Innovation and lobbying in response to regulatory uncertainty

Andrew Hultgren*

July 20, 2026

Abstract

What is the impact of regulation on innovation when firms can lobby the regulator? I review over 7,000 epidemiological studies to assemble novel data on shocks to firm regulatory expectations: 104 initial discoveries of chemical harms affecting 60 different industries over four decades. I find shocked firms slowly increase innovation expenditures by 1.8% of pre-period market capitalization — a total R&D increase that is net of any within-firm substitution across innovative activities. However, firms also increase lobbying by 11% of pre-period levels. Because lobbying shifts regulatory expectations, it can distort firm innovation incentives. Empirically, I find firms with low preexisting lobbying capacity make additional R&D investments at triple the rate of high lobbying capacity firms. This suggests firms may treat innovation and lobbying as substitutes, implying firms' option to lobby may reduce their innovation response. Lobbying has ambiguous welfare implications, but the social value of foregone innovation is potentially large.

JEL Codes: Q55, Q58, D72, D22, O31, O33, I18

*I am indebted to Maximilian Auffhammer, Michael Greenstone, Solomon Hsiang, and James Sallee for their support and feedback on this project. I thank Ashish Arora, Severin Borenstein, Thomas Covert, Lucas Davis, Meredith Fowlie, Eyal Frank, Koichiro Ito, Jennifer Kao, Larry Karp, Ashley Langer, Jorge Lemus, Jeffrey Perloff, David Popp, Gordon Rausser, Joseph Shapiro, Leo Simon, Reed Walker, Brian Wright, and seminar participants at AEA at ASSA, LSE/Imperial Environmental Economics Workshop, Stanford Political Economics of Environmental Sustainability Conference, UChicago EPIC Seminar, UC Berkeley ERE and Climate Lunch seminars, and UCSB Occassional Workshop for helpful comments and suggestions. I thank Sarah Dang, Hannah Maeder, Beau Smit, and Xiaoting Sun for excellent research assistance. All errors and omissions are my own. Contact: University of Illinois Urbana-Champaign Department of Agricultural and Consumer Economics, 332 Mumford Hall, 1301 West Gregory Dr. Urbana, IL 61801 (email: ahultgr@illinois.edu).

1 Introduction

Innovation is a critical driver of economic growth and human well-being (Romer, 1990), and environmental regulation can drive innovation, even changing the direction of technical change (Schumpeter, 1942; Porter, 1996; Popp, Newell, and Jaffe, 2010; Acemoglu et al., 2012). While distortions to innovation associated with appropriability of rents are widely recognized (Arrow, 1962), innovation incentives are also affected by a firm's option to lobby over future regulation. Faced with potential environmental regulation, forward-looking firms have an incentive to innovate (should regulation come to pass), but they must trade off this incentive against their option to lobby the regulator. This paper asks what is the effect of the possibility of environmental regulation on firm innovation expenditures and on firm lobbying?

To fix ideas, consider a well-documented example (Benedick, 1991; Dugoua, 2023). In the 1960s DuPont was the primary manufacturer of chlorofluorocarbons (CFCs) – a non-toxic set of products in widespread use with no known harms. However, in 1974, two atmospheric chemists posited that CFCs could *theoretically* catalyze the destruction of Earth's ozone layer. DuPont was suddenly faced with a shock to its expectations over the probability of future CFC regulation. This shock could induce DuPont to innovate, seeking to capture the market for substitute products should regulation come to pass. But DuPont also had the option to lobby the regulator, changing the probability of future regulation. In fact, DuPont did both.

In this study, I expand beyond this single example to estimate firm innovation and lobbying behaviors across a wide range of chemicals over multiple industries and decades. I employ for identification newly-constructed data on external information shocks: discoveries by human-health researchers of novel environmental externalities, linked to firms at the detailed industry-by-year level. Specifically, a loophole in U.S. chemicals regulation allowed tens of thousands of chemicals to enter the U.S. economy without testing for chronic, low-dose health harms. Epidemiological researchers were left to test for such harms, and I scrape over 7,000 studies from this literature to create a novel dataset of first-time scientific discoveries of harms from widely used industrial chemicals. These discoveries were made by independent epidemiologists (not firm scientists) who target their studies based on factors including chemical characteristics and costliness of testing, rather than changes in firm or other economic outcomes (Section 3).

As a measure of innovation, I focus on firm R&D expenditures reported in Compustat. These

expenditures are a useful measure of innovation for three main reasons. First, they represent the key financial input to knowledge production over which firm managers must optimize in their budgeting decisions. That is, they represent the key allocative decision firm managers actually face. Second, estimated effects represent *aggregate* changes in firm innovation effort, net of any within-firm substitution across innovation activities. Third, R&D expenditures are denominated in monetary units, meaning their value scales with expected payoffs to the innovative activity, thus avoiding issues associated with measurement of the relative value of innovation outputs such as patents.[1] A fourth, more pragmatic, motivation also exists: lobbying is also observed as a firm-level expenditure, and is thus more readily comparable to R&D expenditures than to other innovation outcomes.

A simple conceptual framework motivates the problem, in which a monopolist facing regulation can invest in clean or dirty innovation, and can also lobby for or against regulation. Of course, I do not observe clean or dirty innovation in the R&D expenditures data, but its inclusion in the conceptual framework serves to highlight that total firm innovation could increase or decrease under a regulatory threat, depending on whether substitution into clean innovation dominates substitution out of dirty innovation.[2] The firm invests in clean and dirty innovation until the respective expected marginal payoff equals the marginal investment cost. Lobbying can distort expected innovation payoffs by shifting probability mass between regulatory states, with the direction of firm lobbying controlled by the sign of the regulatory profit wedge.

Taking the model to data, I estimate firms' dynamic innovation and lobbying expenditure responses to the first-time discovery of an externality affecting their detailed industry. Because firms may still be responding to one discovery when they are exposed to another, I employ a distributed lag model in my estimates (rather than an event study model) to explicitly allow for overlapping impulse-responses to multiple treatments.[3] I further employ a first-differenced estimator, implicitly conditioning out firm fixed effects, and additionally control for sector-

[1] A broad literature has made substantial progress on this issue of differential patent values (Hall and Harhoff, 2012) though complexities remain (Higham, De Rassenfosse, and Jaffe, 2021). This is not to detract from the literature studying firm patenting, which is able to observe and classify patented technologies, something the expenditure data do not permit.

