, and permit requirements that a regulated facility must meet. EPA itemizes these requirements: each Clean Water Act permit and each applicable Clean Air Act regulation carries its own limits, monitoring schedule, and reporting duties. A facility that holds more permits, or falls under more regulations, therefore carries more obligations. I use this structure to measure obligations by counting the regulatory items that apply to a firm. The counts measure which requirements apply to a firm according to the regulator's records, not whether the firm meets them; whether a firm complies is a separate question, recorded in the EPA's violation records. The counts can understate a firm's obligations if a requirement applies but was never recorded, for example when a facility discharges without the permit it should hold.

Measuring compliance obligations. For each matched firm and year, I construct three counts, summarized in Table 1.

1. *Regulated facilities*: the number of the firm's facilities regulated under the Clean Air Act or the Clean Water Act. Each regulated facility carries its own set of requirements.

2. *Water permits*: the number of the firm's Clean Water Act discharge permits, issued under the National Pollutant Discharge Elimination System (NPDES), in effect during the year, that is, with an effective date on or before the year and an expiration date on or after it, and not terminated. Each permit sets discharge limits and requires the facility to monitor its discharges and file Discharge Monitoring Reports on the schedule the permit specifies.

3. *Air regulations*: the number of federal Clean Air Act programs and subparts that apply to the firm's facilities, such as specific New Source Performance Standards and National Emission Standards for Hazardous Air Pollutants. Each subpart imposes its own emission standards and monitoring, recordkeeping, and reporting requirements.

Water permits are my preferred measure, because permit records carry effective and expiration dates, so I observe which permits are in force in each year. The other two measures rely on facilities' current status and serve as supporting evidence.

I then estimate how each count varies with the firm's sales, controlling for year effects. If obligations scaled one-for-one with sales, the elasticity would equal one. Because the counts enter

Table 1. Measures of compliance obligations

Measure	Source	What one unit represents	Varies over time?
Regulated facilities	ECHO, FRS facility registry	a facility subject to Clean Air Act or Clean Water Act requirements	through the set of operating facilities
Water permits	ICIS-NPDES permit records	a discharge permit with its own limits and reporting schedule	yes: effective and expiration dates
Air regulations	ICIS-Air program subparts	an applicable air regulation with its own standards and reporting	no: current status only

Note: Counts are aggregated from facilities to firms through the facility–firm match. Appendix A documents the files, variables, and construction of each measure.

in logarithms, each estimate uses firm-years holding at least one obligation of the relevant type, so the elasticities describe how many obligations regulated firms carry, not whether a firm is regulated.

Figure 1 shows the result for water permits and air regulations. Across groups of firms whose sales range from about one three-hundredth of the geometric-mean firm's to more than thirty times it, obligations stay between roughly 0.4 and 1.7 times that firm's. The data lie far flatter than the dashed line that obligations proportional to sales would follow. The elasticity of water permits with respect to sales is 0.095 (standard error 0.016), and that of air regulations 0.134 (0.026). The number of regulated facilities gives a similar elasticity of 0.104 (0.014) (Table B.1). A firm with ten times the sales of another holds only about 1.2 to 1.4 times as many compliance obligations. Within firms over time, the elasticity of water permits is 0.068 (0.025).[3]

Because permit counts are small integers and many regulated firms hold exactly one permit, I also estimate a zero-truncated negative binomial count model of water permits, with industry and year effects (Appendix B). I parameterize the latent mean as proportional to sales raised to a power b, which allows obligations to vary nonlinearly with size; a version that also allows an additive constant places that constant at zero. The estimated exponent is $b = 0.45$ (standard error 0.05), and the restriction $b = 1$, under which the size-dependent part of obligations grows in proportion to sales, is strongly rejected within this specification. Because the sample contains only firms with at least one permit, the elasticity of the observed conditional mean is smaller than b:

[3]Water permits are the cleanest of the three measures, because permit records carry effective and expiration dates. For regulated facilities and air regulations, I count only facilities currently recorded as operating. The EPA files report each facility's current status but no closure dates, so this restriction also drops facilities that operated earlier in the sample and later closed. Including all facilities regardless of status gives elasticities of 0.120 (0.013) for regulated facilities and 0.149 (0.024) for air regulations. The air-regulation records carry no dates, so I do not report a within-firm estimate for them. Appendix B reports all estimates.

about 0.11 at the median sales of the smallest third of firms, 0.17 at the median firm, and 0.23 at the median of the largest third. Both the log–log regressions and the count model therefore imply that obligations rise substantially less than proportionally with firm size.

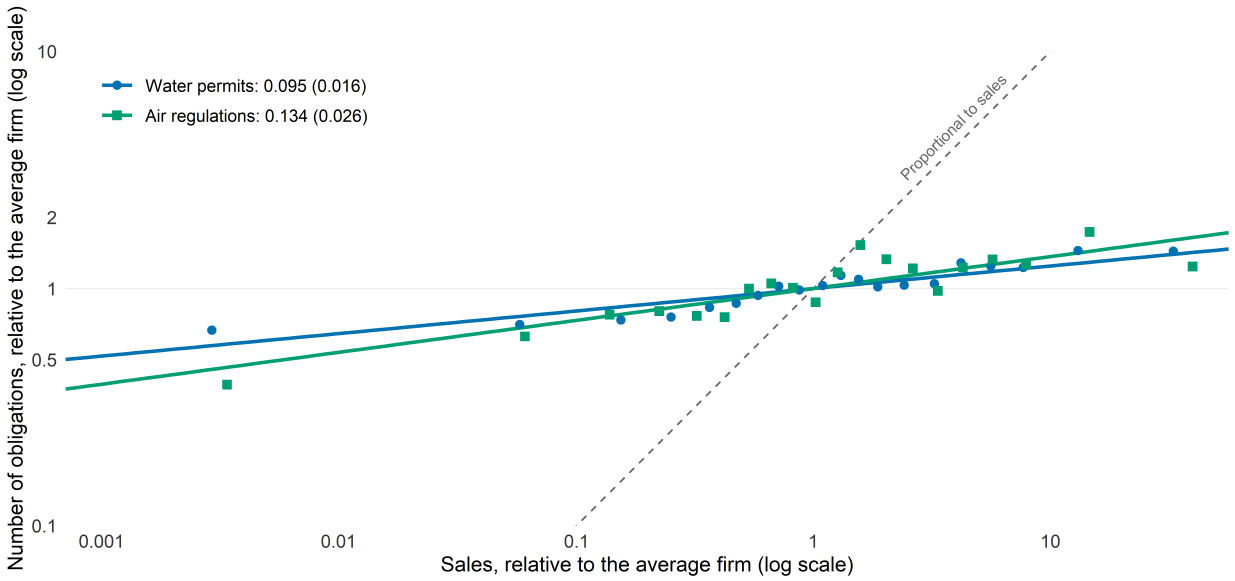


Figure 1. Compliance obligations and firm sales

Note: Each point is the average for one of 20 groups of firm-years, formed by sorting on sales so that every group contains the same number of firm-years. Blue points show water permits and green points show air regulations. Lines are fitted slopes estimated on all firm-years: 0.095 (standard error 0.016) for water permits and 0.134 (0.026) for air regulations. The dashed line shows obligations growing in proportion to sales. Both axes are net of year effects and shown relative to the geometric-mean firm on a logarithmic scale. Air regulations count only currently operating facilities. Appendix B describes the construction. Sample: listed firms matched to EPA-regulated facilities, 2009–2025.

This implies that smaller firms carry substantially more compliance obligations per dollar of sales, as Figure 2 shows. Dividing firms into thirds by employment, the largest third have about 26 times the median sales of the smallest third, but only about 2.4 times as many water permits on average. Smaller firms hold fewer permits in absolute terms, but far more per dollar of sales: 2.1 water permits per $1 billion of sales in the smallest third against 0.2 in the largest. Air regulations show the same pattern, falling from 11.1 to 0.8 per $1 billion of sales.[4] Regulated facilities behave similarly (Table B.3). In absolute terms, air regulations are higher in the middle third than in the top third, but per dollar of sales they fall across all three.

[4]Figure 2 divides the mean count in each third by median sales. Taking instead the median of each firm-year's own ratio of obligations to sales gives lower levels but the same pattern: water permits per $1 billion of sales fall from 2.0 in the smallest third to 0.12 in the largest, and air regulations from 6.2 to 0.53. Under either construction, obligations per dollar of sales fall by roughly an order of magnitude from the smallest to the largest third of firms.

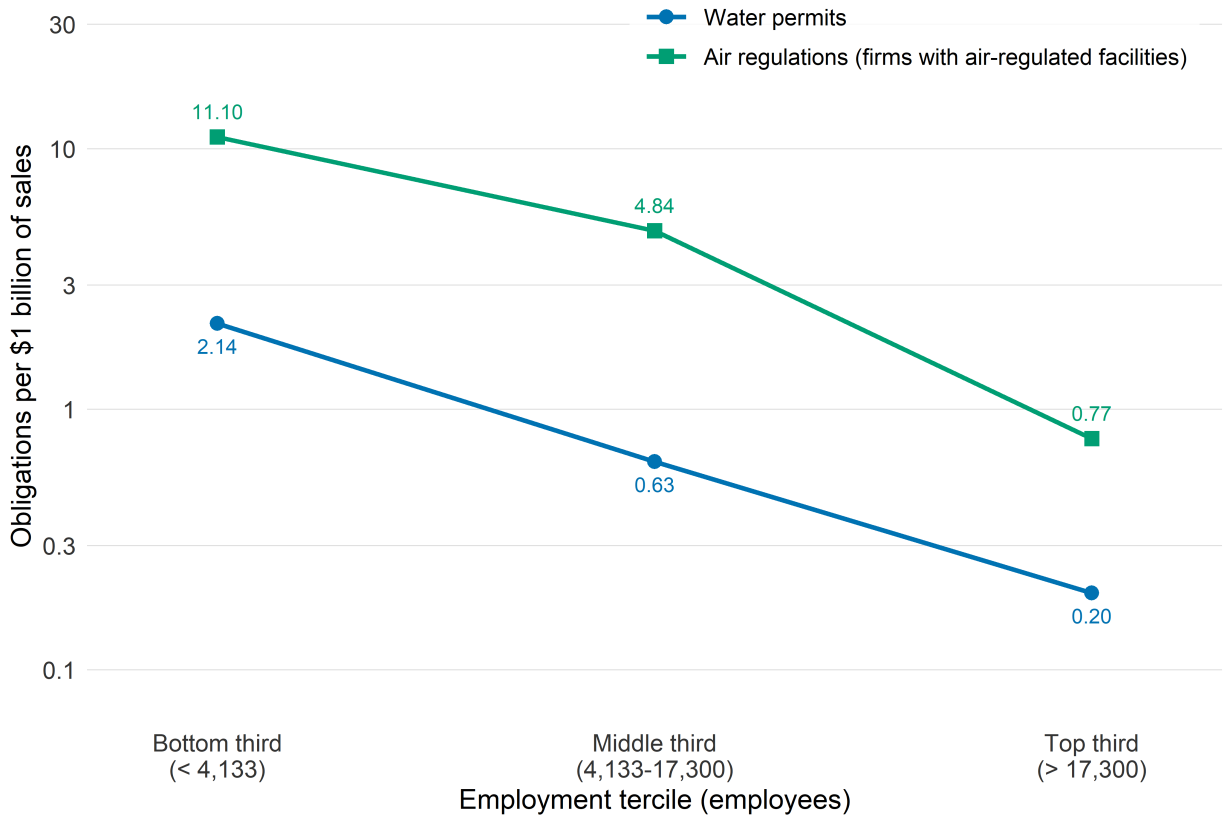


Figure 2. Compliance obligations per dollar of sales, by firm size

Note: Firm-years are divided into thirds by employment. Points show the mean number of water permits and of air regulations divided by median sales within each third, per $1 billion of sales, on a logarithmic scale. I use the median for sales because a few firm-years with near-zero sales make the mean unrepresentative in the smallest third; obligation counts are bounded below by one, so their means are not affected in this way. Table B.3 reports the mean counts and median sales behind each point. Air regulations are computed for firm-years with at least one air-regulated facility. Sample: listed firms matched to EPA-regulated facilities, 2009–2025.

The institutional structure provides a natural explanation for this pattern. Under the Clean Water Act, point-source discharges of pollutants into U.S. waters generally require a discharge permit, and the permit's requirements attach to the facility and its discharge rather than to the owner's sales. A permit is therefore a cost of operating a regulated facility, which a larger firm spreads over more output. Every regulated operation needs at least one permit, which is why the number of permits has a floor at one.[5]

These are listed firms, and even the smallest third includes firms with up to about 4,100 employees. Firms that are not listed, which operate most regulated facilities (Fact 2), are not in the sample. The gradient is steep among listed firms, but the data do not reveal whether it continues among smaller, privately held firms.

These counts measure how many obligations a firm carries, not what they cost. A larger facility may face more intensive monitoring under the same permit. The fact I use is therefore narrow: a firm's compliance obligations grow much less than proportionally with its sales. The institutional structure explains why: obligations attach to facilities and permits, every regulated operation carries at least one, and larger firms operate somewhat more facilities, but far fewer than their sales would imply. Evidence on the cost of regulatory compliance is consistent with economies of scale in compliance. Trebbi et al. (2023) find that compliance costs as a share of the wage bill peak among firms with about 500 employees and fall for larger firms, and Singla (2023) finds that increases in regulatory costs weigh more heavily on small firms than on large ones.

3.3 Fact 2: Listed Firms Account for a Small Share of Regulated Facilities

Listed firms are few and, on average, far larger than firms that are not listed (Davis et al., 2006). To see how much of the regulated population they account for, I restrict attention to active, privately owned and corporate facilities that hold Clean Water Act permits, excluding government, municipal, and federal facilities.[6]

Figure 3 reports, by industry, the share of these Clean Water Act–permitted facilities held by the listed firms in my sample. In most industries the share is below five percent, and across all

[5] Permits are not equally burdensome. Facilities that EPA classifies as major dischargers face heavier monitoring than minor ones, and some small dischargers are covered by simpler general permits. Larger firms operate major facilities more often, so a simple count may understate how much their obligations rise with size.

[6] The EPA records identify ownership type only for facilities with Clean Water Act permits, so the restriction applies to that universe. I count a facility as active if ECHO flags it as active; because the flag is blank rather than negative for most facilities, this restriction is conservative. I also exclude construction (NAICS 236), where most permits cover temporary storm-water discharges at construction sites, and waste management (NAICS 562), which contains many municipal operations. Appendix A reports results that include both.

industries it is 2.0 percent; the highest share in the figure, 9 percent, is in computer and electronic products. Among smaller industries, shares are higher where a few national firms operate many facilities: in couriers and messengers, two listed firms hold 39 percent of the industry's regulated facilities (Table A.8). Among the listed firms themselves, facilities are concentrated: the top ten percent of firms hold 54 percent of the matched facilities. Including facilities regardless of status gives nearly identical figures (1.8 and 54 percent). Because the name match is conservative, these shares are lower bounds. Table A.8 in the appendix reports the full breakdown by industry.

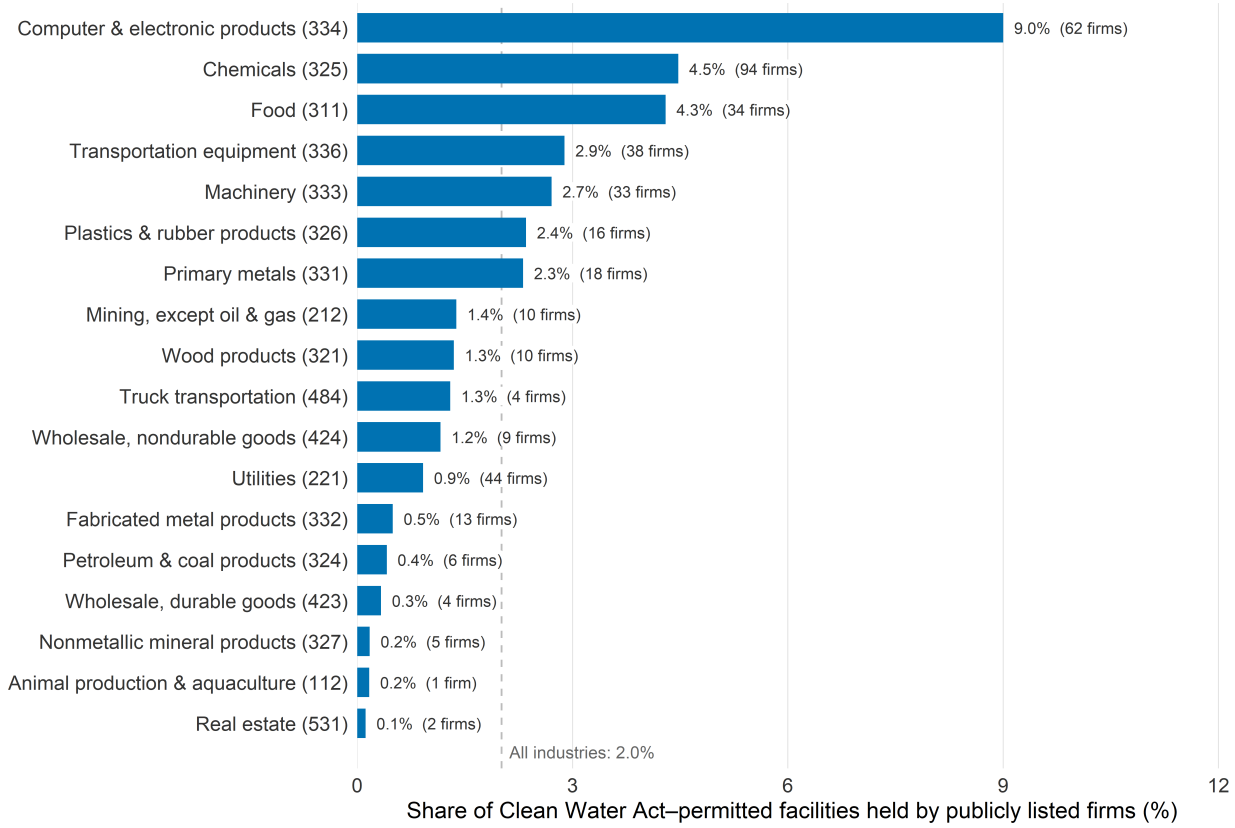


Figure 3. Share of Clean Water Act–permitted facilities held by listed firms, by industry

Note: Bars show the percentage of active, privately owned and corporate facilities holding Clean Water Act permits that belong to listed firms in my sample, for three-digit NAICS industries with at least 1,000 such facilities, excluding NAICS 236 and 562. Labels give the industry and its three-digit NAICS code; numbers at the end of each bar give the number of listed firms in the industry. The dashed line marks the share across all industries. Four industries in which no listed firm in my sample holds a regulated facility are omitted. Table A.8 reports all industries with at least 500 facilities. Shares are lower bounds because the name match is conservative.

Regulated industries therefore combine a small number of large, listed firms with many other operators. The data do not measure the output shares of these other operators, and the EPA

records carry no owner identifier with which to group their facilities into firms. The pattern is consistent with the market structure in Section 4, with a few incumbents and a fringe of smaller firms, but it does not establish that structure on its own.