[2] The conceptual framework also includes the direction of lobbying, even though the data report aggregate lobbying, as the sign of aggregate lobbying response is likewise ambiguous.

[3] Schmidheiny and Siegloch (2023) propose a modified event study design to accommodate multiple treatments, but they note that the conventional event study design cannot do so. Further, they show their modified design is equivalent to estimating a distributed lag model such as that employed here.

year fixed effects to account for differential drivers of innovation and lobbying across sectors over time. My identifying assumption is that within-firm changes in exposure to pollutant discoveries by external scientists are plausibly exogenous, conditional on sector-year shocks. Extended placebo tests for both innovation and lobbying outcomes indicate that pre-period effects are flat even when examining up to ten periods before a given discovery, suggesting that these pollutant discoveries are plausibly exogenous and that treated firms do not anticipate pollutant discoveries before they occur.[4]

In response to a new discovery of a chemical harm, I find that firms increase their R&D investments by a total of 1.8% relative to pre-period market capitalization over a six year period. However, they also increase total lobbying investments by 11% of pre-period levels over a roughly two year period.[5] The use for identification of first-time discoveries of chemical harms mitigates concerns over forward-looking anticipation effects in these estimates (Rittenhouse and Zaragoza-Watkins, 2018). Further, these estimates represent changes in aggregate firm actions and not simply within-firm substitution across clean or dirty innovation (likewise, lobbying) investments. Importantly, while regulatory-induced increases in firm *clean* innovation *outputs* are well-documented (e.g. Lanjouw and Mody, 1996; Popp, 2006; Calel and Dechezleprêtre, 2016; Brunel, 2019; Dugoua, 2023), the possibility of a net decrease in *total* innovation inputs *or* outputs — driven by substitution out of other innovation activities — has been difficult to exclude (Dechezleprêtre and Sato, 2017; Lu and Pless, 2026).

Testing for heterogeneity in the dynamics of firm responses, I find suggestive evidence that preexisting lobbying capacity is an important descriptor of heterogeneity in the innovation response. Firms with low preexisting lobbying capacity make additional R&D investments at triple the rate of their high lobbying capacity counterparts. A symmetric test for heterogeneity in firm lobbying responses yields consistent findings, though with much more noise. I also find suggestive evidence that lobbying spillovers may play an important role in firm innovation responses. Across all tests employed, results are consistent with innovation and lobbying as

[4]The lobbying data contain a range of firm-trending behaviors, and thus lobbying regressions contain a firm fixed effect which absorbs firm trends in a first differences estimator. No such fixed effect is necessary in the innovation specification, and estimates with ten years of leads and lags are robust with or without it (results available upon request).

[5]Discovery of a novel externality could shift firm innovation actions through expectations over future demand or future regulation. The finding that firms substantially increase their lobbying efforts strongly suggests these discoveries shift regulatory expectations.

substitutes in firm actions, on average. This suggests that when faced with potential regulation, the option to lobby may induce these firms to innovate less than they otherwise would have. Drawing on the literature valuing firm innovation investments (Hall, 1999; Bloom, Schankerman, and Van Reenen, 2013; Jones and Summers, 2020), the social value of in-sample foregone innovation implied by this study is about $144 billion.[6]

This paper most directly contributes to the literature on environmental regulation and firm innovation (e.g., Popp, 2006; Aghion et al., 2016; Calel and Dechezleprêtre, 2016; Brunel, 2019; Calel, 2020; Brown, Martinsson, and Thomann, 2022; Dugoua, 2023; Dugoua and Gerarden, 2025; Lu and Pless, 2026).[7] The majority of this literature studies changes in clean innovation outputs, documenting important increases in clean patenting when firms face environmental regulation but leaving open the question of effects on total clean and non-clean innovation (Dechezleprêtre and Sato, 2017). Evidence on total innovation *inputs* (R&D expenditures) — the input to knowledge production over which firm managers must make allocation decisions — is particularly limited. Calel (2020) employs a matched DiD estimator for UK firms and finds R&D expenditures increased for firms that entered regulation under the EU-ETS from 2005–2008; Brown, Martinsson, and Thomann (2022) employ a cross-country DiD design, and findings suggest lagged SO_x taxes from 1990–2012 increased firm R&D expenditures in industries with high SO_x pollution intensities.[8]

This study extends this literature by: 1) providing broad evidence on U.S. firm R&D expenditures, with over 100 treatments spanning over 400 treated firms across 60 distinct industries and 44 distinct years; 2) reviewing over 7,000 epidemiological studies to assemble novel data on over 100 external information shocks that precede regulatory events, mitigating against treatment endogeneity and firm anticipation effects in estimation; and 3) providing novel suggestive

[6]For context, firms affected by a scientific discovery at any point in my sample represent $2.4 trillion in market capitalization in the year 2020.

[7]See Popp (2019) for a survey of this literature.

[8]I am aware of three studies that provide evidence on total innovation outputs (patents) and environmental regulation. Aghion et al. (2016) study the automobile industry under fuel price shocks, finding that increases in clean patenting are nearly exactly offset by decreases in dirty patenting, while "grey" patenting does not respond. Calel and Dechezleprêtre (2016) employ a matched difference-in-differences (DiD) estimator under the introduction of the EU Emissions Trading Scheme (EU-ETS), and find increases in both clean and other (non-clean) patenting. Dugoua and Gerarden (2025) focus on the subset of inventors with clean-energy patents, and find natural gas price shocks increase clean, dirty, and grey energy patenting among these inventors, which is partially offset by decreases in non-energy patenting. Additionally, Aghion, Bergeaud, and Van Reenen (2023) study labor (not environmental) regulation in France, finding labor regulation reduces firm patenting.

evidence on an important source of heterogeneity in firm innovation responses — a firm's ability to lobby the regulator.

As suggested by the final point above, this study also contributes to the political economy literature on firm lobbying. Empirical work has ranged from estimating the effect of corporate lobbying on politician voting behavior (e.g. Ansolabehere, De Figueiredo, and Snyder Jr, 2003; Kang, 2016; Meng and Rode, 2019), to whether lobbyists sell subject matter expertise or access to politicians (i Vidal, Draca, and Fons-Rosen, 2012; Bertrand, Bombardini, and Trebbi, 2014), to charitable giving as lobbying (Bertrand et al., 2020), to returns to corporate lobbying (Kang, 2016), to correlations between lobbying and climate change regulations (Kim, Urpelainen, and Yang, 2016; Brulle, 2018). Findings tend to be consistent with lobbying effects on policymaking as on-net distortionary, though information gains cannot be ruled out and are less studied because information transmission is difficult to observe (Bombardini and Trebbi, 2020).

To my knowledge, the only work mentioning political influence broadly as a source of heterogeneity in firm innovation is that of Akcigit, Baslandze, and Lotti (2023), who study firm growth and political connections (hiring of local politicians) among Italian firms.[9] They descriptively find market-leading firms are more politically connected but less likely to patent overall.[10] Distinct from the present study, Akcigit, Baslandze, and Lotti (2023) do not study firm lobbying expenditures, innovation expenditures, or responses to regulatory shocks. Even so, their cross-sectional documentation of lower patenting among politically connected Italian firms is directionally consistent with evidence in this study suggesting innovation and lobbying may be substitutes in firm responses to the prospect of future regulation.

The tradeoffs between innovation and lobbying faced by firms apply to a range of contexts. For example, they apply to health economics and the regulation of new pharmaceutical drugs found to have previously-unknown side effects, or to finance and the regulation of high-frequency trading over the price volatility it introduces, or to network economics and the regulation of technology firms over questions regarding the spread of misinformation.

[9]Similarly motivated work studies regulatory capture and firm growth (Slinko, Yakovlev, and Zhuravskaya, 2005), business leaders running for political office as a channel for firm influence (Gehlbach, Sonin, and Zhuravskaya, 2010), and effects of data-sharing via Chinese state procurement contracts on firm AI innovation (Beraja, Yang, and Yuchtman, 2023).

[10]Akcigit, Baslandze, and Lotti (2023) causally study a number of additional outcomes, finding political connections causally increase firm survival rates, employment, and revenues, but not productivity, with negative growth implications overall.

The remainder of the paper proceeds as follows: Section 2 motivates the problem with a simple conceptual framework, Section 3 describes the construction of the novel dataset of epidemiological discoveries and gives details on the innovation and lobbying expenditure data, Section 4 presents a descriptive analysis, Sections 5 and 6 describe the empirical models and results, Section 7 gives a welfare interpretation, and Section 8 concludes.

2 Conceptual framework

A simple conceptual framework motivates the problem. A monopolist earns rents in the production of good d with convex costs, and innovates to reduce marginal costs. Production or consumption of d has a negative health effect on exposed populations, but this harm is unknown to all actors. Then, independent health researchers discover that good d has a health harm, causing the regulator to consider a ban. The firm now has three margins of adjustment: 1) it can adjust investments in "dirty" innovation (ρ_d) targeting payoffs in the unregulated state; 2) it can innovate (ρ_c) in a "clean" substitute good c, targeting payoffs in the regulated state; and 3) it can make lobbying investments $\lambda = \{\lambda_f, \lambda_a\}$ to influence the posterior probability of regulation η.[11] The firm endogenously chooses to lobby either in favor of regulation (λ_f) to increase its probability of enactment, or against regulation (λ_a) to decrease its probability. Innovation and lobbying have decreasing (in magnitude) marginal returns to investment.

The firm plays a two-stage game. In stage 1, the firm makes innovation and lobbying investments. Then the regulator decides whether or not to ban good d. In stage 2, the firm sets production decisions q_c or q_d based on the regulatory decision. The firm's second stage knowledge stock is fixed by its first stage innovation investment, thus second stage profits can be written as a reduced-form function of first stage innovation.

The stage 2 solution for each regulatory state $r \in \{c, d\}$ is given by the firm's monopolist action: $q_r^* = \text{argmax}_{q_r} \pi(q_r, \rho_r)$, which implies the payoff $\pi_r^* = \pi_r^*(\rho_r)$.[12] The stage 1 problem

[11] For simplicity, this conceptual framework takes the stringency of potential regulation (a ban) as given. If, e.g., taxation is the regulatory instrument considered, then firms may lobby over both regulatory stringency and probability. See Gowrisankaran, Langer, and Zhang (2022) for an example of recent work which models firm technology adoption in response to both the stringency and probability of regulation (but does not model firm lobbying).

[12] This excludes a "strong" Porter hypothesis result where good c dominates good d in the unregulated state. The strong Porter hypothesis has generally been rejected in empirical studies (Ambec et al., 2013).

is:

$$\max_{\rho_c,\rho_d,\lambda} \ \eta(\lambda) \times \pi_c^*(\rho_c) \ + \ \big(1-\eta(\lambda)\big) \times \pi_d^*(\rho_d) - \sum_{c,d} \rho_{(\cdot)} - \lambda$$

$$\text{where } \lambda \equiv \begin{cases} \lambda_f & \text{if } \Delta\pi > 0 \\ \lambda_a & \text{if } \Delta\pi < 0. \end{cases}$$

That is, the firm chooses innovation and lobbying investments in order to maximize the expected value of second stage profits net of first stage investment costs. Here, $\Delta\pi = \pi_c^* - \pi_d^*$ is the profit wedge between regulatory states, and λ describes the firm's endogenous direction of lobbying as detailed below. The first order conditions are given by: $\eta\pi_c^{*\prime} = 1$, $(1-\eta)\pi_d^{*\prime} = 1$, and $\Delta\pi\partial\eta/\partial\lambda = 1$. The first two FOCs state that the firm invests in clean and dirty innovation until the expected marginal payoff equals the marginal cost of the investment. The last FOC embeds the firm's endogenous decision to lobby for or against regulation. Because the terms $\Delta\pi$ and $\partial\eta/\partial\lambda_l$ multiply to unity at the optimum, they must be the same sign. Thus, if the firm is a "regulatory loser" ($\Delta\pi < 0$) then it lobbies against regulation ($\partial\eta/\partial\lambda_a < 0$). Conversely, if the firm is a "regulatory winner"[13] ($\Delta\pi > 0$) then it lobbies for regulation ($\partial\eta/\partial\lambda_f > 0$). Thus we can define λ as above based on the sign of $\Delta\pi$.