3.4 What the Facts Imply for the Model

The two facts discipline the model in Section 4. Fact 1 shows that compliance obligations rise substantially less than proportionally with firm size. This motivates a compliance technology, $e_j = g(q_j)$, in which compliance is allowed to be concave in output rather than proportional to it. In the model, proportional compliance is the benchmark in which industry compliance depends only on market output. With concave compliance, splitting a given output across more, smaller firms can instead raise industry compliance; Section 5 gives the precise condition under which this occurs. The data do not identify a unique function g, because I observe sales rather than physical output, and only for listed firms. Fact 2 shows that listed firms account for a small share of regulated facilities, a pattern consistent with an industry that combines a small set of large incumbents with a broad population of other operators. That the fringe enters and exits freely is an assumption, standard in the analysis of oligopoly with free entry (Anderson et al., 2020; Peitz and Sato, 2026), rather than an implication of these data.

Two features of the data lie outside the model. First, noncompliance is common: 25.1 percent of Clean Water Act permit-quarters between 2009 and 2025 carry a significant or reportable noncompliance code. The model abstracts from this margin and treats compliance obligations as binding; Häckner and Herzing (2020) study the choice to comply directly. Second, the data do not identify how one firm's compliance affects another's costs. The model therefore takes the compliance spillover as a parameter and studies its consequences for either sign.

4 Model

I study an oligopoly with free entry in which firms must comply with a regulation and in which each firm's compliance cost depends on how much its rivals comply. The setting follows the aggregative approach to oligopoly with free entry of Anderson et al. (2020). I add one element: rivals' compliance, which enters every firm's costs through a compliance spillover of either sign.

4.1 Firms and Demand

Firms produce a homogeneous good. Let $q_j \geq 0$ denote firm j's output and Q market output, defined below. Inverse demand is $P(Q)$.

Assumption 1 (Demand)**.** *P is twice continuously differentiable on $\{Q : P(Q) > 0\}$, with $P'(Q) < 0$ and $P'(Q) + qP''(Q) < 0$ for all $q \in [0, Q]$.*

Assumption 1 is standard in Cournot oligopoly (cf. Novshek, 1985). It ensures that the product-market component of a firm's incentive to produce is decreasing in market output, since $P'(Q) + qP''(Q) < 0$.

The market has two kinds of firms. There are $M_I \geq 1$ *incumbents*, indexed by $i = 1, \ldots, M_I$, with constant marginal costs c_i. There is also a *fringe* of identical potential entrants with marginal cost c_f. Each fringe firm that becomes active pays an entry cost $K > 0$. As in Anderson et al. (2020) and Peitz and Sato (2026), I treat the mass of active fringe firms, $M_f \geq 0$, as a continuous variable and ignore integer constraints. In a symmetric equilibrium every fringe firm produces the same output q_f, so market output is

$$Q \;=\; \sum_{i=1}^{M_I} q_i \;+\; M_f\, q_f.$$

4.2 Compliance

Each active firm j exerts compliance effort $e_j \geq 0$, which represents the obligations it must meet: monitoring, reporting, and the requirements attached to its permits. I abstract from the decision whether to comply and treat each active firm as meeting the obligations that the regulation imposes on it, so compliance effort is not an additional strategic choice. It depends on the firm's output through a compliance technology.

Definition 1 (Compliance technology)**.** *Firm j's compliance effort is $e_j = g(q_j)$, where $g : [0, \infty) \to [0, \infty)$ is continuous, twice continuously differentiable on $(0, \infty)$, nondecreasing, and satisfies $g(q) > 0$ for $q > 0$. Compliance is* proportional *if $g(q) = rq$ for some $r > 0$.*

Four forms of g recur in the paper:

- *proportional*, $g(q) = rq$: obligations grow one-for-one with output;
- *fixed*, $g(q) = \bar{e}$: every active firm bears the same obligation, whatever its output;

- *fixed plus proportional*, $g(q) = \bar{e} + rq$;
- *power*, $g(q) = Aq^b$ with $0 < b < 1$: obligations grow less than proportionally with output.

The last three are concave examples of sub-proportional compliance. They are consistent with the scaling pattern documented in Fact 1, under the mapping from obligation counts to compliance effort and from sales to output.

Industry compliance is

$$E \;=\; \sum_{i=1}^{M_I} g(q_i) \;+\; M_f\, g(q_f),$$

and the compliance of firm j's rivals is

$$E_{-j} \;=\; E - g(q_j). \tag{1}$$

Proportional compliance is special. With $g(q) = rq$, industry compliance equals rQ: it depends on market output alone, not on how that output is divided among firms. With nonproportional compliance, industry compliance generally also depends on how output is divided. For example, in a symmetric market in which N firms each produce Q/N, industry compliance is $N\, g(Q/N)$. Its derivative with respect to N at a given Q is $g(q) - q\, g'(q)$, evaluated at $q = Q/N$. If g is concave with $g(0) \geq 0$, this derivative is nonnegative, and it is strictly positive if $g(0) > 0$ or g is strictly concave. Entry, which divides output among more firms, then raises industry compliance even when market output does not change. Section 5 gives the precise condition for the asymmetric incumbent–fringe market, and shows that this dependence drives the results.

Firm j's compliance cost is

$$\kappa(e_j)\, s(E_{-j}; \phi), \tag{2}$$

where κ is the stand-alone compliance cost and $s(E_{-j}; \phi) > 0$ captures the compliance spillover: how rivals' compliance changes the cost of meeting one's own obligations.

Assumption 2 (Compliance cost and compliance spillover)**.** *(i) κ is twice continuously differentiable and nondecreasing, with $\kappa(0) = 0$, $\kappa'(e) > 0$ and $\kappa''(e) \geq 0$ for $e > 0$. (ii) s is continuously differentiable, with $s(E_{-j}; 0) = 1$ for all E_{-j}. (iii) $\partial s / \partial E_{-j}$ has the sign of $-\phi$, and equals zero only if $\phi = 0$: rivals' compliance lowers a firm's compliance cost when $\phi > 0$ and raises it when $\phi < 0$.*

I call the spillover *cooperative* when $\phi > 0$ and *rivalrous* when $\phi < 0$. When $\phi = 0$, compliance

is private: each firm's compliance cost depends only on its own effort. Section 4.5 discusses the sources of each sign. It is convenient to write $C(q) \equiv \kappa(g(q))$ for the stand-alone compliance cost of producing q. Because g may be concave, C need not be convex even though κ is; Section 5 imposes the curvature condition on C that each firm's problem requires.

4.3 Timing and Equilibrium

The game has two stages (Figure 4). In the first stage, fringe firms decide whether to enter. In the second stage, all active firms choose outputs simultaneously, and each firm's compliance effort follows from its output through g. Firm j's operating profit is

$$\pi_j \;=\; \big[P(Q) - c_j\big]q_j \;-\; \kappa\big(g(q_j)\big)\, s(E_{-j};\phi), \tag{3}$$

and each active fringe firm also pays the entry cost K.

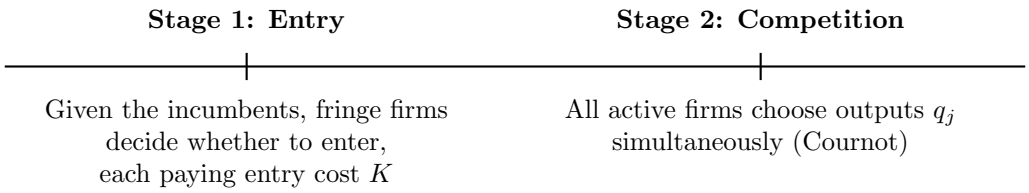


Figure 4. Timing of the game

Definition 2 (Free-entry equilibrium)**.** *A free-entry equilibrium consists of outputs* $(q_1, \ldots, q_{M_I}, q_f)$ *and a fringe mass* $M_f \geq 0$ *such that (i) given* M_f*, each active firm's output maximizes its operating profit* (3)*, taking rivals' outputs as given; (ii) if* $M_f > 0$*, each active fringe firm earns zero profit net of the entry cost,* $\pi_f - K = 0$*; if* $M_f = 0$*, then* $\pi_f - K \leq 0$*.*

I focus on equilibria with an active fringe, $M_f > 0$. The fringe is then the market's corrective force: its entry and exit adjust until the zero-profit condition holds.

For a given fringe mass M_f, let $\widehat{\pi}_f(M_f)$ denote a fringe firm's operating profit evaluated at the second-stage equilibrium that M_f induces.

Assumption 3 (Stability)**.** *At every free-entry equilibrium with* $M_f > 0$*,*

$$\frac{d}{dM_f}\,\widehat{\pi}_f(M_f) \;<\; 0.$$

Assumption 3 requires that additional entry lowers the equilibrium profitability of a fringe firm, so that entry is self-limiting. Conditions of this kind are standard in the analysis of free entry in oligopoly (cf. Seade, 1980). Under a rivalrous spillover, both the product-market channel and the compliance-cost channel work toward reducing fringe profitability as entry rises. Under a cooperative spillover, the assumption also limits the strength of the cost-saving spillover relative to the competitive effect of entry.

4.4 The Entry–Compliance Curve

The free-entry condition ties market output to industry compliance. Given market output Q and industry compliance E, a fringe firm chooses its output, and its rivals comply by E less its own compliance. Let $\Pi_f(Q, E)$ denote the resulting operating profit of a fringe firm; Section 5 shows that this function is well defined under regularity conditions.

Definition 3 (Entry–compliance curve)**.** *The* entry–compliance curve *is the set of pairs* (Q, E) *at which a fringe firm earns zero profit net of the entry cost:*

$$\mathcal{Z} \;=\; \big\{(Q, E) : \Pi_f(Q, E) - K = 0\big\}.$$

Every free-entry equilibrium with an active fringe lies on the entry–compliance curve. The curve is built from the fringe's primitives alone: it contains no characteristic of the incumbents. Section 5 shows that its slope has the sign of the spillover. The fixed technology makes this explicit. With $g(q) = \bar{e}$, a firm's compliance cost does not depend on its own output, so for an active firm with an interior output the first-order condition is $P(Q) + P'(Q)\, q_j - c_j = 0$, which gives

$$q_j(Q) \;=\; \frac{P(Q) - c_j}{-P'(Q)}. \tag{4}$$

Each firm's output depends on market output alone. Define a fringe firm's operating profit before compliance costs, $v_f(Q) = [P(Q) - c_f]\, q_f(Q)$, which is decreasing under Assumption 1. A fringe firm's rivals comply by $E - \bar{e}$, so

$$\Pi_f(Q, E) \;=\; v_f(Q) \;-\; \kappa(\bar{e})\, s(E - \bar{e}; \phi). \tag{5}$$

Lemma 1 (Slope of the entry–compliance curve)**.** *Under the fixed technology and Assumptions 1–2,*

along the entry–compliance curve,

$$\frac{dE}{dQ} = \frac{v_f'(Q)}{\kappa(\bar{e})\, s_E(E - \bar{e}; \phi)}$$

for $\phi \neq 0$, where s_E denotes the derivative of s with respect to its first argument. The curve slopes upward if $\phi > 0$ and downward if $\phi < 0$. If $\phi = 0$, the zero-profit condition pins down market output independently of industry compliance.

Proof. Implicit differentiation of $v_f(Q) - \kappa(\bar{e})\, s(E - \bar{e}; \phi) = K$ gives $v_f'(Q)\, dQ - \kappa(\bar{e})\, s_E(E - \bar{e}; \phi)\, dE = 0$. The numerator $v_f'(Q)$ is negative. By Assumption 2(iii), $s_E < 0$ if $\phi > 0$ and $s_E > 0$ if $\phi < 0$. Hence $dE/dQ > 0$ if $\phi > 0$ and $dE/dQ < 0$ if $\phi < 0$. If $\phi = 0$, $s_E = 0$ and the condition reduces to $v_f(Q) = K + \kappa(\bar{e})$. □

4.5 Discussion of the Assumptions

Assumptions 1 and 3, together with Assumption 2(i), are technical conditions standard in Cournot oligopoly with free entry; the curvature condition on C in Section 5 is their counterpart for compliance costs. The remaining assumptions carry economic content, and I discuss the evidence and reasoning behind each.

Compliance obligations are set by the regulation. I abstract from the decision whether to comply and treat each active firm as meeting the obligations that the regulation imposes on it. The model therefore concerns the cost of satisfying compliance obligations, not the choice to violate them. Noncompliance does occur: 25.1 percent of Clean Water Act permit-quarters between 2009 and 2025 carry a significant or reportable noncompliance code (Appendix A.5). Häckner and Herzing (2020) study the decision to comply directly, in an oligopoly model in which firms choose whether to comply.

Sub-proportional compliance. Fact 1 shows that compliance obligations grow much less than proportionally with firm size, with a substantial mass of regulated firm-years at one water permit. The compliance technology in Definition 1 allows this pattern without imposing a particular form. In the data, obligations attach to facilities and permits, and larger firms operate somewhat more facilities. I do not model facilities separately: g is a reduced-form representation of how a firm's obligations vary with its scale. Proportional compliance serves as the benchmark, not as a descrip-

tion of the data. Section 7 disciplines the shape of g with the count-model estimates in Section 3.2, under the assumption that obligation counts measure compliance effort and sales measure output.

Compliance spillovers of either sign. The empirical literature documents cross-firm responses to regulatory enforcement in both directions (Shimshack and Ward, 2005; Evans et al., 2018; Dasgupta et al., 2023). These papers identify responses in compliance behavior, not in compliance costs. Evidence on costs comes from auditing, where regulation-driven demand from public companies raised audit fees for other clients of the same auditors (Duguay et al., 2020). I represent one possible source of such interdependence in reduced form, through compliance costs. Regulator reputation and direct strategic responses in the product market provide alternative channels (Shimshack and Ward, 2005; Evans et al., 2018), and I do not claim that the cost channel is the only one.[7]

Two mechanisms give the spillover its sign. First, compliance requires specialized compliance inputs, such as environmental consultants and engineers, testing laboratories, and monitoring equipment. If these inputs are supplied at a price that rises with industry demand, rivals' compliance raises the price a firm pays to meet its own obligations, and the spillover is rivalrous. Second, compliance know-how diffuses: shared reporting templates, industry guidance, and common infrastructure lower the cost of meeting a given obligation as more firms comply, and the spillover is cooperative. The sign of ϕ reflects which mechanism dominates in a given industry. I take ϕ as a parameter and study the consequences of both signs.

The form of the spillover. I use the multiplicative form $\kappa(e_j)\, s(E_{-j}; \phi)$ for two reasons. It keeps compliance costs positive for any level of rivals' compliance, and it nests private compliance as the case $\phi = 0$. Appendix D shows that this form arises from a market for compliance inputs whose price depends on the industry's obligations: a capacity-constrained supply of compliance inputs yields a rivalrous spillover, and shared fixed costs of supplying them yield a cooperative one.

Free entry of the fringe. That fringe firms enter and exit freely is an assumption, standard in the analysis of oligopoly with free entry (Anderson et al., 2020; Peitz and Sato, 2026). Fact 2 is consistent with a market structure that combines a few large firms with many other operators, but it does not measure entry and exit.

[7] Organizational sociology offers a further explanation for why compliance behavior converges across firms: imitation under uncertainty, or mimetic isomorphism (DiMaggio and Powell, 1983). My model does not rely on imitation; the compliance spillover operates through the cost of meeting compliance obligations.

4.6 A Linear Example

For illustration, I use linear demand, $P(Q) = \alpha - \beta Q$, a stand-alone compliance cost $\kappa(e) = \vartheta e^m/m$ with $\vartheta > 0$ and $m \geq 2$, and an exponential spillover,

$$s(E_{-j}; \phi) = e^{-\phi E_{-j}}.$$

This specification satisfies Assumptions 1 and 2. With the fixed technology $g(q) = \bar{e}$, fringe output is $q_f(Q) = [P(Q) - c_f]/\beta$, and with $\bar{\kappa} = \kappa(\bar{e})$ the entry–compliance curve is

$$\frac{\big[\alpha - \beta Q - c_f\big]^2}{\beta} \; - \; \bar{\kappa}\, e^{-\phi(E-\bar{e})} \; = \; K,$$

or equivalently, taking the root with a positive price–cost margin,

$$Q(E) \; = \; \frac{\alpha - c_f - \sqrt{\beta\big(K + \bar{\kappa}\, e^{-\phi(E-\bar{e})}\big)}}{\beta}.$$

When $\phi > 0$, the term $e^{-\phi(E-\bar{e})}$ falls as industry compliance rises, so $Q(E)$ increases; when $\phi < 0$, it rises, so $Q(E)$ decreases. The example thus displays the two slopes of Lemma 1 in closed form. With other technologies the curve need not have a closed form, but Section 5 shows that its slope has the same sign.

5 Equilibrium and Comparative Statics

This section studies how shocks to incumbents affect long-run market outcomes. I begin with proportional compliance, $g(q) = rq$. In that case, I show that free entry completely absorbs shocks to incumbents at the level of market output, as in the neutrality result of Anderson et al. (2020): market output, price, industry compliance, and consumer surplus are unchanged, while incumbents' outputs and the fringe mass adjust. I then turn to a general compliance technology. The main result shows that a shock to incumbents changes industry compliance whenever entry changes how much compliance is needed to supply a given output, and that the compliance spillover then moves market output. Two examples, fixed obligations and a power technology, make the mechanism concrete. The results hold for general demand and compliance-cost functions satisfying the stated conditions. Appendix C contains the proofs.

5.1 Proportional Compliance: Free-Entry Neutrality

The analysis rests on writing each firm's output as a function of market output. Under proportional compliance, $g(q) = rq$, firm j's compliance effort is $e_j = rq_j$, its rivals comply by $E_{-j} = r(Q - q_j)$, and its operating profit is

$$\pi_j \;=\; \big[P(Q) - c_j\big]q_j \;-\; \kappa(rq_j)\, s\big(r(Q - q_j);\phi\big). \tag{6}$$

When firm j chooses its output, it takes its rivals' output $Q - q_j$ as given. Its rivals' compliance $r(Q - q_j)$ is therefore fixed in its own problem, and its first-order condition is

$$G_j(q_j, Q) \;\equiv\; P(Q) + P'(Q)\, q_j - c_j - r\,\kappa'(rq_j)\, s\big(r(Q - q_j);\phi\big) \;=\; 0. \tag{7}$$

The condition depends only on the firm's own output and market output. Under Assumptions 1 and 2, firm j's profit is strictly concave in its own output (Appendix C), so this condition characterizes its best response. I impose three conditions on the primitives.

Assumption 4 (Per-unit regularity)**.** *For $0 < q \le Q$ with $P(Q) > 0$:*

(U1) $\dfrac{\partial G_j}{\partial q_j} \;=\; P'(Q) - r^2\kappa''(rq)\, s + r^2\kappa'(rq)\, s_E \;<\; 0;$

(U2) $P'(Q)\, q - r\,\kappa(rq)\, s_E \;<\; 0;$

(U3) $\dfrac{\partial G_j}{\partial Q} \;=\; P'(Q) + qP''(Q) - r^2\kappa'(rq)\, s_E \;<\; 0.$

Here s and s_E are evaluated at $r(Q - q)$. Condition (U1) ensures that, for a given level of market output, at most one positive output is consistent with the first-order condition, and that the backward reply defined below is well behaved; it is equivalent to the requirement of Novshek (1985) that each firm's best reply has slope greater than -1. Together with (U1), condition (U2) ensures that a fringe firm's operating profit falls as market output rises. Condition (U3) states that quantities are strategic substitutes: a firm produces less when market output is higher.

Each condition holds automatically for a sign of the spillover. By Assumption 2(iii), s_E has the sign of $-\phi$. Condition (U1) therefore holds whenever $\phi \ge 0$, and conditions (U2) and (U3) hold whenever $\phi \le 0$. When the spillover has the other sign, each condition limits how strongly rivals' compliance changes a firm's costs, relative to the slope and curvature of demand and the convexity of compliance costs. For the parameter values used in the linear illustration below, all three conditions hold over the relevant equilibrium domain.