Cross-partial derivatives on expected profits shed insight on whether a firm's innovation and lobbying are complements or substitutes in this model. Take the case of a firm for which the regulatory profit wedge $\Delta\pi < 0$ (a regulatory loser), so that the firm lobbies against regulation. In this case, $\frac{\partial^2 E[\pi]}{\partial\lambda_a\partial\rho_c} = \frac{\partial\eta}{\partial\lambda_a}\frac{\partial\pi_c^*}{\partial\rho_c} < 0$, and $\frac{\partial^2 E[\pi]}{\partial\lambda_a\partial\rho_d} = -\frac{\partial\eta}{\partial\lambda_a}\frac{\partial\pi_d^*}{\partial\rho_d} > 0$. That is, lobbying and clean innovation are substitutes, while lobbying and dirty innovation are compliments. The intuition is straightforward: lobbying against regulation shifts probability mass out of the regulated state where ρ_c pays off, and into the unregulated state where ρ_d pays off. The opposite patterns holds if $\Delta\pi > 0$, so that the firm lobbies for regulation.

Comparative statics Empirically, this study examines firm responses to shocks over the probability of future regulation. To conduct associated comparative statics, I introduce a minor adjustment to the conceptual framework, making $\eta(\cdot)$ a function of a common prior η_0 over the background probability of new regulation in any market. That is, consider $\eta = \eta\big(\lambda, \eta_0\big)$. Shocks

[13]While this model excludes a strong Porter hypothesis, $\Delta\pi > 0$ could still result if, e.g., the regulatory threat sufficiently increases second-period demand for good c relative to good d. In a multi-firm context, profit wedges can also be positive if regulation harms competitors more than the focal firm (Salop and Scheffman, 1983).

to the probability of future regulation can now be modeled as shocks to η_0, yielding

$$\frac{\partial \rho_c^*}{\partial \eta_0} = KA\frac{\eta'}{\eta}\frac{\pi_c^{*\prime}}{\pi_c^{*\prime\prime}}\Delta\pi\frac{\partial \eta}{\partial \eta_0}$$

$$\frac{\partial \rho_d^*}{\partial \eta_0} = -KA\frac{\eta'}{1-\eta}\frac{\pi_d^{*\prime}}{\pi_d^{*\prime\prime}}\Delta\pi\frac{\partial \eta}{\partial \eta_0}$$

$$\frac{\partial \rho_{tot}^*}{\partial \eta_0} = \frac{\partial(\rho_c^* + \rho_d^*)}{\partial \eta_0} = KA\left(\frac{\eta'}{\eta}\frac{\pi_c^{*\prime}}{\pi_c^{*\prime\prime}} - \frac{\eta'}{1-\eta}\frac{\pi_d^{*\prime}}{\pi_d^{*\prime\prime}}\right)\Delta\pi\frac{\partial \eta}{\partial \eta_0} \tag{1}$$

$$\frac{\partial \lambda^*}{\partial \eta_0} = A\left[\frac{\partial^2\eta/\partial\lambda\partial\eta_0}{\partial\eta/\partial\eta_0} - \left(\left(\frac{\pi_c^{*\prime}}{\Delta\pi}\right)^2\frac{\pi_c^{*\prime}}{\pi_c^{*\prime\prime}} + \left(\frac{\pi_d^{*\prime}}{\Delta\pi}\right)^2\frac{\pi_d^{*\prime}}{\pi_d^{*\prime\prime}}\right)\right]\Delta\pi\frac{\partial \eta}{\partial \eta_0} \tag{2}$$

$$K \equiv \frac{\eta''}{\eta'} - \frac{\partial^2\eta/\partial\lambda\partial\eta_0}{\partial\eta/\partial\eta_0}$$

where $A > 0$, and primes refer to derivatives with respect to the firm's relevant first-stage choice variable. These expressions can be signed (Appendix A) as follows.[14] For a *regulatory winner*, clean innovation and lobbying increase with the regulatory shock ($\frac{\partial \rho_c^*}{\partial \eta_0}, \frac{\partial \lambda^*}{\partial \eta_0} > 0$) while dirty innovation decreases ($\frac{\partial \rho_d^*}{\partial \eta_0} < 0$). Further, clean innovation gains dominate dirty losses in the empirically-relevant parameter space[15] so that total innovation investments increase.

For a *regulatory loser*, comparative statics are ambiguous as the exogenous shift of probability mass into the regulated state competes with firm lobbying against regulation. This competition is captured by the parameter K, which differences curvature in the lobbying function against a "curvature-like" measure of the relative change in the marginal effect of lobbying operating through the shock η_0 to η. For $K < 0$, innovation results follow those of the regulatory winner (for whom $K < 0$ always), including the increase in total innovation. The sign of the lobbying comparative static is ambiguous, and will be positive if and only if the term in square brackets in Eq. 2 is negative.[16] However, if the shock to η_0 has a large relative impact on

[14] I assume $\text{sign}(\partial^2\eta/\partial\lambda\partial\eta_0) = \text{sign}(\eta')$, which occurs if the shock to η_0 causes the regulator to become more responsive to lobbying in general (e.g. because the regulator is "more uncertain" about regulation). This is consistent with the empirical setting of this paper, where the regulator is unlikely to have a scientific basis for new regulation before the first discovery of a health harm.

[15] The term in parenthesis in Eq. 1 is likely negative in the empirical context of this study. The pre-shock probability of regulation η is likely small, and pre-shock curvature in clean innovation payoffs is likely larger in magnitude than that of dirty innovation (due to low clean innovation investments relative to dirty innovation), both of which lead to $\frac{\eta'}{\eta}\frac{\pi_c^{*\prime}}{\pi_c^{*\prime\prime}} < \frac{\eta'}{1-\eta}\frac{\pi_d^{*\prime}}{\pi_d^{*\prime\prime}} < 0$.

[16] That is, if the "curvature-like" effect of the shock to η_0 is larger in magnitude than the sum of the clean

η' (such that $K > 0$), the regulatory loser unambiguously increases lobbying while decreasing clean innovation and increasing dirty innovation. Total innovation likely decreases in this case.

These results are summarized in Table 1, which focuses on total innovation as reported in the Compustat data. Two additional details merit brief mention. First, the magnitude of the total innovation comparative static will be heterogeneous in firm lobbying returns (Eq. 1) and likewise for the lobbying comparative static with respect to firm innovation returns (Eq. 2). However, the direction of the heterogeneity cannot be generally signed in either case. Second, lobbying by an outside interest group shifts the probability of regulation for all parties. Thus comparative statics over η_0 can also be interpreted as spillover effects from exogenous external lobbying. Empirical tests suggest these factors are present in the data (Appendix E).[17]

	Conceptual framework predictions			
	Regulatory winner	Regulatory loser		
		Strong shock	Weak shock	Finding
Main effects				
$\partial\rho^*_{tot}/\partial\eta_0$ †	+	−	ambiguous	Fig. 2
$\partial\lambda^*/\partial\eta_0$	+	+	ambiguous	Fig. 3
Complementarities				
ρ_c, λ	+	−	−	n/a
ρ_d, λ	−	+	+	n/a
ρ_{tot}, λ	ambiguous	ambiguous	ambiguous	Figs. 4, 5††
Spillovers				
Lobby spillovers to R&D	yes	yes	yes	App. E
R&D spillovers to lobbying	n/a	n/a	n/a	App. E

Table 1: **Conceptual framework predictions and empirical findings.** Cases: "Regulatory winner" ($\Delta\pi > 0$), "regulatory loser" ($\Delta\pi < 0$), "strong shock" ($K > 0$), "weak shock" ($K < 0$). The direction of lobbying, λ, changes with the sign of $\Delta\pi$. R&D spillovers marked "n/a" as they are not modeled for a monopolist. See text for details.
†*Note: predictions in the first two columns of this row are for the empirically-relevant parameter space.*
††*Note: Figures 4 & 5 test for innovation and lobbying complementarities in firm responses to an η_0 shock which, like the first-stage complementarities, are formally ambiguous.*

and dirty innovation payoff curvatures, weighted by the squares of the respective marginal innovation returns as a fraction of the profit wedge.

[17]Innovation spillovers are excluded in a monopolist model, and empirical tests for innovation spillovers to firm lobbying responses are inconclusive (Appendix E).

3 Constructing a novel dataset of scientific discoveries

As discussed previously, regulatory events are often announced years in advance of their implementation, leading forward-looking firms to respond and potentially biasing research designs around the timing of regulatory proceedings (Rittenhouse and Zaragoza-Watkins, 2018). To address this concern, I construct a novel dataset of exogenous shocks to the probability of future regulation: initial scientific discoveries by external scientists of previously unknown harms from commonly used industrial chemicals. These discoveries result from the work of academic and public-sector epidemiological researchers, not private sector researchers.[18] Construction of this dataset required systematically scraping the epidemiological literature for publications of potential low-dose, chronic-exposure human health harms from chemicals, and reviewing over 7,000 individual publications to identify 104 initial discoveries of previously unknown human health harms. I discuss the data set construction in detail here.

3.1 Institutional context for introducing new chemicals in the U.S.

One might ask: why were the low-dose harms whose discoveries I document not already known at the time a given chemical was introduced into the economy? Under Section 5 of the 1976 Toxic Substances Control Act (TSCA), the EPA was to regulate chemicals if it *"...finds that there is a reasonable basis to conclude that... a chemical substance or mixture... presents or will present an unreasonable risk of injury to health or the environment..."* (15 USC 2605(a)).[19] This may sound like a strong regulatory mandate, however several shortcomings in the TSCA new chemical review process rendered it largely ineffective (Schmidt, 2016; Gerlach, 2016). First, the 62,000 chemicals already in industrial use in 1976 were exempted from testing and were simply grandfathered in under the act. Second, EPA had limited ability to mandate that firms generate toxicity data for new chemicals. EPA could only require toxicity data if a chemical potentially presented an "unreasonable health risk" based on existing data. But there is no such existing data for new chemicals, and EPA could not require such data from firms —

[18] All but one of the scientific discoveries I identify come from the epidemiological research community, not private sector researchers. Results are robust to dropping any discovery.

[19] https://www.govinfo.gov/content/pkg/USCODE-2009-title15/html/USCODE-2009-title15-chap53-subchapI-sec2605.htm

who have no private incentive to produce it.[20] Thus, there was often insufficient data to prove the potential for an unreasonable health risk. Third, if EPA did not make a ruling on a new chemical within a 90-day period following notification, that chemical entered the economy by default without further regulatory requirements.[21]

By 2011, EPA had issued regulations under TSCA covering only nine of the more than 85,000 chemicals in circulation in the U.S. economy (Schmidt, 2016). Low-dose harms from new or grandfathered (pre-1976) chemicals were not required to be studied, and the U.S. has been called the "Wild West" of chemical regulation (Gerlach, 2016). Given this regulatory context, it has been left to independent epidemiologists to identify these low-dose health effects, and so I draw from that literature for my dataset construction.

3.1.1 How epidemiologists target chemicals for study

In documenting the human health impacts of chemical exposure, the epidemiological literature typically forms a key component of the scientific basis for environmental regulation of pollution externalities. Thus, the epidemiological community openly documents how they target chemicals for study. For example, the Centers for Disease Control and Prevention (CDC) produces an annual report on chemicals exposure in the U.S. population (CDC, 2022), where they discuss their chemical selection for study on the basis of data suggesting exposure in the population, potential seriousness of health effects, costliness of testing and monitoring exposure, potential efficacy of public health interventions, and availability of testing methods and samples. Notably, chemicals are not targeted for study based on characteristics of the firms that produce or use them, business cycles, or even the availability of substitutes.