Lemma 2 (Outputs as functions of market output)**.** *Under proportional compliance, Assumptions 1–2, and condition (U1) of Assumption 4, let $b_j(0)$ denote firm j's output when it alone serves the market. For each level of market output $Q \geq b_j(0)$, firm j has a unique output $h_j(Q) \geq 0$ consistent with its best reply, with $h_j(Q) > 0$ if and only if $G_j(0, Q) > 0$. Where $h_j(Q) > 0$, the function h_j is continuously differentiable, with $1 - h_j'(Q) > 0$. Under condition (U3), h_j is also nonincreasing.*

The functions h_j are the backward replies of Novshek (1985): the output that firm j chooses when market output equals Q. A fringe firm's operating profit, when market output is Q, is then

$$\Pi_f(Q) \;=\; \big[P(Q) - c_f\big]\, h_f(Q) \;-\; \kappa\big(r\, h_f(Q)\big)\, s\Big(r\big(Q - h_f(Q)\big); \phi\Big). \tag{8}$$

This expression contains no characteristic of the incumbents. That observation drives the first result.

Proposition 1 (Free-entry neutrality under proportional compliance)**.** *Suppose compliance is proportional, $g(q) = rq$, and that Assumptions 1–2 and conditions (U1) and (U2) of Assumption 4 hold.*

(i) The fringe's operating profit $\Pi_f(Q)$ is strictly decreasing in market output where $h_f(Q) > 0$. Hence there is at most one market output Q^* satisfying $\Pi_f(Q^*) = K$, and every free-entry equilibrium with an active fringe has that market output. Lemma 3 below gives conditions under which such an equilibrium exists.

(ii) If an active-fringe equilibrium exists, Q^* depends only on demand, the fringe's marginal cost c_f, the entry cost K, and the compliance primitives r, κ, s, and ϕ. It does not depend on the number of incumbents or on their costs.

(iii) Consequently, any change that affects only the incumbents, such as a change in their costs, their number, or a merger among them, leaves unchanged the market price $P(Q^*)$, industry compliance $E^* = rQ^*$, each fringe firm's output $h_f(Q^*)$, and consumer surplus, provided the fringe remains active. Such a change is absorbed through adjustments in the incumbents' outputs and the fringe mass $M_f = \big[Q^* - \sum_i h_i(Q^*)\big]/h_f(Q^*)$, while the market outcomes listed above remain unchanged.

The logic is the one behind the neutrality results of Anderson et al. (2020). When compliance is proportional to output, a fringe firm's rivals comply in proportion to their output, so both its

revenue and its compliance costs depend only on its own output and market output. The fringe's zero-profit condition therefore involves market output alone, and it pins that output down. If incumbents become less efficient, they produce less, and fringe firms enter until market output returns to Q^*. The compliance spillover does not change this conclusion: it changes the level of Q^*, through its effect on the fringe's costs, but not the response of Q^* to incumbents.

The linear example illustrates. With $\alpha = 20$, $\beta = 1$, $\vartheta = 1$, $m = 2$, $c_f = 4$, $K = 4$, and $r = 0.3$, market output under a rivalrous spillover ($\phi = -0.3$) equals 13.63 in each of five incumbent configurations: one incumbent with cost 1.0; two with costs (1.0, 1.0), (2.5, 2.5), or (1.0, 3.0); and three with costs (1.0, 1.5, 2.0). Across these configurations the fringe mass ranges from 0.85 to 4.88. By Proposition 1, market output is independent of incumbents' costs in any configuration in which the fringe remains active. The spillover affects the level: market output is 13.87 when compliance is private ($\phi = 0$) and 13.95 under a cooperative spillover ($\phi = 0.3$). In each case it is again the same across the five configurations.

Proposition 1 characterizes equilibria with an active fringe. The next result gives conditions under which such an equilibrium exists. Let Q^0 denote market output when only incumbents are active, the solution of $Q = \sum_i h_i(Q)$; let $b_f(0)$ denote the output of a fringe firm serving the market alone; let $\hat{Q} = \max\{Q^0, b_f(0)\}$; and let $\bar{Q}$ denote the output at which price equals the fringe's marginal cost, $P(\bar{Q}) = c_f$.

Lemma 3 (Existence and uniqueness under proportional compliance)**.** *Suppose compliance is proportional, Assumptions 1–2 and 4 hold, and $\bar{Q}$ exists. If entry is profitable at $\hat{Q}$, that is, $\Pi_f(\hat{Q}) > K$, then there exists a unique free-entry equilibrium with an active fringe, and its market output satisfies $\hat{Q} < Q^* < \bar{Q}$.*

5.2 General Compliance: The Main Result

With a general compliance technology, a firm's output depends on both market output and industry compliance. Firm j takes its rivals' outputs, and hence their compliance $E_{-j} = E - g(q_j)$, as given. Writing $C(q) = \kappa(g(q))$, its first-order condition is

$$G_j(q_j, Q, E) \;\equiv\; P(Q) + P'(Q)\, q_j - c_j - C'(q_j)\, s\big(E - g(q_j); \phi\big) \;=\; 0,$$

which defines its output as a function $h_j(Q, E)$. A free-entry equilibrium with an active fringe is the intersection of two curves in (Q, E) space. The first is the entry–compliance curve,

$$\Pi_f(Q,E) \;\equiv\; \big[P(Q) - c_f\big]\, h_f(Q,E) \;-\; C\big(h_f(Q,E)\big)\, s\big(E - g(h_f(Q,E)); \phi\big) \;=\; K, \tag{9}$$

which contains no characteristic of the incumbents. The second is the adding-up curve. Solving $E = \sum_i g(h_i) + M_f\, g(h_f)$ for the fringe mass and substituting into $Q = \sum_i h_i + M_f\, h_f$ gives

$$A(Q,E;c) \;\equiv\; \sum_{i=1}^{M_I} h_i(Q,E) \;+\; \frac{E - \sum_i g\big(h_i(Q,E)\big)}{g\big(h_f(Q,E)\big)}\, h_f(Q,E) \;-\; Q \;=\; 0. \tag{10}$$

The incumbents enter only through the adding-up curve. A shock to incumbents therefore shifts the adding-up curve and moves the equilibrium along the entry–compliance curve.

How the adding-up curve shifts depends on one object. Holding market output and industry compliance fixed, suppose incumbent i produces one unit less. The fringe must supply that unit, which requires compliance $g(q_f)/q_f$ per unit of fringe output, while the incumbent saves compliance $g'(q_i)$ at the margin. The difference,

$$\Omega_{if} \;\equiv\; \frac{g(q_f)}{q_f} \;-\; g'(q_i), \tag{11}$$

is the change in the compliance needed to supply a given output when output moves from incumbent i to the fringe. I call it the *entry–compliance wedge*. Under proportional compliance, $\Omega_{if} = r - r = 0$.

Regularity. I impose the following conditions, which generalize Assumption 4. For each firm, s and s_E are evaluated at its rivals' compliance.

Assumption 5 (Regularity)**. (R0)** *For each firm and every own output $q > 0$ with $P(Q) > 0$,* $2P'(Q) + qP''(Q) - C''(q)\, s < 0$.

(R1) *At the equilibrium, $P'(Q) - C''(q)\, s + C'(q)\, g'(q)\, s_E < 0$ for each active firm.*

(R2) *At the equilibrium, $P'(Q)\, q_f - C(q_f)\, g'(q_f)\, s_E < 0$.*

(R3) *At the equilibrium, $\partial \Pi_f / \partial Q < 0$: holding industry compliance fixed, a fringe firm earns less when market output is higher.*

(R4) *At the equilibrium, the second stage, which determines market output and industry compliance for a given fringe mass, is regular in the sense of Appendix C.*

Condition (R0) makes each firm's problem strictly concave in its own output. Because $C = \kappa \circ g$ need not be convex when g is concave, (R0) restricts how strongly compliance costs can fall with scale. Condition (R1) ensures that each firm's output is a well-defined function of market output and industry compliance. It holds when $\phi \geq 0$ and C is convex. Condition (R2) holds when $\phi \leq 0$, and, given (R1), so does (R3). When the spillover has the other sign, (R1)–(R3) limit how strongly rivals' compliance changes a firm's costs. Condition (R4) is a regularity condition on the second stage; it holds in every example below.

Proposition 2 (Shocks to incumbents with general compliance)**.** *Suppose Assumptions 1–3 and 5 hold, and the equilibrium has an active fringe. Consider an increase in the marginal cost c_i of an active incumbent with interior output. Then*

$$\frac{dQ^*}{dc_i} \;=\; -\,\frac{A_c\,\Pi_E}{D}, \qquad \frac{dE^*}{dc_i} \;=\; \frac{\Pi_Q\,A_c}{D},$$

where Π_Q and Π_E are the partial derivatives of Π_f, $A_c = (\partial h_i/\partial c_i)\,[h_f/g(h_f)]\,\Omega_{if}$ is the shift of the adding-up curve, and $D > 0$. Consequently:

(i) industry compliance changes with the sign of the wedge,

$$\operatorname{sgn}\bigl(dE^*/dc_i\bigr) \;=\; \operatorname{sgn}\bigl(\Omega_{if}\bigr);$$

(ii) market output changes with the sign of the wedge times the sign of the spillover,

$$\operatorname{sgn}\bigl(dQ^*/dc_i\bigr) \;=\; \operatorname{sgn}\bigl(\Omega_{if}\bigr)\cdot\operatorname{sgn}(\phi).$$

In particular, market output is unchanged if $\Omega_{if} = 0$, as under proportional compliance, or if $\phi = 0$.

The mechanism has three steps. First, when an incumbent becomes less efficient, its output falls. Holding market output fixed, replacing that output through the fringe changes the compliance required to supply it by the wedge Ω_{if}. Second, if the fringe needs more compliance per unit than the incumbent saved, $\Omega_{if} > 0$, industry compliance rises in equilibrium. Third, the higher industry compliance changes every fringe firm's costs through the spillover. Under a cooperative spillover, it

lowers the fringe's costs, so the fringe expands beyond the point that restores market output, and market output rises. Under a rivalrous spillover, it raises the fringe's costs, so the fringe stops short, and market output falls. The model's stability assumption signs the denominator: the appendix shows that Assumption 3 holds if and only if $D > 0$.

Proposition 2 identifies two conditions under which a shock to incumbents reaches market output: a nonzero wedge and a nonzero spillover. The wedge is positive under a simple condition.

Corollary 1 (A sufficient condition for a positive wedge)**.** *If g is concave and incumbent i produces at least as much as a fringe firm, $q_i \geq q_f > 0$, then $\Omega_{if} \geq 0$. The inequality is strict if $g(0) > 0$ or g is strictly concave on $[0, q_i]$.*

Concave compliance means that output supplied by fringe firms requires more compliance per unit than output supplied at the margin by incumbents. Under (R1), an incumbent with lower marginal cost than a fringe firm produces more at a given (Q, E). Under the conditions of Corollary 1, the direction of the effect is therefore set by the sign of the spillover alone.

5.3 Examples

Fixed obligations. With the fixed technology $g(q) = \bar{e}$, the wedge is $\Omega_{if} = \bar{e}/q_f > 0$. Each firm's output depends on market output alone, as in (4), and the results take an explicit form.

Corollary 2 (Fixed obligations)**.** *Under the fixed technology, suppose Assumptions 1–3 hold and the equilibrium has an active fringe. After an increase in the marginal cost of an active incumbent, the fringe mass and industry compliance rise, and*

$$\frac{dQ^*}{dc_i} = \frac{\kappa(\bar{e})\,\bar{e}\,s_E\big(E^* - \bar{e}; \phi\big)}{|P'(Q^*)|\,h_f(Q^*)\,\Phi'(Q^*)},$$

where $\Phi(Q) = v_f(Q) - \kappa(\bar{e})\,s(\varepsilon(Q) - \bar{e}; \phi) - K$, $\varepsilon(Q)$ is the industry compliance consistent with market output Q, and stability is equivalent to $\Phi'(Q^*) < 0$.

Under the fixed technology, every fringe firm brings the same obligation, so industry compliance is proportional to the number of active firms. The formula shows directly that the response vanishes as the fixed obligation $\bar{e}$ vanishes. With both a fixed and a proportional component, $g(q) = \bar{e} + rq$, the wedge is again $\bar{e}/q_f$, and the appendix shows that dQ^*/dc_i tends to zero as $\bar{e} \to 0$, provided the active-fringe equilibrium converges continuously to the equilibrium under proportional compliance. The general result thus connects continuously to Proposition 1.

The fixed technology also clarifies the effect of a merger. When incumbents a and b merge into a single firm m, the merger removes one active firm, and at a given market output increases the fringe mass needed to replace the lost incumbent output. With fixed obligations, industry compliance at a given market output changes by

$$\Delta\varepsilon(Q) \;=\; \bar{e}\left[-1+\frac{h_a(Q)+h_b(Q)-h_m(Q)}{h_f(Q)}\right].$$

The merger therefore raises industry compliance if and only if the reduction in the merging incumbents' output exceeds the output of one fringe firm. For two identical incumbents merging without cost savings, the condition becomes $h_i(Q) > h_f(Q)$: the incumbents are larger than fringe firms. When, in addition, $\Delta\varepsilon(Q) > 0$ at every relevant market output and the zero-profit residual Φ is strictly decreasing over the relevant domain, a merger lowers long-run market output under a rivalrous spillover and raises it under a cooperative one. In the aggregative framework without compliance spillovers, by contrast, free entry leaves long-run consumer surplus unchanged after a merger (Anderson et al., 2020).

The linear example illustrates, with $g(q) = \bar{e} = 1$, $m = 2$, $(\alpha, \beta, \vartheta, c_f, K) = (20, 1, 1, 4, 4)$, and two incumbents with marginal cost 1.0. At the margin, each unit of increase in one incumbent's cost raises the fringe mass by about 0.4 to 0.5 firms. It lowers market output by 0.027 per unit under a rivalrous spillover ($\phi = -0.3$) and raises it by 0.008 per unit under a cooperative spillover ($\phi = 0.3$). A merger of the two incumbents raises the number of active firms from 3.45 to 4.57 under the rivalrous spillover and lowers market output by 0.09. Section 7 examines their magnitude in a calibration.

Power technology. With $g(q) = Aq^b$ and $0 < b < 1$, the wedge is

$$\Omega_{if} \;=\; A\big(q_f^{b-1} - b\, q_i^{b-1}\big) \;\geq\; A\,(1-b)\,q_f^{b-1} \;>\; 0 \qquad \text{whenever } q_i \geq q_f.$$

The exponent b therefore governs the size of the wedge. In a symmetric market, $1 - b$ is the elasticity of industry compliance with respect to the number of firms at a given market output; in the incumbent–fringe market, the exact wedge also depends on the outputs of the two kinds of firms. The power technology corresponds to the concave scaling specification estimated in Section 3.2, under the mapping from obligation counts to compliance effort and from sales to output. In the linear example with $g(q) = 0.5\,q^{0.45}$, $m = 3$, and the other parameters as above, the wedge is about

0.25. A unit increase in one incumbent's cost lowers market output by 0.010 at the margin under a rivalrous spillover ($\phi = -0.3$) and raises it by 0.002 under a cooperative spillover ($\phi = 0.3$). The size of the response need not be monotone in b: changing b changes both the entry–compliance wedge and the level of compliance of firms of different sizes, and in the example the response is largest at intermediate values of b.

6 Welfare

This section asks what the entry–compliance feedback implies for welfare. I measure welfare as market surplus: consumer surplus plus producers' operating profits, which are net of compliance costs, less fringe entry costs,

$$W \;=\; CS(Q) \;+\; \sum_{i=1}^{M_I} \pi_i \;+\; M_f\,(\pi_f - K), \qquad CS(Q) \;=\; \int_0^Q P(x)\,dx \;-\; P(Q)\,Q. \tag{12}$$

Two features of this criterion matter for interpretation. First, it treats the modelled compliance expenditures, $\kappa(e_j)\,s(E_{-j};\phi)$, as real resource costs. If a rivalrous spillover partly reflects higher prices paid to scarce consultants, engineers, or testing laboratories, part of that expenditure is a transfer to suppliers of compliance inputs rather than a resource cost, and the criterion overstates the social cost of compliance. Second, the model takes the regulatory regime as given and contains no environmental damages or benefits. The welfare analysis therefore concerns the market-side consequences of entry and of shocks to incumbents, not the full welfare effect of environmental regulation.

6.1 Consumer Surplus

Consumer surplus depends only on market output, and $CS'(Q) = -P'(Q)\,Q > 0$. The results of Section 5 therefore translate directly into statements about consumers.

Corollary 3 (Consumer surplus after shocks to incumbents)**.** *Under the conditions of Propositions 1–2, consider an increase in the marginal cost of an active incumbent.*

(i) Under proportional compliance, consumer surplus is unchanged.

(ii) With a general compliance technology, consumer surplus changes with the sign of $\Omega_{if} \cdot \phi$. In particular, when the entry–compliance wedge is positive, as under the conditions of Corollary 1, consumer surplus rises if $\phi > 0$, falls if $\phi < 0$, and is unchanged if $\phi = 0$.

Part (ii) extends to mergers under the conditions discussed in Section 5.3: when a merger raises the industry compliance needed to supply each level of market output, it lowers consumer surplus under a rivalrous spillover. Free entry then does not preserve consumer surplus after the merger, in contrast to the benchmark of Anderson et al. (2020).

6.2 The Compliance Externality of Entry

I now ask whether free entry produces the socially desirable fringe. Following Mankiw and Whinston (1986), I consider a planner who chooses the fringe mass but cannot control firms' output decisions once they have entered. For a given fringe mass M_f, the second stage determines outputs and industry compliance as in Section 5, and welfare becomes a function $W(M_f)$. At the free-entry equilibrium, the sign of dW/dM_f shows whether a small change in the fringe mass would raise welfare. A negative sign means that free entry produces too large a fringe, locally; a positive sign means it produces too small a fringe.

Proposition 3 (The welfare effect of entry)**.** *Suppose an active-fringe free-entry equilibrium exists and the second-stage equilibrium is locally differentiable with respect to M_f. At that equilibrium,*

$$\frac{dW}{dM_f} = \underbrace{\sum_j \omega_j \left[-P'(Q)\, q_j\right] \frac{dq_j}{dM_f}}_{\textit{business-stealing distortion}} + \underbrace{\sum_j \omega_j \left[-\kappa(e_j)\, s_E(E_{-j};\phi)\right] \frac{dE_{-j}}{dM_f}}_{\textit{compliance externality}}, \tag{13}$$

where the sums run over each incumbent, with weight $\omega_i = 1$, and the representative fringe firm, with weight $\omega_f = M_f$, and the derivatives are taken along the second-stage equilibrium. Consequently:

(i) If entry lowers each firm's output ($dq_j/dM_f \le 0$ for all j, with strict inequality for some j), the business-stealing distortion is negative.

(ii) If entry raises each firm's rivals' compliance ($dE_{-j}/dM_f > 0$ for all j), the compliance externality has the sign of ϕ.

(iii) Under both conditions, free entry produces a locally excessive fringe when $\phi \le 0$. When $\phi > 0$, the compliance externality offsets the business-stealing distortion, and free entry produces a locally insufficient fringe if and only if the positive compliance externality exceeds the business-stealing distortion in absolute value.

Under fixed obligations, $g(q) = \bar{e}$, and Assumptions 1–3, both conditions hold automatically.

The first term is the business-stealing mechanism of Mankiw and Whinston (1986), in the form it takes in this model. Each firm's price exceeds its private marginal cost, including the marginal cost of its own compliance, by the Cournot markup $-P'(Q)\,q_j$. When an additional fringe firm enters and causes rivals to produce less, society loses that markup on each unit forgone, a loss that the fringe firm does not take into account. Amir et al. (2014) generalize the comparison between free and socially optimal entry under business-stealing and business-enhancing competition.

The second term is the compliance externality introduced by the model. An additional fringe firm brings its own compliance, and changes its rivals' output and hence their compliance. Both change the compliance of every other firm's rivals, and through the spillover the cost of meeting every other firm's own obligations. The fringe firm does not take this into account either. Under a rivalrous spillover, entry raises other firms' compliance costs, which reinforces the business-stealing distortion: free entry is excessive on two counts. Under a cooperative spillover, entry lowers other firms' compliance costs, which works against the business-stealing distortion. The structure resembles Basak and Petrakis (2021), where a demand-side network externality can reverse the excess-entry result; here the externality operates through the costs of meeting compliance obligations and can take either sign.

The fixed-plus-proportional example illustrates how large the compliance externality can be. With $g(q) = 1 + 0.3q$, $m = 2$, two incumbents with marginal cost 1.0, and $(\alpha, \beta, \vartheta, c_f, K) = (20, 1, 1, 4, 4)$, a rivalrous spillover ($\phi = -0.3$) makes the compliance externality as large as the business-stealing distortion itself, -9.21 against -9.10, so that the social cost of the marginal fringe firm roughly doubles. A cooperative spillover ($\phi = 0.3$) makes the compliance externality positive, $+0.55$, and offsets about a tenth of the business-stealing distortion. Section 7 examines whether the reversal in part (iii) arises under the empirically disciplined compliance technology. In that calibration it does, but only in the most fragmented markets considered, or when strong cooperative spillovers are combined with high compliance costs, elastic demand, and a large incumbent sector.

7 A Quantitative Illustration

Sections 5 and 6 establish the signs of the effects. This section asks how large they are under parameter values disciplined, where possible, by my data and by published estimates. The data discipline the shape of the compliance technology and parts of the market structure. They do not identify the compliance spillover, and I am not aware of a published estimate of this compliance

spillover for environmental compliance; the closest evidence, from auditing (Duguay et al., 2020), concerns a different regulation and input market. The exercise is therefore an illustration: it shows the magnitudes the model implies at measured compliance costs, and how strong a spillover must be for them to matter.

7.1 Specification and Calibration

I use the linear example of Section 4.6 with two incumbents of equal cost and the power technology $g(q) = Aq^b$. I set the curvature of the stand-alone compliance cost to $m = 1/b$, so that the composite compliance cost $C(q) = \kappa(g(q))$ is linear in output. This choice makes $C''(q) = 0$, so that (R0) follows directly from the demand assumption, and it shuts down economies of scale in the cost of compliance, so that the results reflect only the channel the paper studies: how compliance aggregates across firms. I normalize price and market output to one at the baseline and choose units of compliance effort so that a fringe firm's compliance equals one. Marginal costs then follow from each firm's first-order condition, the scale of compliance costs from the compliance-cost share below, and the entry cost from the fringe's zero-profit condition. The shock is a 10 percent increase in both incumbents' marginal costs, a common illustrative scale; for shocks up to 10 percent, the responses are approximately proportional to the size of the shock. Table 2 lists the parameters and their sources, and Appendix B.5 documents the calibration targets.

Table 2. Calibration

Parameter	Baseline	Range	Source
Compliance exponent b	0.17	0.11–0.45	Count model, Appendix B.2
Compliance cost, share of revenue	1%	0.4–2%	U.S. Census Bureau (2008); sensitivity range
Number of active firms, $N = M_I + M_f$	20	5–100	ECHO facilities per state and six-digit industry
Incumbents' output share	0.4	0.2–0.6	Compustat and Bureau of Economic Analysis (BEA); Davis et al. (2006)
Demand elasticity	1.5	1–3	Ryan (2012); Fowlie et al. (2016)
Spillover elasticity η	–	−0.5 to 0.5	Not identified; sensitivity parameter

Note: The spillover elasticity η is the elasticity of a fringe firm's compliance cost with respect to its rivals' compliance at the baseline equilibrium; $\eta > 0$ corresponds to a rivalrous spillover ($\phi < 0$) and $\eta < 0$ to a cooperative one ($\phi > 0$).

Compliance technology. The model's g describes the compliance of active, regulated firms. The closest empirical object is the elasticity of expected permits among firms that hold at least one, which the count model estimates at 0.17 at the median firm (Appendix B.2). I use 0.17 as the

baseline and report results for 0.11, the value in the smallest third of firms, and 0.45, the exponent of the latent mean.

Compliance costs. The most recent comprehensive measure of environmental compliance costs in U.S. manufacturing is the 2005 Pollution Abatement Costs and Expenditures (PACE) survey. It puts pollution-abatement operating costs, including depreciation of abatement capital, at 0.44 percent of shipments across manufacturing, and at 0.8 to 1.1 percent in chemicals, paper, petroleum, and primary metals (U.S. Census Bureau, 2008). These heavily regulated industries dominate my matched sample. I therefore use 1 percent as an order-of-magnitude benchmark for heavily regulated manufacturing and report results from 0.4 to 2 percent. Operationally, I choose ϑ so that a fringe firm's baseline compliance cost equals 1 percent of its revenue; the survey ratios provide an order-of-magnitude benchmark for this normalization rather than a direct estimate for fringe firms. Broader evidence on regulatory compliance, which measures labor spent on complying with all regulation rather than environmental compliance costs, is consistent with costs of this order (Trebbi et al., 2023). Because the effects below scale approximately in proportion to this share, readers who prefer a different value can rescale them.

Market structure. In the ECHO data, the median regulated facility shares its state and six-digit industry with 23 facilities, and with 5 to 52 facilities in the five most heavily regulated manufacturing industries. Because many of these products are traded across states, a state-level count understates the number of competitors, while ownership of several plants by one firm overstates the number of independent firms. For comparison, the regional cement markets studied by Ryan (2012) contain 4.75 plants on average, and between 1 and 20. I refer to $N = M_I + M_f$ as the number of active firms, although the fringe is represented by a continuous mass. I use $N = 20$ as an illustrative baseline disciplined by these facility counts, not as a direct estimate of the number of firms, and report results for 5 and 100. For the incumbents' output share, sales of listed firms matched to regulated facilities equal 61 to 70 percent of manufacturing gross output in the BEA accounts, an upper bound because Compustat sales include foreign sales and other lines of business, while listed firms account for about a third of private employment (Davis et al., 2006). I use 0.4.

Demand. Structural models of the U.S. cement industry that, like mine, assume Cournot competition in a homogeneous good estimate market demand elasticities of about 0.9 to 2.0 (Fowlie et al., 2016) and, in Ryan (2012), a preferred estimate of 2.96 from a specification with market fixed

effects, with specifications without fixed effects ranging from about 2.0 to 3.2. I use 1.5 and report results for 1 and 3. Models with spatial differentiation estimate much lower aggregate elasticities (Miller and Osborne, 2014), but they are not directly comparable, since market power there arises from distance rather than from Cournot conduct.

The spillover. The empirical literature on enforcement spillovers measures how one plant's sanction changes other plants' violations (Shimshack and Ward, 2005). It concerns compliance behavior rather than compliance costs, and cannot be mapped into η. Evidence on compliance costs from auditing (Duguay et al., 2020) measures fee changes across types of clients rather than the elasticity in my model, so it does not discipline η either. I therefore treat η as a sensitivity parameter, centered on the neutral case $\eta = 0$, and report the value needed for the effects to reach a given size. Appendix D gives η an interpretation: under a capacity-constrained supply of compliance inputs, it is the inverse of their supply elasticity, so that $\eta = 0.3$ corresponds to a supply elasticity of about 3.

7.2 Shocks to Incumbents

Table 3 reports the long-run effects of the cost shock. With private compliance, market output and price are unchanged, as Proposition 2 implies when $\phi = 0$: the shock is absorbed entirely by entry. With a spillover, entry still responds strongly. The fringe grows by about 43 percent and industry compliance by about 35 percent, whatever the sign of the spillover. With a rivalrous spillover of $\eta = 0.3$, long-run output falls by 0.17 percent and price rises by 0.12 percent; with a cooperative spillover of the same size, output rises by 0.16 percent.

Table 3. Long-run effects of a 10 percent increase in incumbents' marginal costs

	Cost shock				Merger	Welfare
η	Output	Price	Fringe mass	Industry compliance	Output	Externality / business stealing
−0.5	+0.26%	−0.17%	+43%	+36%	+0.17%	−0.21
−0.3	+0.16%	−0.11%	+43%	+35%	+0.10%	−0.12
0	0	0	+43%	+35%	0	0
0.3	−0.17%	+0.12%	+43%	+35%	−0.11%	+0.12
0.5	−0.30%	+0.20%	+42%	+35%	−0.19%	+0.20

Note: Baseline calibration of Table 2. "Merger" merges the two incumbents into one firm with the same marginal cost. The last column is the ratio of the compliance externality to the business-stealing distortion in Proposition 3, evaluated at the baseline equilibrium; a negative ratio means the externality offsets business stealing. All equilibria satisfy the stability condition of Assumption 3.

A simple relationship explains these magnitudes. The fringe's zero-profit condition equates price

with a fringe firm's average cost. A shock to incumbents can therefore change the long-run price only by changing the fringe's average cost, and in the model the only channel through which it can do so is the compliance spillover. A change in industry compliance of $\Delta E/E$ changes a fringe firm's compliance cost by approximately $\eta\,\Delta E/E$, which changes its average cost by the compliance-cost share times that amount. Market output then moves by the demand elasticity times the change in price:

$$\frac{\Delta Q}{Q} \;\approx\; -\varepsilon_D \times (\text{compliance-cost share}) \times \eta \times \frac{\Delta E}{E}. \tag{14}$$

At the baseline, with $\eta = 0.3$, this gives $-1.5 \times 0.01 \times 0.3 \times 0.35 \approx -0.16$ percent, against the exact -0.17 percent. Across the sensitivity analysis, the approximation is within about 12 percent of the exact result when the market has at least 20 firms. The four factors show what sets the size of the effect: entry raises industry compliance substantially, and the compliance-cost share, the demand elasticity, and the spillover elasticity determine how much of that change reaches long-run output.

The same relationship indicates how strong a spillover must be for the effect to matter. At the baseline, a rivalrous spillover of $\eta = 0.18$ reduces long-run output by 0.1 percent. Extending the calculation beyond the reported sensitivity range, a rivalrous spillover of $\eta = 0.80$ reduces it by 0.5 percent. An effect of half a percent of output therefore requires that a firm's compliance cost rise nearly one-for-one with its rivals' compliance, or a compliance-cost share well above the baseline.

7.3 Mergers and Entry

A merger of the two incumbents moves long-run output in the direction of the spillover, by 0.11 percent under a rivalrous spillover of $\eta = 0.3$ (Table 3). Under private compliance, as in Anderson et al. (2020), it leaves long-run output unchanged.

The compliance externality of entry is 12 to 20 percent of the business-stealing distortion at the baseline for $|\eta|$ between 0.3 and 0.5. Under a rivalrous spillover it adds to the business-stealing distortion, so that free entry is more excessive than business stealing alone would imply. Under a cooperative spillover it offsets part of that distortion. Within the sensitivity ranges, the offset is large enough to make entry locally insufficient only in the most fragmented markets considered, with 100 active firms, cooperative spillovers, and compliance costs of at least 1 percent of revenue.

7.4 Sensitivity

Table 4 varies one parameter at a time. The effects scale with the compliance-cost share and the demand elasticity, as (14) predicts, and increase with the incumbents' output share, because a larger incumbent sector draws in more entry after the shock. In this calibration, the effects are larger at the lower values of b considered: moving from $b = 0.45$ to $b = 0.11$ raises the response of industry compliance from 27 to 36 percent. Across all variants, long-run output changes by between 0.05 and 0.45 percent for $|\eta| = 0.3$. Every equilibrium is stable, and an active fringe exists in every case I examine, including spillover elasticities up to 1.5 in absolute value (Figure 5).

Table 4. Sensitivity of the long-run output effect

	Output		Industry	Approximation	Externality /
Variant	$\eta = 0.3$	$\eta = -0.3$	compliance	/ exact	business stealing
Baseline	−0.17%	+0.16%	+35%	0.90	0.12
$b = 0.11$	−0.18%	+0.16%	+36%	0.90	0.13
$b = 0.45$	−0.13%	+0.12%	+27%	0.92	0.08
Compliance 0.4% of revenue	−0.07%	+0.06%	+35%	0.90	0.05
Compliance 2% of revenue	−0.34%	+0.32%	+34%	0.90	0.24
$N = 5$	−0.11%	+0.11%	+19%	0.77	0.03
$N = 100$	−0.20%	+0.18%	+40%	0.93	0.61
Demand elasticity 1	−0.07%	+0.07%	+22%	0.92	0.08
Demand elasticity 3	−0.45%	+0.40%	+44%	0.89	0.24
Incumbents' share 0.2	−0.05%	+0.05%	+11%	0.93	0.09
Incumbents' share 0.6	−0.25%	+0.22%	+49%	0.88	0.18

Note: Each row changes one parameter from the baseline in Table 2. "Output" is the long-run change in market output after a 10 percent increase in incumbents' marginal costs. "Industry compliance" is its change for $\eta = 0.3$. "Approximation / exact" is the ratio of (14) to the exact output change for $\eta = 0.3$. The last column is evaluated at $\eta = 0.3$.

Figure 5 varies jointly the two parameters that set the size of the output effect: the compliance-cost share, which the data measure only for 2005, and the spillover elasticity, which no study estimates. Within the range of compliance-cost shares in the 2005 survey, long-run output falls by up to about 0.7 percent at $\eta = 1$ and up to about 1.2 percent at $\eta = 1.5$. Under capacity-constrained supply (Appendix D), $\eta = 1.5$ corresponds to a supply elasticity of compliance inputs of about 0.67, an inelastic short-run supply such as the supply of auditors after the Sarbanes–Oxley Act (Duguay et al., 2020). Larger effects require both a larger compliance-cost share and a stronger spillover. A 1 percent fall in long-run output requires a compliance-cost share of about 6 percent at $\eta = 0.3$, about 3.6 percent at $\eta = 0.5$, and about 2 percent at $\eta = 0.8$; at the upper end of the 2005 range, it requires $\eta \approx 1.3$. The channel therefore matters most where compliance is a larger share of costs than in 2005 manufacturing and where compliance inputs are scarce.

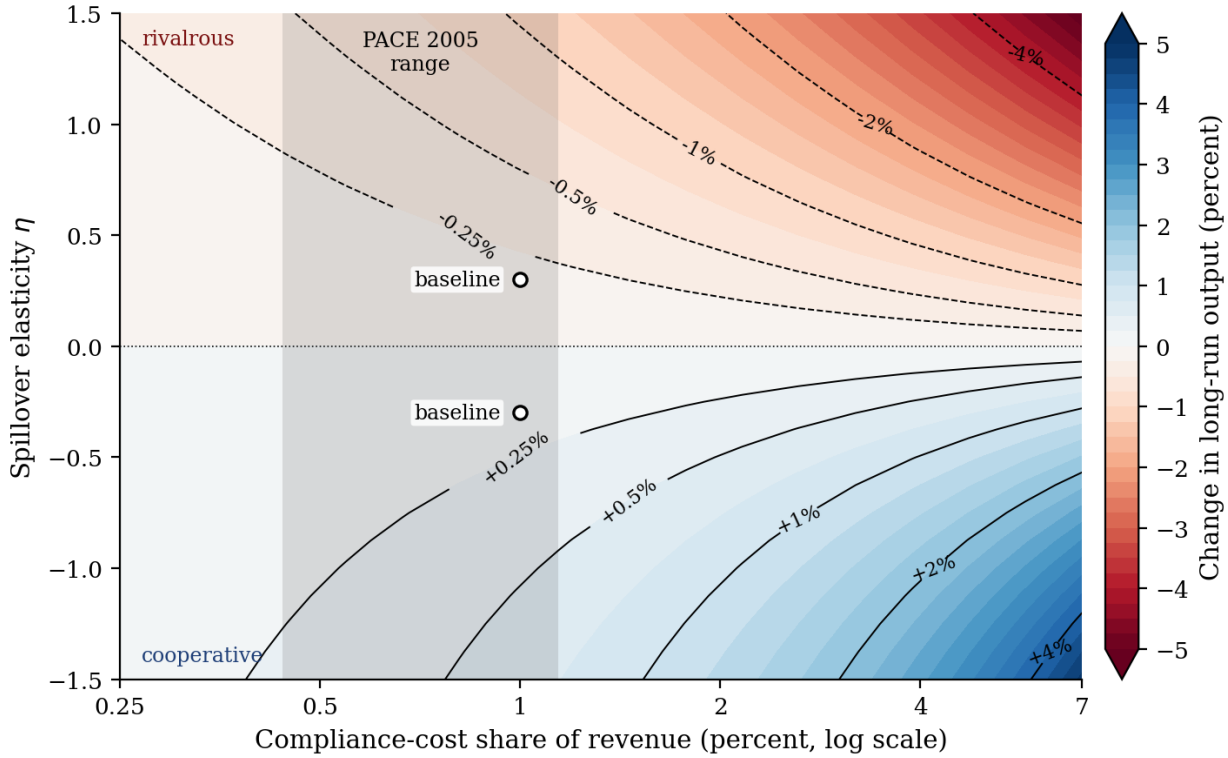


Figure 5. Long-run output effect by compliance-cost share and spillover elasticity

Note: Each point is the change in long-run market output after a 10 percent increase in incumbents' marginal costs, with all other parameters at the baseline of Table 2. Positive η is a rivalrous spillover and negative η a cooperative one; the figure extends the range of Table 4 to $|\eta| = 1.5$. For $\eta \geq 2$, the fringe has no interior equilibrium at the highest compliance-cost shares, so the figure stops at 1.5. The horizontal axis is on a logarithmic scale. The shaded band is the range of pollution-abatement operating costs as a share of shipments in the 2005 PACE survey, from all manufacturing (0.44 percent) to primary metals (1.14 percent) (U.S. Census Bureau, 2008). Circles mark the baseline, a compliance-cost share of 1 percent with $\eta = \pm 0.3$. Contours are computed from 500 equilibria; every equilibrium has an active fringe and satisfies the stability condition of Assumption 3.

Taken together, the exercise shows that the entry–compliance channel moves long-run output away from the zero that free-entry neutrality implies, with a sign set by the spillover. Within the range of measured compliance costs, long-run output falls by about 0.2 percent at the baseline and by up to about 1.2 percent when the supply of compliance inputs is inelastic. The effect grows approximately in proportion to compliance costs, so it is larger in industries where compliance is a larger share of costs. The main open question is the size of the spillover itself, which my data do not identify.

8 Conclusion

This paper asks when free entry preserves consumer surplus after shocks to large incumbent firms. Using EPA compliance records linked to firm sales from Compustat, I document that environmental compliance obligations grow far less than proportionally with firm size, and that listed firms hold a small share of regulated facilities. In a model with incumbents and a fringe of smaller firms that enter freely, I show that an incumbent cost shock changes long-run output if and only if the entry–compliance wedge is nonzero and compliance spills over across firms. Free-entry neutrality therefore holds when compliance is proportional to output, whatever the spillover, and for marginal cost shocks when compliance is private. When both conditions hold, the direction of the output response follows the sign of the spillover, and entry creates a compliance externality that can reinforce or reverse the classic excess-entry result. Within the range of measured compliance costs, long-run output falls by about 0.2 percent at the baseline and by up to about 1.2 percent when the supply of compliance inputs is inelastic, and the effect grows with the compliance-cost share.