3.2 Scraping and evaluating scientific publications

The Toxics Release Inventory (TRI; EPA, 2021) provides an independent list of chemicals to scrape. The search phrase I use requires that the chemical name be present in each matching

[20] Indeed, given firms face legal exposure if they know of a health harm and do not disclose it, firms actually have a strong disincentive to even investigate low-dose harms from their products.

[21] In 2016, forty years after its initial passage, Congress amended TSCA to address the grandfathering and data requirement issues discussed above (see Denison (2017) for a detailed overview). The discoveries I study here relate to chemicals introduced into the economy before the 2016 TSCA amendments. And, while improved screening capacity by EPA should improve detection of health harms, there is no reason to believe such detection will be perfect.

publication. The search phrase gives additional weight to exact matches and near matches of the following set of terms: *endocrine, estrogenic, carcinogen, cancer, chronic, toxic, human, health*. The last six terms are self-explanatory. The first two terms "endocrine" and "estrogenic" describe chemicals that mimic human hormones and their interactions with the human hormonal ("endocrine") system, which can lead to various cancers (e.g. breast cancer) and other health effects. I analyze TRI chemicals in reverse order from their date of addition to the list, and have reviewed results for 78 chemicals.

This process returned the top 100 hits for each chemical, giving a total of 7,800 scientific publications. These 100 findings for each chemical were reviewed in chronological order – from oldest to newest – and the *first* finding was reported of either: 1) a low-dose, chronic exposure health effect in laboratory animal testing, or 2) a human health effect associated with routine chemical use following manufacturer guidelines. For laboratory testing, I define "low dose" as under 1,000 ppm (about 65 mg/kg in rats and 140 mg/kg in mice) and "chronic" as a minimum of four weeks of laboratory exposure,[22] though many studies last for one to two years.

This process resulted in 93 scientific discoveries of novel, low-dose, chronic exposure health effects. These 93 discoveries exceed the number of TRI chemicals reviewed (78 chemicals) because some chemicals have distinct discoveries of different harms associated with them. For example, public health researchers at University of California San Francisco discovered in 1994 that California agricultural workers in routine workplace environments were being exposed to the chemical cyanazine (CAS No. 21725-46-2) in quantities that represented a substantial fraction of the estimated LD_{50} dosage – the exposure level that would be lethal for humans 50% of the time (Woodruff, Kyle, and Bois, 1994). Later, in 2005, a researcher at the contract research firm Covance — who was also the Editor-in-Chief of Journal of Applied Toxicology — concluded that a category of breast-cancer-related findings in rat studies, which had widely been viewed as irrelevant for humans, should be re-evaluated for their risk to humans (Harvey, 2005). Harvey (2005) extensively discusses cyanazine as an example requiring such re-evaluation for its breast cancer risks in humans.[23] So, cyanazine in my data has two discoveries associated

[22]For reproductive studies in rats and mice only one week of exposure is permitted, since total gestation is only about 21 days.

[23]Harvey (2005) states (emphasis added): "Prolactin-induced mammary carcinogenesis in rodents, particularly rats, is often stated to be of low toxicological relevance to humans. This opinion appears to have developed from a number of lines of cited evidence... However, recent evidence now suggests that prolactin has a major role in human breast cancer, and the similarity of mechanism with the rodent suggests that prolactin-mediated

it. One in 1994 related to the discovery of toxicity concerns associated with routine use of the chemical by agricultural workers, and another in 2005 associated with low-dose breast cancer in which an Editor-in-Chief sought to overturn common and recent conclusions that specific findings in rat studies were not relevant to humans.

Finally, an additional 11 discoveries compiled from early work on this project are also included in the analysis (see Appendix G), for a total of 104 (93 + 11) initial scientific discoveries of previously unknown health effects from commonly used industrial chemicals.

Matching discoveries to firms For competitive reasons, firms do not generally disclose details of their chemical inputs or outputs. Thus, I match chemicals to granular industries. Then, in the Compustat data, I create a measure of firm-specific exposure using Compustat Segment data on firm sales by industry.[24] The chemicals database PubChem[25] contains a section titled "Use and Manufacturing" which lists the industries in which a particular chemical is used. I draw on this field to match each chemical to the industry identifiers available in the Compustat data (four-digit SIC codes) and in the Center for Responsive Politics (CRP) lobbying data ("catcode" industry codes). Across all industries – not just those affected by the scientific discoveries I identify – there are 312 four-digit SIC codes and 459 lobbying catcodes, so the two sets of industry definitions are of similar granularity.

3.3 Data sources for firm innovation and lobbying investments

Data on firm innovation and lobbying outcomes used in this study have been used in a wide range of previous studies (see Arora, Belenzon, and Sheer, 2021; Bombardini and Trebbi, 2020, for relevant reviews). Details relevant to this analysis are described below.

mammary carcinogenesis **in rodents could be of much higher toxicological relevance to humans than previously thought**... For example, there have been a number of recent studies on chloro-s-triazine herbicides (e.g. cyanazine...) where the authors suggest that the mechanism of mammary tumours in the rat... is 'thought to have low relevance to humans' (Bogdanffy et al., 2000) or of 'no biological relevance to humans' (Stevens et al., 1999)... It should be pointed out that, regardless of any species- or rat-strain-specific effect of cyanazine on luteinizing hormone (LH), an effect directly on prolactin through dopamine inhibition may be relevant to humans at least in inducing prolactin secretion..."

[24]This exposure measure follows Bloom, Schankerman, and Van Reenen (2013).

[25]https://pubchem.ncbi.nlm.nih.gov

Innovation data Firm innovation data were obtained from the Compustat Fundamentals dataset, accessed through Wharton Research Data Services. This dataset contains firm-level annual observations of key financial data from 1950–2021.[26] The reported financial data are wide-ranging and include innovation expenditures (XRD), market capitalization ($MKVALT$)[27], net income (NI), and sales ($SALE$) as well as firm four-digit SIC sector codes (SIC). The data cover all publicly-traded firms in the U.S. and are compiled from required SEC 10-K filings. Mergers are tracked and a common identifier for the parent company persists ($GVKEY$), allowing the researcher to track firms through mergers and name changes over time. All data are adjusted for inflation using the Bureau of Labor Statistics Consumer Price Index (CPI).