The results carry two implications. For regulatory design, the way compliance obligations scale with output affects how markets respond to shocks to their largest firms. For merger enforcement, the results suggest that whether entry offsets a merger depends not only on how much entry occurs but also on how entry changes compliance costs. The main open question is the size of the compliance spillover. Appendix D suggests one way to measure it: through the supply elasticity of compliance inputs, such as testing, monitoring, and auditing services.

References

Akey, P. and Appel, I. (2021). The limits of limited liability: Evidence from industrial pollution. *Journal of Finance*, 76(1):5–55.

Amir, R., De Castro, L., and Koutsougeras, L. (2014). Free entry versus socially optimal entry. *Journal of Economic Theory*, 154:112–125.

Anderson, S. P., Erkal, N., and Piccinin, D. (2020). Aggregative games and oligopoly theory: Short-run and long-run analysis. *RAND Journal of Economics*, 51(2):470–495.

Anderson, S. P. and Peitz, M. (2020). Media see-saws: Winners and losers in platform markets. *Journal of Economic Theory*, 186:104990.

Bartel, A. P. and Thomas, L. G. (1987). Predation through regulation: The wage and profit effects of the occupational safety and health administration and the environmental protection agency. *Journal of Law and Economics*, 30(2):239–264.

Basak, D. and Petrakis, E. (2021). Social efficiency of entry: Implications of network externalities. *Journal of Economics & Management Strategy*, 30(4):820–829.

Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy*, 76(2):169–217.

Blundell, W., Gowrisankaran, G., and Langer, A. (2020). Escalation of scrutiny: The gains from dynamic enforcement of environmental regulations. *American Economic Review*, 110(8):2558–2585.

California State Water Resources Control Board (2014). National pollutant discharge elimination system general permit for storm water discharges associated with industrial activities, Order 2014-0057-DWQ, and fact sheet. Section XIV (Compliance Groups); Fact Sheet, pp. 66–67.

Coates, J. C. and Srinivasan, S. (2014). SOX after ten years: A multidisciplinary review. *Accounting Horizons*, 28(3):627–671.

Code of Federal Regulations (2024a). 40 CFR part 122 — EPA administered permit programs: The national pollutant discharge elimination system. Title 40 – Protection of Environment; Section 122.46 (duration of permits) and Section 122.48 (monitoring and reporting).

Code of Federal Regulations (2024b). 40 CFR section 123.45 — noncompliance and program reporting by the Director. Title 40 – Protection of Environment, Part 123 – State Program Requirements.

Dasgupta, S., Huynh, T. D., and Xia, Y. (2023). Joining forces: The spillover effects of EPA enforcement actions and the role of socially responsible investors. *Review of Financial Studies*, 36(9):3781–3824.

d'Aspremont, C. and Jacquemin, A. (1988). Cooperative and noncooperative R&D in duopoly with spillovers. *American Economic Review*, 78(5):1133–1137.

Davis, S. J., Haltiwanger, J., Jarmin, R., and Miranda, J. (2006). Volatility and dispersion in business growth rates: Publicly traded versus privately held firms. *NBER Macroeconomics Annual*, 21:107–179.

DiMaggio, P. J. and Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, 48(2):147–160.

Duguay, R., Minnis, M., and Sutherland, A. (2020). Regulatory spillovers in common audit markets. *Management Science*, 66(8):3389–3411.

Erkal, N. and Piccinin, D. (2010). Cooperative R&D under uncertainty with free entry. *International Journal of Industrial Organization*, 28(1):74–85.

European Parliament and Council of the European Union (2006). Regulation (EC) no 1907/2006 concerning the registration, evaluation, authorisation and restriction of chemicals (REACH). Official Journal of the European Union, L 396.

Evans, M. F., Gilpatric, S. M., and Shimshack, J. P. (2018). Enforcement spillovers: Lessons from strategic interactions in regulation and product markets. *Journal of Law and Economics*, 61(4):739–769.

Fowlie, M., Reguant, M., and Ryan, S. P. (2016). Market-based emissions regulation and industry dynamics. *Journal of Political Economy*, 124(1):249–302.

Garicano, L., Lelarge, C., and Van Reenen, J. (2016). Firm size distortions and the productivity distribution: Evidence from france. *American Economic Review*, 106(11):3439–3479.

Gray, W. B. and Shimshack, J. P. (2011). The effectiveness of environmental monitoring and enforcement: A review of the empirical evidence. *Review of Environmental Economics and Policy*, 5(1):3–24.

Häckner, J. and Herzing, M. (2017). The effectiveness of environmental inspections in oligopolistic markets. *Resource and Energy Economics*, 48:83–97.

Häckner, J. and Herzing, M. (2020). The equilibrium compliance rate among regulated firms. *International Review of Law and Economics*, 63:105911.

Helland, E. (1998). The enforcement of pollution control laws: Inspections, violations, and self-reporting. *The Review of Economics and Statistics*, 80(1):141–153.

Khazabi, M. and Quyen, N. (2012). R&D spillovers, innovation, and entry. MPRA Paper No. 39460, University of Ottawa.

Laffont, J.-J. and Tirole, J. (1991). The politics of government decision-making: A theory of regulatory capture. *The Quarterly Journal of Economics*, 106(4):1089–1127.

Li, Z. and Lyon, T. P. (2026). Learning or leaking? enforcement spillover effects within and beyond the firm. *Business & Society*, 65(3):544–578.

Liu, W., Luan, J., and Xie, Q. (2026). Environmental regulation, abatement strategies, and industrial concentration: Theory and evidence from China. *Energy Economics*, 158:109319.

Liu, X., Kirwan, B., and Martens, A. (2018). Regulatory compliance, information disclosure and peer effects: Evidence from the Mexican gasoline market. *Journal of Regulatory Economics*, 54(1):53–80.

Mankiw, N. G. and Whinston, M. D. (1986). Free entry and social inefficiency. *RAND Journal of Economics*, 17(1):48–58.

Miller, N. H. and Osborne, M. (2014). Spatial differentiation and price discrimination in the cement industry: Evidence from a structural model. *RAND Journal of Economics*, 45(2):221–247.

Novshek, W. (1985). On the existence of Cournot equilibrium. *Review of Economic Studies*, 52(1):85–98.

Pashigian, B. P. (1984). The effect of environmental regulation on optimal plant size and factor shares. *Journal of Law and Economics*, 27(1):1–28.

Peitz, M. and Sato, S. (2026). Asymmetric platform oligopoly. *RAND Journal of Economics*, 57(1):60–77.

Potoski, M. and Prakash, A. (2005). Covenants with weak swords: Iso 14001 and facilities' environmental performance. *Journal of Policy Analysis and Management*, 24(4):745–769.

Ryan, S. P. (2012). The costs of environmental regulation in a concentrated industry. *Econometrica*, 80(3):1019–1061.

Salop, S. C. and Scheffman, D. T. (1983). Raising rivals' costs. *American Economic Review*, 73(2):267–271.

Seade, J. (1980). On the effects of entry. *Econometrica*, 48(2):479–489.

Sharma, V. (2026). Price floors, market structure, and Bertrand equilibrium under subadditive costs. Working paper, The Ohio State University.

Shimshack, J. P. and Ward, M. B. (2005). Regulator reputation, enforcement, and environmental compliance. *Journal of Environmental Economics and Management*, 50(3):519–540.

Singla, S. (2023). Regulatory costs and market power. SSRN Working Paper 4368609.

Stafford, S. L. (2002). The effect of punishment on firm compliance with hazardous waste regulations. *Journal of Environmental Economics and Management*, 44(2):290–308.

Stigler, G. J. (1971). The theory of economic regulation. *The Bell Journal of Economics and Management Science*, 2(1):3–21.

Tirole, J. (1996). A theory of collective reputations (with applications to the persistence of corruption and to firm quality). *Review of Economic Studies*, 63(1):1–22.

Trebbi, F., Zhang, M. B., and Simkovic, M. (2023). The cost of regulatory compliance in the United States. NBER Working Paper w30691, National Bureau of Economic Research.

U.S. Census Bureau (2008). Pollution abatement costs and expenditures: 2005. Technical Report MA200(05), U.S. Government Printing Office, Washington, DC.

U.S. Department of Justice and Federal Trade Commission (2023). Merger guidelines. Section 3.2, Entry and Repositioning.

U.S. Environmental Protection Agency (1994). Pesticide registration notice 94-9: Announcing the formation of two industry-wide task forces: Agricultural reentry task force and outdoor residential exposure task force.

U.S. Environmental Protection Agency (2024a). ECHO quick start guide. Enforcement and Compliance History Online.

U.S. Environmental Protection Agency (2024b). Enforcement and Compliance History Online (ECHO). Accessed 2025.

U.S. Environmental Protection Agency (2024c). ICIS-NPDES download summary and data element dictionary. Enforcement and Compliance History Online (ECHO).

U.S. Environmental Protection Agency (2024d). National enforcement and compliance initiatives.

U.S. Government Accountability Office (2025). Persistent chemicals: DOD needs to provide congress more information on costs associated with addressing PFAS. Technical Report GAO-25-107401, U.S. Government Accountability Office, Washington, DC.

Appendices

A Data Construction

Throughout the paper I distinguish between facilities (physical sites regulated under EPA programs and tracked in ECHO) and firms (legal entities reporting financial information in Compustat). A firm owns one or more facilities. I aggregate facility-level records to the firm-year before merging with firm financials.

A.1 Data Sources

EPA compliance records. I draw facility-level records from the U.S. EPA's Enforcement and Compliance History Online (ECHO) system. ECHO consolidates records from program-specific databases in a common framework keyed on the Facility Registry Service identifier `REGISTRY_ID`. I use the following files from the ECHO download distribution.

1. `ECHO_EXPORTER.csv`: the facility master file, with the universal identifier `REGISTRY_ID`, facility name, location, NAICS codes, the active-status flag `FAC_ACTIVE_FLAG`, and the federal-facility flag `FAC_FEDERAL_FLG`.

2. `FRS_PROGRAM_LINKS.csv`: the bridge table that maps `REGISTRY_ID` to program-specific facility identifiers, which lets me link a facility's air and water records.

3. `ICIS-AIR_FACILITIES.csv`: the Clean Air Act facility list, including each facility's operating status (`AIR_OPERATING_STATUS_CODE`).

4. `ICIS-AIR_PROGRAM_SUBPARTS.csv`: the Clean Air Act programs and subparts that apply to each facility.

5. `ICIS_FACILITIES.csv` (NPDES download): the Clean Water Act facility list, including the facility type code that identifies ownership (for example, privately owned, corporation, municipal, state, or federal).

6. `ICIS_PERMITS.csv` (NPDES download): Clean Water Act discharge permits, with their effective, expiration, and termination dates.

7. `NPDES_QNCR_HISTORY.csv`: the Quarterly Noncompliance Report history, which records each permit's compliance status by quarter.

Table A.1 summarizes the ECHO facility universe.

Table A.1. ECHO facility universe

Statistic	Count
Total facilities in `ECHO_EXPORTER`	3,141,165
Unique `REGISTRY_ID` values	3,095,510
Facilities with a non-blank NAICS code	1,527,789
Facilities flagged active (`FAC_ACTIVE_FLAG = Y`)	1,351,898
Facilities flagged major NPDES dischargers	24,911

Source: U.S. EPA ECHO Exporter file.
Note: Counts cover all facilities in the ECHO universe regardless of activity status.

Linking air and water records. The `FRS_PROGRAM_LINKS` table links program-specific facility identifiers to the universal `REGISTRY_ID`, so that I can combine a facility's Clean Air Act and Clean Water Act records. Table A.2 shows its coverage across EPA programs. I use the links for the NPDES, AIR, and ICIS programs, which cover Clean Water Act and Clean Air Act records; the table describes each program.

Table A.2. FRS program coverage

Program	Description	Records	Unique facilities
RCRAINFO	Resource Conservation and Recovery Act (hazardous waste)	1,578,620	1,476,646
NPDES	National Pollutant Discharge Elimination System (water dischargers)	1,162,783	994,147
SFDW	Safe Drinking Water Information System (public water systems)	676,905	452,826
AIR	Clean Air Act stationary sources (ICIS-Air)	279,292	265,828
ICIS	Integrated Compliance Information System (legacy NPDES records)	229,937	207,791
EIS	Emissions Inventory System	211,197	199,026
TRIS	Toxics Release Inventory	74,769	73,731
Other	Six smaller EPA program databases	115,510	104,366
Total		**4,329,013**	

Source: U.S. EPA Facility Registry Service program links.
Note: Each record links a program-specific identifier to a universal `REGISTRY_ID`; a facility can appear under several programs. I use the NPDES, AIR, and ICIS links.

Air and water facility universes. Table A.3 summarizes the two program-level facility files. The operating-status codes in the Clean Air Act file determine which facilities count as operating

in Appendix A.3; the facility-type codes in the Clean Water Act file identify ownership for the Fact 2 sample in Appendix A.4.

Table A.3. Clean Air Act and Clean Water Act facility files

Statistic	Count
Clean Air Act facilities (ICIS-AIR_FACILITIES)	
Total facilities	279,124
Operating	184,001
Permanently closed	77,044
Temporarily closed	4,017
Other or missing status	14,062
Clean Water Act permit records (ICIS_FACILITIES)	
Total permit records	1,187,641
Unique facilities	979,486
Privately owned facility (POF)	422,320
Corporation (COR)	214,826
Municipal or water district	65,423
State government	19,396
Federal facility	6,770
Other, unknown, or blank type	458,906

Source: U.S. EPA ICIS-Air and ICIS-NPDES facility downloads.
Note: Operating status reflects each facility's current status at the time of the data release; the files carry no closure dates. Clean Water Act records are one per permit, so a facility can appear more than once. "Municipal or water district" combines municipal and water-district (MWD) and municipality (CTG) codes.

Firm financials. From WRDS Compustat Fundamentals Annual I take sales (`sale`), employment (`emp`), and the identifiers needed for the match (`gvkey`, `conm`, `naics`). Table A.4 traces the sample selection.

Table A.4. Compustat sample selection

Step	Firm-years	Unique firms
Raw Compustat Fundamentals Annual	156,631	20,510
U.S.-incorporated (`loc = USA`)	133,586	17,492
Sample window (`fyear` 2009–2025)	133,586	17,492
Positive sales (`sale` > 0)	81,311	9,532
Final Compustat base	**81,311**	**9,532**

Source: WRDS Compustat Fundamentals Annual, standard industrial universe (`datafmt = STD`, `indfmt = INDL`, `consol = C`).
Note: The positive-sales filter removes mostly financial-services records that lack an industrial sales construct.

A.2 Facility–Firm Match

I match EPA facilities to Compustat firms by fuzzy name comparison. I clean facility names (from `ECHO_EXPORTER`, with `ICIS-AIR_FACILITIES` as a fallback) and Compustat firm names (`conm`) by converting to upper case, replacing punctuation with spaces, and removing common corporate suffixes (Inc., Corp., LLC, LP, Ltd., Company, Co., Holdings, Group, International, PLC, and similar). I drop facility records with the placeholder name "UNKNOWN" (9,852 records). I block candidate pairs by three-digit NAICS industry and the first letter of the cleaned name, which reduces the number of pairwise comparisons by roughly three orders of magnitude, compute the Jaro–Winkler similarity between every facility name and every firm name within a block, and keep the best firm match for each facility.

I keep matches with similarity of at least 0.90. Blocking on three-digit NAICS industry has a cost of its own: it cannot match a facility whose industry differs from its parent's principal industry in Compustat, which can happen for diversified firms. I prefer the conservative side of this trade-off: assigning a facility to the wrong firm would contaminate that firm's obligations directly, whereas missing a true match only reduces the sample. Blocking by three-digit NAICS protects against the most common false positive, facilities with similar names in unrelated industries. Table A.5 reports the outcome.

Table A.5. Facility–firm match

Statistic	Value
Facilities with a valid three-digit NAICS code and name	418,608
Compustat firms with a valid name and three-digit NAICS code	9,434
Facilities matched (Jaro–Winkler ≥ 0.90)	6,109
of which exact matches (similarity ≥ 0.999)	1,759
Listed firms matched (`gvkey`)	1,121
Mean Jaro–Winkler similarity (matched pairs)	0.945
Median similarity	0.929

Note: I keep the best Compustat match for each facility. Because the match is conservative, it misses facilities that listed firms own under different names, such as some subsidiaries.

A.3 Measures of Compliance Obligations

For each matched firm and year, I build three counts of compliance obligations by aggregating over the firm's matched facilities.

Water permits. I link each matched facility to its Clean Water Act discharge permits in `ICIS_PERMITS.csv`. A permit can appear in several versions as it is reissued or modified; I collapse the versions of each permit into a single active span, from its earliest effective date to its latest expiration date, ending earlier if the permit is terminated or retired. A permit counts as active in year t if its span covers year t. The water-permit count for firm i in year t is the number of active permits at the firm's matched facilities. The measure covers 3,137 permits at matched facilities; Table B.1 reports the estimation samples.

Air regulations. For each matched facility, I count the Clean Air Act program and subpart records in `ICIS-AIR_PROGRAM_SUBPARTS.csv`, such as specific New Source Performance Standards and National Emission Standards for Hazardous Air Pollutants. These records describe the regulations that currently apply and carry no dates, so the count does not vary over time for a given facility. The air-regulation count for firm i in year t is the sum over the firm's matched facilities that are currently operating.

Regulated facilities. The regulated-facility count for firm i in year t is the number of the firm's matched facilities that are currently operating.

Operating status. The EPA files record each facility's current status but no closure dates. I treat a facility as operating if its Clean Air Act status code is `OPR` (operating) or, for facilities without an air record, if `FAC_ACTIVE_FLAG` equals `Y`. Of the 6,109 matched facilities, 4,425 (72.4 percent) are operating under this rule. Because the rule relies on current status, it also removes facilities that operated earlier in the sample and later closed. Table B.1 reports the elasticities with and without this restriction.

A.4 Sample for Fact 2

Fact 2 describes the population of regulated facilities, not only matched ones. I start from the 1,465,685 ECHO facilities with a usable three-digit NAICS code and apply four restrictions.

1. *Ownership type.* I keep facilities whose Clean Water Act facility type is privately owned (`POF`) or corporation (`COR`). This removes municipal, state, federal, tribal, and other government facilities. Ownership type is recorded only for facilities with Clean Water Act permits, so the restriction also limits the sample to that universe.

2. *Federal facilities.* I also drop facilities that ECHO flags as federal, by `FAC_FEDERAL_FLG` or a federal agency code.

3. *Active status.* I keep facilities with `FAC_ACTIVE_FLAG` equal to `Y`. The flag is blank rather than negative for most facilities, so this restriction is conservative.

4. *Industries.* In the main results I exclude construction (NAICS 236), where most permits cover temporary storm-water discharges at construction sites, and waste management (NAICS 562), which contains many municipal operations.

Table A.6 reports the facilities removed at each step. Table A.7 shows that the pooled figures change little across these choices, and Table A.8 reports the breakdown by industry.

Missing ownership type. The ownership-type field is blank or coded as unknown in 36 percent of Clean Water Act permit records, and the share varies widely across states, from about 1 percent to over 90 percent. Because states also differ in how completely they report permits and active status, I do not use the sample to compare states. Missing ownership type does not drive Fact 2. Of the facilities whose type is blank or unknown, 17,548 are active, non-federal, and outside construction and waste management; listed firms hold 1.37 percent of them. Adding them to the sample lowers the listed-firm share from 2.01 to 1.89 percent (Table A.7).