Firm innovation investments exhibit both many zeros and a long right tail, so I normalize a firm's innovation spending by its pre-period market capitalization, following Chan, Lakonishok, and Sougiannis (2001).[28] To avoid capturing changes in firm market capitalization when estimating treatment effects, I normalize treated firms by their pre-period average market capitalization over the two years ending before the *first* scientific discovery affecting them. Thus, no information from any treatment enters the normalization for these firms. Untreated firms do not have a clearly defined "baseline" period and so are normalized by their panel-average market capitalization. See Section 6.3 for additional discussion and robustness to alternative normalizations. I winsorize outlier values of firm innovation (in first differences) at the top and bottom 0.5 percentiles of the distribution.

Lobbying data Firm lobbying data were obtained from the Center for Responsive Politics (CRP). In the U.S., any lobbying activities in excess of $200 have been legally required to be reported semi-annually to the Senate Office of Public Records (SOPR). CRP has collected this data from SOPR back to 1998, shortly after lobbying reporting requirements went into effect. I primarily use the CRP "Lobbying" data file in this analysis. These data specify the lobbyist ($REGISTRANT$), the client firm paying for the lobbying ($CLIENT$), the amount spent on

[26] Quarterly data are also available, starting in 1989. Most firms report in the annual dataset, but some firms only report in the quarterly file. Data for these firms were annualized and included in this analysis.

[27] Where $MKVALT$ data are missing in the Compustat data, values are calculated as common shares outstanding ($CSHO$) $\times$ fiscal year closing share price ($PRCC\_F$), following common practice.

[28] Normalizing innovation expenditures by sales is another alternative in the literature, but criticized by Jaffe and Palmer (1997) when used across industries with varying levels of market power. Jaffe and Palmer (1997) study industry aggregate innovation and so are able to take logs; the many zeros in the firm-level panel prevent me from doing this.

lobbying ($AMOUNT$), and a five-character industry identifier ($CATCODE$) which is similar in granularity to an SIC-4 industry identifier. Importantly, the $CLIENT$ identifier is designed to uniquely identify firms over time, similar to the $GVKEY$ identifier in the Compustat data. The lobbying data range from January 1998 to June 2021. All financial data are adjusted for inflation using the Bureau of Labor Statistics CPI data.

Like the innovation data, firm lobbying data exhibit both many zeros and a long right tail. However, unlike the innovation data, the lobbying data only contain information on firm lobbying expenditures. No other firm covariates (e.g. market capitalization, net income, or sales) are present to account for scale effects. This leads me to normalize a firm's lobbying spending by the only measure available: the firm's average pre-period level of lobbying. I avoid changes in the denominator when estimating treatment effects by normalizing treated firms by their pre-period average lobbying spending over the two half-year periods ending before the *first* scientific discovery affecting them. Thus, no information from any treatment enters the normalization for these firms. Untreated firms do not have a clearly defined "baseline" period and so are normalized by their panel-average lobbying expenditures. See Section 6.3 for additional discussion and robustness to alternative normalizations. Consistent with the approach in the innovation data, I winsorize outlier values of firm lobbying (in first differences) at the top and bottom 0.5 percentiles of the distribution.

Institutional implications of U.S. lobbying for information spillovers It is important to note that regulatory proposals often affect many different firms across multiple sectors of the economy, and the public and frequent nature of lobbying expenditure reporting in the U.S. means firm lobbying activities are common knowledge. Such a situation invites coordination among firms with a common lobbying position, and industry trade groups exist in the U.S. in part for such a purpose (e.g., Bombardini and Trebbi, 2012, 2020). Indeed, Bombardini and Trebbi (2012) report that for the 1999–2001 period, trade association lobbying made up roughly 30% of all industry lobbying in the U.S. The public nature of U.S. lobbying reporting and the common role of industry trade groups in coordinating actual firm lobbying activities means that lobbying spillovers are likely important in the U.S. context. In particular, such spillovers are likely to be present in empirical estimates of firm trade-offs between lobbying and other

potential actions (e.g. innovation). The estimates in Appendix E of this study are constructed to test for such information spillovers.

4 Descriptives: Scientific discoveries and firm outcomes

The compilation of first-time discoveries of chemical harms results in a novel dataset of 104 discoveries affecting 60 industries, with discoveries occurring in 44 distinct years and affecting over 400 firms. Here I present an overview of these data, as well as a descriptive discussion of the cross-sectional variation in firm innovation and lobbying investments, both at highly aggregated (SIC division) as well as disaggregated (firm-specific) levels.

Descriptive data on discoveries of potential externalities Table 2 displays decadal and industry patterns in the 104 discoveries of chemical harms. The average rate of new discoveries increased from the 70s through the 90s, then stabilized for a decade in the 2000s before slowing in the 2010s (column 2). Economic growth, regulatory regimes, and market structures have all also varied over this period of expanding knowledge of chemical harms. Addressing these temporal trends in estimation will be critical. However, it is also noteworthy that discoveries of chemical harms are not heavily clustered in any one period, but rather exhibit substantial interannual variation (Appendix Figure B.1). Additionally, a wide range of SIC4 industries were affected by these discoveries across decades (column 3). For example, SIC2 sectors – which themselves each encompass many SIC4 industries – range from crop production (sector 01) to oil & gas extraction (sector 13) to electronics manufacturing (sector 36), to choose a few examples among many. A corresponding range of "catcode" industries in the CRP data on firm lobbying were also affected (column 4). Finally, hundreds of firms in both the Compustat (column 5) and CRP (column 6) datasets were affected by these discoveries in each decade data are available.[29]

Cross sectional variation in innovation and lobbying Table 3 displays highly aggregated firm financial, innovation, and lobbying data (columns 1–4) averaged by SIC division. Averages across all SIC divisions are reported in the bottom row. Columns 5–6 report SIC-division

[29]Note that lobbying reporting only begins at the very end of the 1990s.