Table A.6. Sample restrictions for Fact 2

Step	Facilities
ECHO facilities with a usable three-digit NAICS code	1,465,685
No Clean Water Act facility-type record	−1,176,703
Facility type not privately owned or corporation	−83,768
Privately owned or corporation, but federal	−366
Privately owned or corporation, not flagged active	−81,339
Active, privately owned and corporate facilities	123,509
Excluding construction (236) and waste management (562)	79,276

Note: Starting sample: ECHO facilities with a usable three-digit NAICS code. Most exclusions occur because a facility holds no Clean Water Act permit and therefore has no ownership-type record.

A.5 Noncompliance Records

Section 3.4 reports how often Clean Water Act permits are out of compliance. The figure comes from the Quarterly Noncompliance Report, which assigns each permit-quarter a status code. I

Table A.7. Share of regulated facilities held by listed firms: alternative samples

Status	Industries	Facilities	Listed-firm facilities	Share (%)	Top 10% share (%)
Active only	Excluding 236, 562	79,276	1,594	2.01	54.5
Active only	All	123,509	1,695	1.37	55.3
All statuses	Excluding 236, 562	97,947	1,748	1.78	54.1
All statuses	All	204,848	1,860	0.91	54.8
Including facilities with blank or unknown ownership type					
Active only	Excluding 236, 562	96,824	1,834	1.89	–

Note: Privately owned and corporate facilities holding Clean Water Act permits, excluding federal facilities. "Share" is the percentage of facilities held by listed firms in my sample. "Top 10% share" is the percentage of listed-firm facilities held by the ten percent of listed firms with the most facilities. The first row is the main sample in Section 3.3. Construction (NAICS 236) contains many facilities and almost no listed-firm facilities, which is why including it lowers the pooled share. The last row adds 17,548 active facilities whose ownership type is blank or coded as unknown (`UNK`); treating other non-identifying codes as unknown gives a share of 1.85 percent.

count a permit-quarter as noncompliant if its code indicates significant noncompliance (Category I: codes D, E, S, T, and X) or reportable noncompliance (Category II: code N), following EPA's classification under Section 123.45 of Title 40 of the Code of Federal Regulations (Code of Federal Regulations, 2024b; U.S. Environmental Protection Agency, 2024c). Table A.9 describes these codes. Between 2009 and 2025, 1,520,375 of 6,066,926 permit-quarters (25.1 percent) carry one of them. These are the only compliance-status records I use; the paper's measures of obligations do not depend on them.

B Compliance Obligations and Firm Size

This appendix describes how I estimate the relationship between compliance obligations and firm size, first with log-linear regressions and then with a count model, how I construct Figures 1 and 2, and how the estimates relate to the model.

B.1 Log-Linear Elasticities

Specification. Let O_{it} denote a count of firm i's compliance obligations in year t (water permits, regulated facilities, or air regulations, as defined in Appendix A.3) and S_{it} its sales. I estimate

$$\ln O_{it} \;=\; b \ln S_{it} \;+\; \lambda_t \;+\; \varepsilon_{it}, \tag{15}$$

Table A.8. Regulated facilities and listed-firm ownership by industry

			Listed-firm facilities			Facilities per listed firm		
NAICS	Industry	Facilities	Number	Share (%)	Listed firms	Median	P90	Max
221	Utilities	7,973	73	0.92	44	1.0	2.0	10
212	Mining (except oil and gas)	6,674	92	1.38	10	1.0	15.7	67
327	Nonmetallic mineral products	5,243	9	0.17	5	2.0	2.6	3
332	Fabricated metal products	4,667	23	0.49	13	1.0	2.6	9
237	Heavy and civil engineering	4,569	0	0.00	0	–	–	–
325	Chemicals	4,382	196	4.47	94	1.0	4.0	18
311	Food manufacturing	3,376	145	4.30	34	2.0	10.7	17
423	Wholesale trade, durable goods	2,717	9	0.33	4	1.5	4.1	5
336	Transportation equipment	2,562	74	2.89	38	1.0	4.0	8
321	Wood products	2,008	27	1.34	10	2.0	7.0	7
324	Petroleum and coal products	1,950	8	0.41	6	1.0	2.0	3
326	Plastics and rubber products	1,914	45	2.35	16	2.0	6.5	8
484	Truck transportation	1,853	24	1.30	4	4.5	11.3	14
333	Machinery	1,847	50	2.71	33	1.0	2.8	7
493	Warehousing and storage	1,819	0	0.00	0	–	–	–
424	Wholesale trade, nondurable goods	1,812	21	1.16	9	2.0	3.6	6
531	Real estate	1,753	2	0.11	2	1.0	1.0	1
331	Primary metals	1,646	38	2.31	18	2.0	3.3	4
488	Support activities for transport.	1,298	0	0.00	0	–	–	–
112	Animal production	1,211	2	0.17	1	2.0	2.0	2
238	Specialty trade contractors	1,070	0	0.00	0	–	–	–
334	Computer and electronic products	1,022	92	9.00	62	1.0	2.0	5
811	Repair and maintenance	906	0	0.00	0	–	–	–
322	Paper	902	123	13.64	17	2.0	17.6	39
339	Miscellaneous manufacturing	814	17	2.09	15	1.0	1.6	2
492	Couriers and messengers	771	298	38.65	2	149.0	190.6	201
441	Motor vehicle and parts dealers	699	7	1.00	3	1.0	4.2	5
335	Electrical equipment	667	13	1.95	12	1.0	1.0	2
447	Gasoline stations	585	0	0.00	0	–	–	–
611	Educational services	560	0	0.00	0	–	–	–
713	Amusement and recreation	549	1	0.18	1	1.0	1.0	1
337	Furniture	526	18	3.42	10	1.0	4.0	4
312	Beverage and tobacco products	507	8	1.58	5	1.0	2.6	3
486	Pipeline transportation	500	20	4.00	6	2.0	7.0	9

Note: Active, privately owned and corporate facilities holding Clean Water Act permits, in three-digit NAICS industries with at least 500 such facilities, excluding NAICS 236 and 562. "Listed-firm facilities" belong to listed firms in my sample; the share is the percentage of the industry's facilities they hold. The last three columns describe the number of facilities per listed firm within the industry. Shares are lower bounds because the name match is conservative.

Table A.9. Clean Water Act noncompliance codes

Code	Type	Description
D	Failure to report	Non-receipt of the required Discharge Monitoring Report.
E	Effluent, monthly average limit	Violation of monthly average limits, or a single-event violation at the significant level.
S	Compliance schedule, violation	Enforcement action or permit schedule event more than 90 days late.
T	Compliance schedule, reporting	Enforcement action or permit reporting requirement more than 30 days late.
X	Effluent, non-monthly average limit	Violation of non-monthly average limits.
N	Reportable noncompliance	Noncompliance below the significant threshold but tracked in the report.

Note: Codes D, E, S, T, and X are Category I (significant noncompliance); code N is Category II (reportable noncompliance). Other codes, such as resolved (R), non-reportable violations (V), resolved pending (P, Q), legacy non-receipt for non-major permits (W), and undetermined (U), are not counted as noncompliance. Source: ECHO ICIS-NPDES Download Summary and Data Element Dictionary, U.S. EPA.

where λ_t is a year fixed effect. I cluster standard errors by firm. For water permits, which carry dates, I also estimate a version with firm fixed effects, which uses only changes within a firm over time. Table B.1 reports all estimates.

Because the dependent variable is a logarithm, I estimate equation (15) on firm-years holding at least one obligation of the relevant type. The elasticities therefore describe how the number of obligations varies with sales among firms that are regulated under the corresponding program.

Table B.1. Elasticity of compliance obligations with respect to sales

	Operating facilities only		All statuses	
Measure	Elasticity	Firm-years	Elasticity	Firm-years
Year fixed effects				
Water permits	0.095 (0.016)	3,254	0.095 (0.016)	3,254
Regulated facilities	0.104 (0.014)	9,632	0.120 (0.013)	11,492
Air regulations	0.134 (0.026)	5,834	0.149 (0.024)	6,387
Firm and year fixed effects				
Water permits	0.068 (0.025)	3,192	0.068 (0.025)	3,192

Note: Each cell reports the coefficient b from equation (15), with standard errors clustered by firm in parentheses. "Operating facilities only" counts regulated facilities and air regulations at currently operating facilities; "all statuses" counts all matched facilities. Water permits are dated, so the operating restriction does not apply to them. I report a within-firm estimate only for water permits: air-regulation records carry no dates, and the facility match is fixed, so the other two counts change little within a firm over time.

Interpretation. Exponentiating equation (15) gives $O_{it} = A_t\, S_{it}^{b}\, e^{\varepsilon_{it}}$ with $A_t = e^{\lambda_t}$, so the fitted relationship is

$$\widehat{O}_{it} \;=\; A_t\, S_{it}^{b},$$

a power function of sales, and

$$b \;=\; \frac{d\ln O}{d\ln S} \;=\; \frac{dO/O}{dS/S}$$

is the percentage change in obligations for a one-percent change in sales. If obligations were proportional to sales, b would equal one; if they did not depend on sales, b would equal zero. For two firms in the same year, fitted obligations satisfy

$$\frac{\widehat{O}_2}{\widehat{O}_1} \;=\; \left(\frac{S_2}{S_1}\right)^{b}.$$

With $b = 0.095$, a firm with ten times the sales of another holds $10^{0.095} \approx 1.24$ times as many water permits.

Dividing by sales gives fitted obligations per dollar of sales,

$$\frac{\widehat{O}_{it}}{S_{it}} \;=\; A_t\, S_{it}^{b-1}.$$

Because $b < 1$, obligations per dollar of sales fall as sales rise. With $b = 0.095$, a firm with ten times the sales carries $10^{-0.905} \approx 0.12$ times the water permits per dollar of sales.

B.2 Count Model

Permit counts are small integers: 63.1 percent of firm-years hold exactly one water permit, and the variance of permits is 10.6 times the mean. A log-linear regression summarizes such data with one elasticity of the average log count. To examine the shape of the relationship, I also estimate a count model of water permits.

Specification. Let $Y_{it} \geq 1$ denote firm i's water permits in year t. Because the sample contains only firm-years with at least one permit, I use a zero-truncated negative binomial model. The latent, untruncated count has mean

$$\mu_{it} \;=\; \exp\bigl(\gamma_{n(i)} + \lambda_t\bigr)\, S_{it}^{b}, \tag{16}$$

where $\gamma_{n(i)}$ is a three-digit NAICS industry effect, λ_t a year effect, S_{it} sales in billions of dollars, and b the exponent of interest. The variance is $\mu_{it}+\mu_{it}^2/\alpha$, where $\alpha > 0$ governs overdispersion, and the probability of a zero count is $P_{0,it} = (1 + \mu_{it}/\alpha)^{-\alpha}$. Each observation contributes the negative binomial probability of Y_{it} divided by $1 - P_{0,it}$. I maximize the likelihood directly and cluster standard errors by firm. Eighteen industries in which every firm-year holds exactly one permit (98 firm-years) are pooled into a single industry group. Because this group contains no variation in permit counts, its fitted latent mean approaches the lower boundary and it provides essentially no identifying variation for b; dropping it leaves the estimate unchanged.

Results. Table B.2 reports the estimates. The exponent is $b = 0.45$ (standard error 0.05). A nested specification, in which the latent mean is proportional to $\kappa_0+S_{it}^g$, places the additive constant κ_0 at zero and reproduces the power specification exactly. Within that nested specification, the restriction $g = 1$, under which the size-dependent part of permits grows in proportion to sales, is rejected: the likelihood-ratio statistic is 173 with one degree of freedom. A zero-truncated Poisson version of the model fits far worse, with a log-likelihood of $-4{,}483.6$ against $-3{,}739.5$, consistent with the overdispersion in the data (Table B.2).

The observed elasticity. The exponent b describes the latent mean, which includes firms that would hold no permit. Among firms that hold at least one permit, the conditional mean is $\mathbb{E}[Y_{it} \mid Y_{it} \geq 1] = \mu_{it}/(1 - P_{0,it})$, and its elasticity with respect to sales is

$$\frac{d \ln \mathbb{E}[Y_{it} \mid Y_{it} \geq 1]}{d \ln S_{it}} \;=\; b \left[1 - \frac{\alpha\, \mu_{it}\, P_{0,it}}{(\alpha + \mu_{it})\,(1 - P_{0,it})}\right],$$

which is smaller than b and rises with μ_{it}, because truncation at one permit binds most where expected counts are small. At the median sales of the smallest, middle, and largest employment thirds, it equals 0.11, 0.16, and 0.23; at the median firm it equals 0.17. The log-linear elasticity of 0.095 in Table B.1 is lower still, because it describes the average of log counts, which equal zero for the 63 percent of firm-years with one permit. The three numbers therefore describe the same data at three levels: the latent mean, the conditional mean among permit holders, and the average log count.

Robustness. The estimate of b is stable across the specifications in Table B.2: dropping the pooled industries, measuring size by employment instead of sales, and using each firm's average

sales over the sample, which removes year-to-year variation within firms. The last result indicates that the estimate reflects differences in scale across firms rather than annual fluctuations in sales.

Table B.2. Count-model estimates of water permits and firm size

Specification	b	α	Log-lik.	Firm-years	Firms	Cond. elasticity
Baseline (sales)	0.446 (0.050)	0.636	−3,739.5	3,254	541	0.165
Dropping pooled industries	0.446 (0.050)	0.636	−3,739.5	3,156	514	0.166
Employment instead of sales	0.436 (0.051)	0.679	−3,523.4	3,051	511	0.174
Firm-average sales	0.454 (0.049)	0.683	−3,735.6	3,254	541	0.171
Nested specification, latent mean proportional to $\kappa_0 + S^g$						
g free	$\hat{\kappa}_0 = 0$, $\hat{g} = 0.446$; log-likelihood −3,739.5					
$g = 1$ imposed	log-likelihood −3,826.0; likelihood-ratio statistic 173.1, $\chi^2(1)$					
Poisson (baseline sample)	0.321 (0.047)	–	−4,483.6	3,254	541	–

Note: Zero-truncated negative binomial models of water permits, with three-digit NAICS industry and year effects in the latent mean (16). Standard errors clustered by firm in parentheses. α is the negative binomial size parameter; the variance is $\mu + \mu^2/\alpha$. "Cond. elasticity" is the elasticity of $\mathbb{E}[Y \mid Y \geq 1]$ with respect to size, evaluated at the median firm. Eighteen industries in which every firm-year holds exactly one permit are pooled into one industry group in all rows except the second, which drops them. The last row is a zero-truncated Poisson model with the same latent mean.

B.3 Construction of Figures 1 and 2

Construction of Figure 1. Figure 1 is a binned scatter plot of equation (15). For each measure, I first regress $\ln O_{it}$ and $\ln S_{it}$ separately on year fixed effects and keep the residuals, so that both variables are net of year effects. I center each residual at its mean. I then sort firm-years into 20 equal-sized groups by residual log sales and plot the mean of each variable within each group. By the Frisch–Waugh–Lovell theorem, the slope of the fitted line through the residuals equals the estimate of b in equation (15). Both axes use the same logarithmic scale and are normalized so that zero corresponds to the sample mean in logs, that is, to the geometric-mean firm, so the dashed 45-degree line shows the relationship that obligations proportional to sales would produce.

Construction of Figure 2. Figure 2 compares obligations per dollar of sales across firms of different sizes.

Sample and groups. Let $\mathcal{S}$ denote the set of matched firm-years that hold at least one water permit and report employment, and let L_{it} denote firm i's employment in year t. I pool firm-years across 2009–2025 and compute the empirical terciles of employment over $\mathcal{S}$, with cutoffs $\ell_{1/3} = 4{,}130$

and $\ell_{2/3} = 17{,}300$ employees. The three groups are

$$G_1 = \{(i,t) \in \mathcal{S} : L_{it} < \ell_{1/3}\}, \qquad G_2 = \{(i,t) \in \mathcal{S} : \ell_{1/3} \leq L_{it} < \ell_{2/3}\},$$
$$G_3 = \{(i,t) \in \mathcal{S} : L_{it} \geq \ell_{2/3}\}.$$

Because I pool firm-years, a firm can move between groups over time.

The plotted measure. For each group k, the plotted value is

$$B_k = \frac{\overline{O}_k}{\widetilde{S}_k}, \qquad \overline{O}_k = \frac{1}{|G_k|} \sum_{(i,t) \in G_k} O_{it}, \qquad \widetilde{S}_k = \text{median}\{S_{it} : (i,t) \in G_k\},$$

where O_{it} is the count of water permits, sales S_{it} are measured in billions of dollars, and B_k is therefore obligations per \$1 billion of sales. I use the median for sales because a few firm-years with near-zero sales make the mean unrepresentative in the smallest group. I use the mean for obligation counts, which are bounded below by one and so cannot be dominated by near-zero values in the same way.

For air regulations, I apply the same formula within the subgroups $G_k^A = \{(i,t) \in G_k :$ firm i has at least one air-regulated facility in year $t\}$, computing both the mean count and the median sales on G_k^A. The numerator and denominator therefore always come from the same firm-years. Table B.3 reports the group sizes, mean counts, and median sales for each group.

Why I group by employment rather than sales. The measure divides by sales. If I formed the groups by sales, I would sort firms on the denominator of the ratio. Any temporary fluctuation in a firm's sales, or error in measuring them, would then push firm-years with unusually low sales into the bottom group and raise their measured obligations per dollar of sales, even if obligations did not depend on firm size at all. Grouping by employment, a separate measure of size, avoids sorting on the denominator. Employment and sales are strongly related, so the groups still separate small firms from large ones: median sales rise from \$0.67 billion in the smallest third to \$17.5 billion in the largest.

The figure plots B_1, B_2, and B_3 on a logarithmic vertical axis. On this scale, equal vertical distances correspond to equal proportional changes, so the two series can be compared directly even though air regulations are more numerous than water permits.

Regression and group comparisons. The two figures describe the same relationship in different ways, and their magnitudes differ for a mechanical reason. The largest third of firms have

Table B.3. Compliance obligations by employment tercile

	Bottom third	Middle third	Top third
Employment, L	$L < 4{,}130$	$4{,}130 \leq L < 17{,}300$	$L \geq 17{,}300$
Firm-years	1,031	1,032	1,029
Firms	234	206	153
Median sales ($ million)	671.1	3,330.2	17,539.0
Water permits			
Mean	1.44	2.09	3.46
Share with exactly one permit	0.77	0.64	0.49
90th percentile	2	4	6
Per $1 billion of sales	2.14	0.63	0.20
Median of firm-year ratios, per $1 billion	1.99	0.44	0.12
Regulated facilities			
Mean	3.09	6.81	9.99
Firm-years	844	929	967
Air regulations (firm-years with an air-regulated facility)			
Mean	11.39	18.07	14.08
Firm-years	544	649	801
Median sales ($ million)	1,023.5	3,737.6	18,256.0
Per $1 billion of sales	11.13	4.83	0.77
Median of firm-year ratios, per $1 billion	6.22	1.84	0.53

Note: Matched firm-years holding at least one water permit and reporting employment, divided into thirds by employment. Each statistic uses the firm-years in the group for which the relevant count is available; air regulations are computed for firm-years with at least one air-regulated facility. "Per $1 billion of sales" divides the mean count by median sales in the same firm-years, as in Figure 2; "median of firm-year ratios" is the median, within the group, of each firm-year's own count divided by its sales. Firm counts sum to more than the number of distinct firms because firms can move between groups over time.

about 26 times the median sales of the smallest third. If obligations followed the power function $O = AS^b$ exactly, the largest third would hold $26^{0.095} \approx 1.36$ times as many water permits, and their obligations per dollar of sales would satisfy

$$\frac{B_3}{B_1} \approx \left(\frac{\widetilde{S}_3}{\widetilde{S}_1}\right)^{b-1} = 26^{-0.905} \approx 0.05,$$

roughly a twenty-fold fall. In the group comparison, the ratio of mean counts is about 2.4 and the observed fall in obligations per dollar of sales is about eleven-fold (from 2.14 to 0.20). The regression describes the typical firm in logarithms, whereas the group comparison uses mean counts, which a few large firms with many permits pull upward. Both show that obligations rise far less than sales, so that obligations per dollar of sales fall sharply with firm size.

B.4 Link to the Model

In the model of Section 4, a firm's compliance effort is $e_j = g(q_j)$. The comparative statics depend on the entry–compliance wedge Ω_{if} (Section 5); concave, sub-proportional compliance generates a positive wedge when incumbents are at least as large as fringe firms. The estimates in this appendix concern a different object: counts of water permits, related to sales, across listed firms. They correspond to the model only under the assumptions that obligation counts are proportional to compliance effort and that sales are proportional to output. Under these assumptions, the latent mean (16) corresponds to the power technology $g(q) = Aq^b$, and every estimate in this appendix, from the log-linear elasticity of 0.095 to the latent exponent of 0.45, implies that compliance grows substantially less than proportionally with output.

The estimates do not pin down a single value of the exponent for the model. The latent exponent describes a mean that includes firms that would hold no permit, whereas the model's g describes the compliance of active, regulated firms, which is closer to the conditional elasticity among permit holders. Section 7 therefore uses the power technology with $b = 0.17$, the conditional elasticity at the median firm, as the baseline, and reports results from 0.11, the conditional elasticity at the median of the smallest third of firms, to 0.45, the latent exponent.

B.5 Calibration Targets

This subsection documents the targets used in Section 7 other than the compliance exponent, which comes from Appendix B.2. Table B.4 collects them.

Market size. I use the Fact 2 sample of active, privately owned and corporate facilities with Clean Water Act permits (Appendix A.4), keeping the 74,443 facilities with a state and a six-digit NAICS code, which fall into 11,503 occupied state × six-digit industry cells. Because the relevant question is how many competitors a typical producer faces, I weight each cell by its facilities: for each facility, I record the number of facilities in its own cell. The median cell contains only two facilities, but most facilities sit in larger cells. A state-level count understates the number of competitors for products traded across states, and ownership of several plants by one firm overstates the number of independent firms.

Listed firms' output share. For each industry, I divide the Compustat sales of listed firms matched to regulated facilities, assigned by their primary three-digit NAICS code, by gross output in the BEA gross domestic product (GDP) by industry accounts, in current dollars. Compustat sales are worldwide and consolidated across lines of business, whereas BEA gross output measures U.S. production in the industry, so the ratio overstates listed firms' domestic output share; in petroleum and coal products it exceeds one. Because BEA combines food with beverages and tobacco, I compare the combined Compustat industries 311 and 312 with the combined BEA industry.

Compliance costs. From Table 1 of U.S. Census Bureau (2008), I divide pollution-abatement operating costs by the value of shipments reported in the Annual Survey of Manufactures, by three-digit industry. Operating costs include depreciation of abatement capital, about 14 percent of the total across manufacturing. Permits and fees, reported separately in the survey's Table 9, amount to about 0.01 percent of shipments.

Comparison with cement markets. In the regional cement markets studied by Ryan (2012), the number of plants per market averages 4.75 and ranges from 1 to 20 (his Table II), and his preferred demand elasticity is -2.96 (his Table III).

C Proofs

The first part of this appendix treats proportional compliance, $g(q) = rq$; the second proves the general result; the third treats the examples; and the final part proves the welfare results. I write $Q_{-j} = Q - q_j$ for the output of firm j's rivals, and I evaluate s and its derivative s_E at rivals' compliance rQ_{-j} unless stated otherwise.

Table B.4. Calibration targets

	All industries	Food (311)	Paper (322)	Chemicals (325)	Primary metals (331)
A. Facilities in own state × six-digit industry cell, facility-weighted					
Median	23	7	5	11	5
25th–75th percentile	6–110	3–18	2–12	5–22	2–10
B. Sales of matched listed firms / BEA gross output					
2012	0.70	0.52	0.50	0.69	0.26
2019	0.61	0.43	0.51	0.64	0.34
C. Pollution-abatement operating costs / shipments, 2005 (percent)					
Share	0.44	0.29	1.10	0.86	1.14

Note: Panel A: Fact 2 sample; 74,443 facilities in 11,503 occupied cells. In petroleum and coal products (324), the facility-weighted median is 52 (25th–75th percentile 15–73). Panel B: "All industries" is manufacturing (NAICS 31–33); "Food" compares Compustat 311 and 312 with BEA food, beverage, and tobacco products. For petroleum and coal products the ratio exceeds one (1.26 in 2012, 1.22 in 2019), because Compustat sales include foreign sales and other lines of business. Panel C: U.S. Census Bureau (2008), Table 1; in petroleum and coal products the share is 0.79 percent, and in nonmetallic mineral products 0.61 percent.

Proportional compliance

Concavity of each firm's problem. Firm j chooses $q_j \geq 0$ to maximize

$$\pi_j(q_j; Q_{-j}) \;=\; \big[P(q_j + Q_{-j}) - c_j\big]q_j \;-\; \kappa(rq_j)\, s(rQ_{-j}; \phi),$$

taking Q_{-j} as given. Because rivals' compliance rQ_{-j} does not depend on q_j,

$$\frac{\partial^2 \pi_j}{\partial q_j^2} \;=\; 2P'(Q) + q_j P''(Q) \;-\; r^2 \kappa''(rq_j)\, s(rQ_{-j}; \phi).$$

By Assumption 1, $P'(Q) + q_j P''(Q) < 0$, so $2P'(Q) + q_j P''(Q) < P'(Q) < 0$. By Assumption 2, $\kappa'' \geq 0$ and $s > 0$. The second derivative is therefore negative, π_j is strictly concave in q_j, and firm j has a unique best reply $b_j(Q_{-j})$, characterized by the first-order condition (7) when it is positive and by $G_j(0, Q_{-j}) \leq 0$ when it is zero.

Proof of Lemma 2. Where $b_j(Q_{-j}) > 0$, implicit differentiation of the first-order condition gives

$$b_j'(Q_{-j}) \;=\; -\,\frac{\partial^2 \pi_j / \partial q_j \partial Q_{-j}}{\partial^2 \pi_j / \partial q_j^2}, \qquad \frac{\partial^2 \pi_j}{\partial q_j \partial Q_{-j}} \;=\; P'(Q) + q_j P''(Q) - r^2 \kappa'(rq_j)\, s_E.$$

Hence

$$1 + b_j'(Q_{-j}) \;=\; \frac{\partial^2\pi_j/\partial q_j^2 \;-\; \partial^2\pi_j/\partial q_j \partial Q_{-j}}{\partial^2\pi_j/\partial q_j^2} \;=\; \frac{P'(Q) - r^2\kappa''(rq_j)\,s + r^2\kappa'(rq_j)\,s_E}{\partial^2\pi_j/\partial q_j^2}.$$

The numerator is negative by condition (U1) and the denominator is negative by concavity, so $1 + b_j' > 0$. Condition (U1) is thus the requirement of Novshek (1985) that best replies have slope greater than -1. It follows that $Q = b_j(Q_{-j}) + Q_{-j}$ is strictly increasing in Q_{-j}, and for each $Q \geq b_j(0)$ there is a unique $Q_{-j} \geq 0$ with $b_j(Q_{-j}) + Q_{-j} = Q$. Define the backward reply $h_j(Q) = b_j(Q_{-j})$ for this Q_{-j}. By construction, $h_j(Q)$ is the unique output of firm j consistent with its best reply when market output is Q, and $h_j(Q) > 0$ if and only if $G_j(0, Q) > 0$.

Where $h_j(Q) > 0$, the function h_j solves $G_j(h_j(Q), Q) = 0$. Write $G_q = \partial G_j/\partial q_j$ and $G_Q = \partial G_j/\partial Q$, with

$$G_q \;=\; P'(Q) - r^2\kappa''(rq)\,s + r^2\kappa'(rq)\,s_E, \qquad G_Q \;=\; P'(Q) + qP''(Q) - r^2\kappa'(rq)\,s_E,$$

where the s_E terms arise because s is evaluated at $r(Q - q)$. By condition (U1), $G_q < 0$, so the implicit function theorem gives a continuously differentiable h_j with $h_j'(Q) = -G_Q/G_q$. Then

$$1 - h_j'(Q) \;=\; \frac{G_q + G_Q}{G_q}, \qquad G_q + G_Q \;=\; 2P'(Q) + qP''(Q) - r^2\kappa''(rq)\,s,$$

because the two s_E terms cancel. The sum $G_q + G_Q$ is negative by Assumptions 1 and 2, and $G_q < 0$, so $1 - h_j'(Q) > 0$. Under condition (U3), $G_Q < 0$, so $h_j'(Q) < 0$ where $h_j(Q) > 0$; where $h_j(Q) = 0$ it is locally constant. Hence h_j is nonincreasing. ■

A derivative used below. Where $h_f(Q) > 0$, differentiating (8) gives

$$\Pi_f'(Q) \;=\; P'(Q)\,h_f + \big[P(Q) - c_f\big]h_f' \;-\; r\,\kappa'(rh_f)\,s\,h_f' \;-\; r\,\kappa(rh_f)\,s_E\,(1 - h_f'),$$

with $h_f = h_f(Q)$ and s, s_E evaluated at $r(Q - h_f)$. The first-order condition (7) for the fringe gives $P(Q) - c_f - r\kappa'(rh_f)s = -P'(Q)h_f$. Substituting,

$$\Pi_f'(Q) \;=\; \big(1 - h_f'(Q)\big)\Big[P'(Q)\,h_f(Q) - r\,\kappa\big(rh_f(Q)\big)\,s_E\Big]. \tag{17}$$

Proof of Proposition 1. (i) By Lemma 2, $1 - h_f'(Q) > 0$, and by condition (U2) the bracket in (17) is negative. Hence $\Pi_f'(Q) < 0$ wherever $h_f(Q) > 0$. In any free-entry equilibrium with an active fringe, each fringe firm's output is a best reply, so it equals $h_f(Q)$, and zero profit net of the entry cost requires $\Pi_f(Q) = K$. Since $K > 0$, this requires $h_f(Q) > 0$, and on that region Π_f is strictly decreasing. The equation $\Pi_f(Q) = K$ therefore has at most one solution, Q^*, and every such equilibrium has market output Q^*.

(ii) The function Π_f in (8) is built from P, c_f, r, κ, s, and ϕ, through h_f, which solves the fringe's first-order condition (7) and depends on the same primitives. No characteristic of the incumbents enters. Hence, whenever an active-fringe equilibrium exists, its market output Q^*, which solves $\Pi_f(Q^*) = K$, depends only on these primitives and K.

(iii) Price $P(Q^*)$, each fringe firm's output $h_f(Q^*)$, and consumer surplus depend on market output only, and industry compliance is $E^* = \sum_j r q_j = rQ^*$. None depends on the incumbents. Each incumbent produces $h_i(Q^*)$, which depends on its own primitives. Adding up outputs, $Q^* = \sum_i h_i(Q^*) + M_f\, h_f(Q^*)$, which gives the fringe mass stated in the proposition. A change that affects only the incumbents, including a merger among them, changes the functions h_i but not Π_f. It is therefore absorbed through adjustments in the incumbents' outputs and the fringe mass, while Q^* and the market outcomes above remain unchanged, provided the fringe remains active. Other quantities, such as incumbents' profits and total compliance costs, can change. ■

Proof of Lemma 3. Under conditions (U1) and (U3), the incumbents-only market output Q^0 exists and is unique, because $Q - \sum_i h_i(Q)$ is continuous and strictly increasing (Novshek, 1985). Let $\hat{Q} = \max\{Q^0, b_f(0)\}$, where $b_f(0)$ is the output of a fringe firm serving the market alone, so that h_f is defined on $[\hat{Q}, \bar{Q}]$ and Π_f is continuous there. By hypothesis, $\Pi_f(\hat{Q}) > K$. At $\bar{Q}$, where $P(\bar{Q}) = c_f$, the fringe's first-order condition gives $G_f(0, \bar{Q}) = -r\kappa'(0)\, s \le 0$, so $h_f(\bar{Q}) = 0$ and $\Pi_f(\bar{Q}) = -\kappa(0)\, s = 0 < K$. By the intermediate value theorem there exists $Q^* \in (\hat{Q}, \bar{Q})$ with $\Pi_f(Q^*) = K$, and by Proposition 1(i) it is unique.

Set $q_i = h_i(Q^*)$ for each incumbent, $q_f = h_f(Q^*)$, and $M_f = [Q^* - \sum_i h_i(Q^*)]/h_f(Q^*)$. Since $\Pi_f(Q^*) = K > 0$, $h_f(Q^*) > 0$. By condition (U3), each h_i is nonincreasing, so $\sum_i h_i(Q^*) \le \sum_i h_i(Q^0) = Q^0 \le \hat{Q} < Q^*$, and hence $M_f > 0$. By construction, market output equals Q^*, each firm's output is its best reply to its rivals' output, and fringe firms earn zero profit net of the entry cost. This is a free-entry equilibrium with an active fringe. Any other such equilibrium has market output Q^* by Proposition 1(i), hence the same outputs $h_j(Q^*)$ and the same fringe mass. ■

General compliance

In this part, g satisfies Definition 1, $C = \kappa \circ g$, and for each firm s and s_E are evaluated at its rivals' compliance $E - g(q)$. Since $g(q) > 0$ and, by Assumption 2, $\kappa' > 0$ at positive effort, $C(q) > 0$ and $C'(q) \geq 0$ for $q > 0$.

Backward replies. Firm j takes its rivals' outputs, and therefore E_{-j}, as given, so by (R0) its profit is strictly concave in its own output, its best reply is unique, and when positive it solves $G_j(q_j, Q, E) = 0$. The partial derivatives of G_j are

$$G_q = P' - C''\, s + C'\, g'\, s_E, \qquad G_Q = P' + q_j P'', \qquad G_E = -C'\, s_E, \qquad G_c = -1.$$

Holding E fixed, rivals' compliance does not depend on Q, which is why G_Q contains no spillover term. By (R1), $G_q < 0$, so $h_j(Q, E)$ is continuously differentiable where positive, with

$$h_Q = -\frac{G_Q}{G_q} < 0, \qquad h_E = \frac{C'\, s_E}{G_q}, \qquad h_c = \frac{1}{G_q} < 0,$$

where $h_Q < 0$ because $G_Q < 0$ by Assumption 1, and h_E has the sign of ϕ or is zero.

Partial derivatives of fringe profit. Differentiating (9) and using the fringe's first-order condition, $P(Q) - c_f - C'(h_f)\, s = -P'(Q)\, h_f$, gives

$$\Pi_E \;=\; -C(h_f)\, s_E \;+\; h_{fE}\left(C(h_f)\, g'(h_f)\, s_E - P'(Q)\, h_f\right), \tag{18}$$

$$\Pi_Q \;=\; P'(Q)\, h_f\,(1 - h_{fQ}) \;+\; C(h_f)\, g'(h_f)\, s_E\, h_{fQ}. \tag{19}$$

Lemma C.1 (sign of Π_E). Under (R1) and (R2), Π_E has the sign of ϕ, and $\Pi_E = 0$ if and only if $\phi = 0$.

Proof. Since $C(h_f) > 0$, the first term in (18), $-Cs_E$, has the sign of ϕ and is nonzero when $\phi \neq 0$. By (R2), the bracket in the second term is positive, and h_{fE} has the sign of ϕ or is zero, so the second term is either zero or has the sign of ϕ. Both terms vanish when $\phi = 0$. ■

The first term is the direct effect of rivals' compliance on the cost of meeting a firm's own obligations. The second is the induced effect: rivals' compliance changes a fringe firm's output, and hence its own compliance and revenue.

The adding-up curve and the wedge. At a given fringe mass, write the second stage as

$$F(Q,E) \equiv \sum_i h_i + M_f\, h_f - Q = 0, \qquad H(Q,E) \equiv \sum_i g(h_i) + M_f\, g(h_f) - E = 0.$$

The adding-up curve (10) substitutes $M_f(Q,E) = [E - \sum_i g(h_i)]/g(h_f)$ into F. Since $\partial M_f/\partial Q = -H_Q/g_f$ and $\partial M_f/\partial E = -H_E/g_f$, with $g_f = g(h_f)$,

$$A_Q = F_Q - \frac{h_f}{g_f}\, H_Q, \qquad A_E = F_E - \frac{h_f}{g_f}\, H_E.$$

Holding (Q,E) fixed, an increase in incumbent i's cost changes only h_i, so

$$A_c \;=\; h_{ic} \;-\; \frac{g'(h_i)\, h_{ic}}{g_f}\, h_f \;=\; h_{ic}\, \frac{h_f}{g_f}\, \Omega_{if},$$

with Ω_{if} as in (11). Since $h_{ic} < 0$ and $h_f/g_f > 0$, A_c has the sign of $-\Omega_{if}$.

Condition (R4). Condition (R4) states that $\Delta \equiv F_Q H_E - F_E H_Q > 0$ at the equilibrium: the second-stage system, which determines (Q,E) for a given fringe mass, is locally nonsingular with this orientation.

Lemma C.2 (stability and the sign of D). Let $D \equiv \Pi_E A_Q - \Pi_Q A_E$. Under (R4), Assumption 3 holds at the equilibrium if and only if $D > 0$.

Proof. Hold the incumbents fixed and let the fringe mass vary. Differentiating $F = 0$ and $H = 0$ with respect to M_f gives $F_Q Q' + F_E E' = -h_f$ and $H_Q Q' + H_E E' = -g_f$, so

$$Q' \;=\; \frac{g_f F_E - h_f H_E}{\Delta} \;=\; \frac{g_f\, A_E}{\Delta}, \qquad E' \;=\; \frac{h_f H_Q - g_f F_Q}{\Delta} \;=\; -\,\frac{g_f\, A_Q}{\Delta}.$$

A fringe firm's operating profit at the second-stage equilibrium is $\widehat{\pi}_f(M_f) = \Pi_f(Q(M_f), E(M_f))$, so

$$\frac{d\widehat{\pi}_f}{dM_f} \;=\; \Pi_Q\, Q' + \Pi_E\, E' \;=\; \frac{g_f\,(\Pi_Q A_E - \Pi_E A_Q)}{\Delta} \;=\; -\,\frac{g_f\, D}{\Delta}.$$

Since $g_f > 0$ and $\Delta > 0$, $d\widehat{\pi}_f/dM_f < 0$ if and only if $D > 0$. ■

Proof of Proposition 2. Totally differentiating $\Pi_f(Q,E) = K$ and $A(Q,E;c) = 0$ gives

$$\begin{pmatrix} \Pi_Q & \Pi_E \\ A_Q & A_E \end{pmatrix} \begin{pmatrix} dQ^* \\ dE^* \end{pmatrix} = \begin{pmatrix} 0 \\ -A_c \end{pmatrix} dc_i,$$

and Cramer's rule gives $dQ^*/dc_i = -A_c\Pi_E/D$ and $dE^*/dc_i = \Pi_Q A_c/D$. By Lemma C.2, $D > 0$. (i) By (R3), $\Pi_Q < 0$, and A_c has the sign of $-\Omega_{if}$, so dE^*/dc_i has the sign of Ω_{if}. (ii) $-A_c$ has the sign of Ω_{if} and, by Lemma C.1, Π_E has the sign of ϕ, so dQ^*/dc_i has the sign of $\Omega_{if}\,\phi$. ■

Proof of Corollary 1. Since g is concave, its tangent at q_f lies above it, so $g(0) \le g(q_f) - q_f\, g'(q_f)$, and hence

$$\frac{g(q_f)}{q_f} \;\ge\; \frac{g(0)}{q_f} + g'(q_f) \;\ge\; g'(q_f) \;\ge\; g'(q_i),$$

where the second inequality uses $g(0) \ge 0$ and the last uses that g' is nonincreasing and $q_i \ge q_f$. Hence $\Omega_{if} \ge 0$. If g is strictly concave on $[0, q_i]$, the first inequality is strict; if $g(0) > 0$, the second is strict. ■

Examples

Fixed obligations. In this part, $g(q) = \bar{e} > 0$, and $\bar{\kappa} = \kappa(\bar{e})$. Each firm's first-order condition is $G_j(q_j, Q) = P(Q) + P'(Q)q_j - c_j = 0$, with $\partial G_j/\partial q_j = P'(Q) < 0$ and $\partial G_j/\partial Q = P'(Q) + q_j P''(Q) < 0$ by Assumption 1. Hence $h_j(Q) = [P(Q) - c_j]/[-P'(Q)]$ where positive, with $h_j'(Q) < 0$ and $1 - h_j'(Q) > 0$.

The slope of v_f. Differentiating $v_f(Q) = [P(Q) - c_f]\, h_f(Q)$ and using $P(Q) - c_f = -P'(Q)\, h_f(Q)$ gives

$$v_f'(Q) \;=\; P'(Q)\, h_f(Q)\, \big(1 - h_f'(Q)\big) \;<\; 0.$$

The equilibrium conditions. With fixed obligations, outputs add up to $Q = \sum_i h_i(Q) + M_f\, h_f(Q)$, industry compliance is $E = \bar{e}(M_I + M_f)$, and the zero-profit condition is $v_f(Q) - \bar{\kappa}\, s(E - \bar{e}; \phi) = K$. Solving the first for the fringe mass gives the industry compliance consistent with market output Q,

$$\varepsilon(Q) \;=\; \bar{e}\left[M_I + \frac{Q - \sum_i h_i(Q)}{h_f(Q)}\right],$$

and equilibrium market output solves $\Phi(Q) \equiv v_f(Q) - \bar{\kappa}\, s(\varepsilon(Q) - \bar{e}; \phi) - K = 0$.

Stability is equivalent to $\Phi'(Q^*) < 0$. Fix the incumbents and let the fringe mass vary. The adding-up condition defines market output as a function of M_f, with

$$\frac{dQ}{dM_f} \;=\; \frac{h_f(Q)}{1 - \sum_i h_i'(Q) - M_f\, h_f'(Q)} \;>\; 0,$$

since each $h_j' \le 0$. Along this path, industry compliance is $\bar{e}(M_I + M_f) = \varepsilon(Q)$, so a fringe firm's operating profit at the second-stage equilibrium induced by M_f is $\widehat{\pi}_f(M_f) = \Phi(Q(M_f)) + K$. Therefore

$$\frac{d\widehat{\pi}_f}{dM_f} \;=\; \Phi'(Q)\,\frac{dQ}{dM_f},$$

and since $dQ/dM_f > 0$, Assumption 3 holds at an equilibrium if and only if $\Phi'(Q^*) < 0$.

Proof of Corollary 2. Write $\Phi(Q; c_i) = v_f(Q) - \bar{\kappa}\, s(\varepsilon(Q; c_i) - \bar{e}; \phi) - K$, making explicit that ε depends on incumbent i's cost through h_i. Since incumbent i is active, $\partial h_i/\partial c_i = 1/P'(Q) = -1/|P'(Q)|$, so

$$\varepsilon_c \;\equiv\; \frac{\partial \varepsilon}{\partial c_i} \;=\; -\,\frac{\bar{e}}{h_f(Q)}\,\frac{\partial h_i}{\partial c_i} \;=\; \frac{\bar{e}}{|P'(Q)|\, h_f(Q)} \;>\; 0.$$

Then $\partial\Phi/\partial c_i = -\bar{\kappa}\, s_E\, \varepsilon_c$, and the implicit function theorem, which applies because $\Phi'(Q^*) < 0$, gives

$$\frac{dQ^*}{dc_i} \;=\; -\,\frac{\partial\Phi/\partial c_i}{\Phi'(Q^*)} \;=\; \frac{\bar{\kappa}\, s_E\, \varepsilon_c}{\Phi'(Q^*)} \;=\; \frac{\bar{\kappa}\, \bar{e}\, s_E}{|P'(Q^*)|\, h_f(Q^*)\, \Phi'(Q^*)}.$$

The factors $\bar{\kappa}$, $\bar{e}$, $|P'|$, and h_f are positive and $\Phi'(Q^*) < 0$, so dQ^*/dc_i has the sign of $-s_E$, which by Assumption 2(iii) is the sign of ϕ. This gives the formula in Corollary 2, whose sign is that of ϕ.

For industry compliance and the fringe mass, differentiate the zero-profit condition, which does not involve c_i, along the equilibrium path: $v_f'(Q^*)\, dQ^* = \bar{\kappa}\, s_E\, dE^*$. Combining with $dE^* = \varepsilon_Q\, dQ^* + \varepsilon_c\, dc_i$ and $\Phi'(Q^*) = v_f'(Q^*) - \bar{\kappa}\, s_E\, \varepsilon_Q$ gives

$$\frac{dE^*}{dc_i} \;=\; \frac{v_f'(Q^*)\, \varepsilon_c}{\Phi'(Q^*)}.$$

Both $v_f'(Q^*)$ and $\Phi'(Q^*)$ are negative and $\varepsilon_c > 0$, so $dE^*/dc_i > 0$ for every value of ϕ. Since $E^* = \bar{e}(M_I + M_f^*)$, also $dM_f^*/dc_i = \bar{e}^{-1}\, dE^*/dc_i > 0$. ■

Fixed plus proportional compliance: the limit as $\bar{e} \to 0$. Let $g(q) = \bar{e} + rq$ with $\bar{e} > 0$, so that $\Omega_{if} = \bar{e}/q_f > 0$ and Proposition 2 applies. For the limit, it is convenient to eliminate the fringe mass differently. Since $E = \bar{e}(M_I + M_f) + rQ$ whenever outputs add up, the pair $\Pi_f(Q, E) = K$

and

$$\tilde{A}(Q,E;c) \;\equiv\; \sum_i h_i(Q,E) + \left[\frac{E-rQ}{\bar{e}} - M_I\right] h_f(Q,E) - Q \;=\; 0$$

has the same solutions as the pair $\Pi_f = K$ and $A = 0$, so it gives the same comparative statics. Its partial derivatives are $\tilde{A}_c = h_{ic}$, $\tilde{A}_Q = F_Q - (r/\bar{e})\, h_f$, and $\tilde{A}_E = F_E + h_f/\bar{e}$, and with $\tilde{D} = \Pi_E \tilde{A}_Q - \Pi_Q \tilde{A}_E$,

$$\frac{dQ^*}{dc_i} \;=\; -\,\frac{\tilde{A}_c\,\Pi_E}{\tilde{D}} \;=\; -\,\frac{\bar{e}\,h_{ic}\,\Pi_E}{\bar{e}\,(\Pi_E F_Q - \Pi_Q F_E) \;-\; h_f\,(\Pi_Q + r\,\Pi_E)}.$$

Suppose that, as $\bar{e} \to 0$, the active-fringe equilibrium converges to the equilibrium under proportional compliance, with the fringe mass bounded. Then h_{ic}, Π_E, F_Q, F_E, and h_f converge, with h_f converging to a positive limit. At $\bar{e} = 0$, industry compliance is $E = rQ$, so $\Pi_Q + r\,\Pi_E$ is the derivative of the fringe profit $\Pi_f(Q, rQ)$ under proportional compliance, which is negative by Proposition 1(i). The denominator therefore converges to a positive limit and the numerator to zero, so $dQ^*/dc_i \to 0$, and $|dQ^*/dc_i|$ is of order $\bar{e}$ near $\bar{e} = 0$. ■

Power technology. For $g(q) = Aq^b$ with $0 < b < 1$, g is strictly concave and $g(0) = 0$, so Corollary 1 gives $\Omega_{if} > 0$ whenever $q_i \geq q_f$. Directly, $\Omega_{if} = A(q_f^{b-1} - b\,q_i^{b-1})$, and $q_i^{b-1} \leq q_f^{b-1}$ gives $\Omega_{if} \geq A(1-b)\,q_f^{b-1} > 0$.

Welfare

Proof of Corollary 3. Consumer surplus satisfies $CS'(Q) = P(Q) + P'(Q)\,Q - P(Q) = -P'(Q)\,Q > 0$, so it moves with market output. Part (i) follows from Proposition 1(iii), and part (ii) from Proposition 2(ii); its final statement, under a positive wedge, follows from Corollary 1. ■

Proof of Proposition 3. Write $\mathcal{C}_j = \kappa(g(q_j))\, s(E_{-j};\phi)$ for firm j's total compliance cost, including the spillover; it differs from the stand-alone cost $C(q_j) = \kappa(g(q_j))$. Since $\pi_j = [P(Q) - c_j]\,q_j - \mathcal{C}_j$, welfare (12) equals gross surplus minus all costs:

$$W \;=\; \int_0^Q P(x)\,dx \;-\; \sum_{i=1}^{M_I}\big(c_i\,q_i + \mathcal{C}_i\big) \;-\; M_f\,\big(c_f\,q_f + \mathcal{C}_f + K\big).$$

Let a prime denote the derivative with respect to M_f along the second-stage equilibrium, which exists because the second stage is locally differentiable in M_f. Since $Q = \sum_i q_i + M_f\,q_f$, we have

$Q' = \sum_i q_i' + q_f + M_f\, q_f'$, and therefore

$$W' \;=\; \big[P(Q)\, q_f - c_f\, q_f - \mathcal{C}_f - K\big] \;+\; \sum_{i=1}^{M_I} \big[(P(Q) - c_i)\, q_i' - \mathcal{C}_i'\big] \;+\; M_f\, \big[(P(Q) - c_f)\, q_f' - \mathcal{C}_f'\big].$$

The first bracket equals $\pi_f - K$, which is zero at the free-entry equilibrium: the marginal fringe firm's private gain from entry cancels. For each firm, $e_j = g(q_j)$, so

$$\mathcal{C}_j' \;=\; \kappa'(e_j)\, g'(q_j)\, s(E_{-j};\phi)\, q_j' \;+\; \kappa(e_j)\, s_E(E_{-j};\phi)\, \frac{dE_{-j}}{dM_f}.$$

Each firm's first-order condition gives $P(Q) - c_j - \kappa'(e_j)\, g'(q_j)\, s(E_{-j};\phi) = -P'(Q)\, q_j$. Substituting,

$$(P(Q) - c_j)\, q_j' - \mathcal{C}_j' \;=\; -P'(Q)\, q_j\, q_j' \;-\; \kappa(e_j)\, s_E(E_{-j};\phi)\, \frac{dE_{-j}}{dM_f}.$$

Summing over incumbents with weight one and over the fringe with weight M_f gives (13).

(i) Since $-P'(Q)\, q_j > 0$ for each active firm, the first sum is negative when each $q_j' \le 0$ with at least one strict inequality. (ii) Since $\kappa(e_j) > 0$ and s_E has the sign of $-\phi$, each term $-\kappa(e_j)\, s_E$ has the sign of ϕ; when each $dE_{-j}/dM_f > 0$, the second sum has the sign of ϕ. (iii) follows from (i) and (ii).

Under fixed obligations, $E_{-j} = E - \bar{e}$ and $E = \bar{e}(M_I + M_f)$, so $dE_{-j}/dM_f = \bar{e} > 0$. Each firm's output is $h_j(Q)$ with $h_j' < 0$, and, as shown in the proof of Corollary 2, $dQ/dM_f > 0$, so $q_j' = h_j'(Q)\, dQ/dM_f < 0$. Both conditions therefore hold. ■

D A Market for Compliance Inputs

This appendix shows that the reduced-form spillover of Section 4 arises from a simple market for compliance inputs, and that the sign of the spillover is set by the supply side of that market.

Setup. Meeting compliance obligations requires compliance inputs, such as testing, monitoring, auditing, and consulting services. Firm j purchases $\kappa(e_j)$ units of these inputs at price p per unit, and takes the price as given. The price depends on the industry's compliance obligations, $p = p(E)$, because the scarce resource in supplying these inputs, such as accredited laboratory capacity, auditor hours, or a shared monitoring program, is used in proportion to the obligations

the industry must meet. Firm j's compliance cost is then

$$p(E)\,\kappa(e_j) \;=\; \bar{p}\,\kappa(e_j)\,\sigma(E), \qquad \sigma(E) \equiv p(E)/\bar{p},$$

where $\bar{p}$ is the price at a reference level of industry compliance. Two supply structures give the two signs of the spillover, as Figure D.1 illustrates.

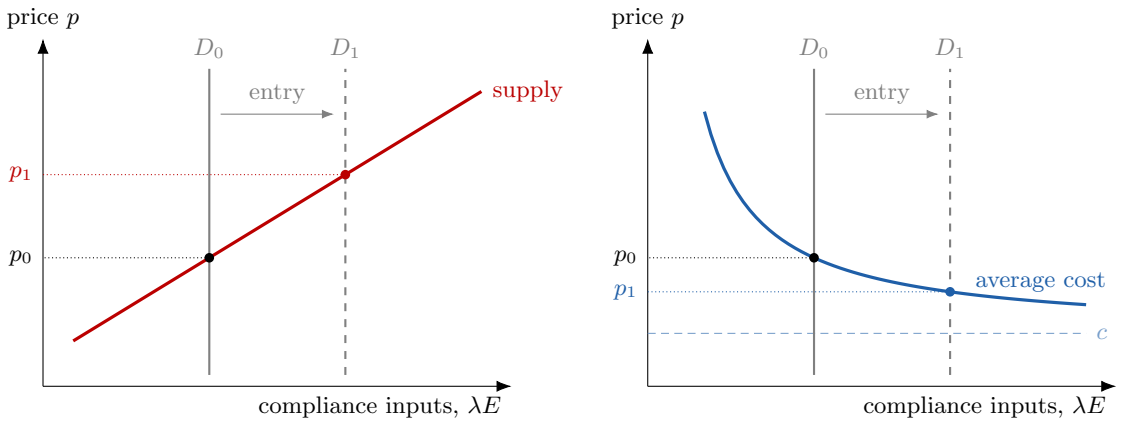
- *Capacity-constrained supply.* Compliance inputs are supplied competitively along an upward-sloping supply curve, with supply elasticity ε_S. Then $p'(E) > 0$, the spillover is rivalrous, and the elasticity of a firm's compliance cost with respect to industry compliance is $\eta = 1/\varepsilon_S$. Auditors after the Sarbanes–Oxley Act (Duguay et al., 2020) and PFAS testing laboratories (U.S. Government Accountability Office, 2025) are examples.

- *Shared fixed costs.* Supplying compliance inputs requires a fixed cost F, such as a regional monitoring program or a jointly developed study, and a constant cost c per unit of obligations, and compliance inputs are priced at average cost, $p(E) = c + F/(\lambda E)$ for some $\lambda > 0$. Then $p'(E) < 0$, the spillover is cooperative, and $\eta = -F/(\lambda E\, p)$, minus the share of the fixed cost in the price. Compliance groups under California's industrial stormwater permit (California State Water Resources Control Board, 2014) and EPA's pesticide task forces (U.S. Environmental Protection Agency, 1994) are examples.

Relation to the reduced form. In Section 4, a firm's compliance cost depends on its rivals' compliance, E_{-j}; here it depends on industry compliance, $E = E_{-j} + g(q_j)$, which includes the firm's own. For a fringe firm, whose own compliance is negligible relative to the industry's, the two coincide. For an incumbent, they differ by its own compliance. The following remark shows that the main result does not depend on this distinction.

Remark 1. *Suppose each firm's compliance cost is $C(q_j)\,\sigma(E)$, where σ is positive and differentiable and $\sigma'(E)$ has the sign of $-\phi$, and firms take E as given. Suppose Assumptions 1–3 hold, (R0) and (R4) hold with s replaced by $\sigma(E)$, and $P'(Q) - C''(q)\,\sigma(E) < 0$ for each active firm. Then the conclusions of Proposition 2 hold:* $\operatorname{sgn}(dE^*/dc_i) = \operatorname{sgn}(\Omega_{if})$ *and* $\operatorname{sgn}(dQ^*/dc_i) = \operatorname{sgn}(\Omega_{if}) \cdot \operatorname{sgn}(\phi)$.

Proof. Because $\sigma(E)$ is fixed in each firm's problem, firm j's output solves $G_j \equiv P(Q) + P'(Q)\,q_j - c_j - C'(q_j)\,\sigma(E) = 0$, with $G_q = P' - C''\sigma < 0$ by hypothesis, $G_Q = P' + q_j P'' < 0$, $G_E = -C'\sigma'$, and $G_c = -1$. Hence $h_Q < 0$, $h_c = 1/G_q < 0$, and $h_E = C'\sigma'/G_q$, which has the sign of ϕ or is zero.



(a) Capacity-constrained supply: price rises, rivalrous spillover

(b) Shared fixed costs: price falls, cooperative spillover

Figure D.1. The market for compliance inputs and the sign of the spillover

Note: The horizontal axis is the quantity of compliance inputs the industry must buy, proportional to industry compliance E. Demand is vertical because obligations are set by regulation rather than by price. Entry of fringe firms raises industry compliance and shifts demand from D_0 to D_1. In panel (a, red), compliance inputs are supplied along an upward-sloping supply curve, so the price rises and each firm's compliance cost rises: a rivalrous spillover, with $\eta = 1/\varepsilon_S$. In panel (b, blue), supplying compliance inputs requires a fixed cost shared across the industry's obligations, so the average cost $c + F/(\lambda E)$ falls and each firm's compliance cost falls: a cooperative spillover. Schematic.

Fringe profit is $\Pi_f(Q, E) = [P(Q) - c_f]\, h_f - C(h_f)\, \sigma(E)$. Using the fringe's first-order condition, $P - c_f - C'\sigma = -P' h_f$,

$$\Pi_E \;=\; -C(h_f)\, \sigma'(E) \;-\; P'(Q)\, h_f\, h_{fE}, \qquad \Pi_Q \;=\; P'(Q)\, h_f\, (1 - h_{fQ}) \;<\; 0.$$

Both terms of Π_E have the sign of ϕ when $\phi \neq 0$, and both vanish when $\phi = 0$, so Π_E has the sign of ϕ, as in Lemma C.1. The adding-up curve is unchanged, so $A_c = h_{ic}\, [h_f / g(h_f)]\, \Omega_{if}$ has the sign of $-\Omega_{if}$, and Lemma C.2 applies without change, so $D > 0$. The proof of Proposition 2 then applies with $\Pi_Q < 0$ in place of (R3). □

In the calibration of Section 7, the two versions give output effects within about 5 percent of each other at the baseline.

Interpreting the spillover elasticity. This interpretation gives η an economic meaning. Under capacity-constrained supply, the baseline value $|\eta| = 0.3$ corresponds to a supply elasticity of compliance inputs of about 3, and the break-even value $\eta = 0.8$ for a 0.5 percent output effect corresponds to a supply elasticity of about 1.25, a fairly inelastic supply. Under shared fixed costs,

$\eta = -0.3$ corresponds to fixed costs that make up 30 percent of the price of compliance inputs. These mappings suggest where future work could measure the spillover: in the supply of compliance inputs such as environmental testing, monitoring, and consulting.