Decade	Count of Discoveries	Industries affected: Compustat	Industries affected: CRP	# Firms affected: Compustat	# Firms affected: CRP
1970 – 1979	9	0100, 2822, 2834, 2851, 2865, 2869, 2879, 2891, 2899	N/A†	123	N/A†
1980 – 1989	20	0100, 2295, 2821, 2822, 2823, 2834, 2843, 2844, 2851, 2865, 2869, 2879, 2891, 2892, 2899, 2911	N/A†	183	N/A†
1990 – 1999	33	0100, 0111, 0115, 0116, 0131, 2086, 2273, 2295, 2821, 2822, 2834, 2841, 2843, 2844, 2851, 2865, 2869, 2873, 2879, 2892, 2899, 2911, 3111, 3411, 3471, 3674, 3842	N/A†	327	N/A†
2000 – 2009	29	0100, 0111, 0115, 0116, 0131, 1381, 1382, 1389, 2819, 2821, 2823, 2834, 2841, 2842, 2843, 2844, 2851, 2869, 2879, 2891, 3087, 3312, 3331, 3334, 3339, 3674, 3679	A3100, A4100, C5000, C5100, C5200, C5400, E1150, G5500, M1400, M1500, M1600, M1700, M2100, M2200, M2250, M3000, M8000	298	238
2010 – 2019	13	0100, 1311, 2819, 2821, 2834, 2843, 2844, 2869, 2879, 2891, 2899, 3087, 3674, 3679	A4100, C5000, C5100, C5200, C5400, H4300, M1000, M1600, M1700, M8000	238	252

Table 2: **Novel discoveries of potential externalities and the industries and firms they affect.** The rate of discovery varies by decade (column 2). A wide range of SIC4 industries were affected by these discoveries (column 3), with SIC2 sectors including crop production (sector 01), oil & gas (sector 13), and manufacturing of: food products (sector 20), textiles (sector 22), chemicals (sector 28), petroleum refining (sector 29), rubber and plastics (sector 30), leather (sector 31), primary metals (sector 33), fabricated metals (sector 34), and electronics (sector 36). A corresponding range of "catcode" industries in the CRP data on firm lobbying were also affected (column 4). In each decade, hundreds of firms in both datasets (columns 5 & 6) were affected by these discoveries. The counts of treated firms reported here reflect firms with sufficient observations to enter the estimating sample.

†Note: The lobbying data start in 1998, so no treated firms enter the data before the year 2000 as none have the necessary number of leads for estimation (see Section 6).

average innovation and lobbying expenditures both normalized by SIC-division average market capitalization. Rows are ordered from the highest to the lowest fraction of market capitalization committed to innovation expenditures (that is, rows are decreasing in column 5).

Several patterns emerge in this data. First, firm average innovation and lobbying (Table 3, columns 3 and 4) both vary substantially by SIC division, but are not highly correlated with each other. Some divisions have large annual innovation and lobbying investments (e.g. Manufacturing), some have high levels of innovation but relatively low levels of lobbying (e.g. Services), some have low levels of innovation but high levels of lobbying (e.g. Finance, Insurance, & Real Estate), and some are low in both measures (e.g. Construction). Second, divisions with higher market capitalizations appear to innovate more than those with lower market capitalizations (columns 2 and 3); however, this correlation is not as strong in the sales data (columns 1 and 3). This descriptively suggests innovators may chase, or seek to protect, rents. Such patterns are harder to discern in firm lobbying investments, though the more limited variation in division average lobbying may be a factor. Third, innovation and lobbying intensities (here reported as ratios of SIC division averages, columns 5–6) appear to be negatively correlated. Innovation intensities are generally higher for low lobbying intensity divisions, and vice versa.

While columns 5 and 6 of Table 3 characterize broad patterns in the data with average innovation and lobbying intensities, such averaging omits within-division, firm-level correlations in innovation or lobbying investments and their market capitalizations. Figure 1 reports firm-level averages of innovation and lobbying intensities for a subset of firms that can be matched across the separate Compustat and CRP datasets (see Appendix E). The figure separates firms in 4-digit SIC industries that are ever exposed to a scientific discovery in the data ("Treated") from those that are not ("Untreated"). Two noteworthy features appear. First, treated and untreated firms are evenly represented across levels of normalized innovation and lobbying investments. Second, the negative correlation between normalized innovation and normalized lobbying observed in aggregate in Table 3 still appears to hold for both treated and untreated firms in this firm-matched subset of the data. This finding is qualitatively similar to Akcigit, Baslandze, and Lotti (2023)'s descriptive analysis of Italian firm patenting and politician-hiring, here suggesting that in the U.S. context both treated and untreated firms may treat innovation and lobbying as substitutes.

SIC Div.	(1) Sales ($M)	(2) Mkt Cap ($M)	(3) R&D ($M)	(4) Lobby ($k)	(5) R&D/MktCap (—)	(6) Lobby/MktCap ($\times 10^{-3}$)
Manufacturing	2,361	3,122	80	188	0.026	0.060
Services	1,005	2,808	51	93	0.018	0.033
Retail Trade	4,129	3,460	27	249	0.008	0.072
Ag, Forestry, and Fishing	932	539	4.0	115	0.007	0.214
Transportation & Public Utilities	3,129	4,502	14	157	0.003	0.035
Mining	1,430	2,474	7.0	216	0.003	0.088
Wholesale Trade	2,775	955	1.5	92	0.002	0.097
Construction	1,327	771	0.5	78	0.001	0.101
Finance, Insurance, & Real Estate	1,903	2,062	0.5	198	0.0002	0.096
All Firms	2,217	2,850	41	111	0.014	0.039

Table 3: **Descriptive statistics of firm financial and lobbying data.** Average sales, market capitalization, R&D expenditures, and lobbying expenditures by SIC division displayed in columns (1)–(4). SIC-division-average R&D and lobbying, both normalized by SIC-division-average market capitalization, displayed in columns (5)–(6). The average scale of lobbying in the economy is about two orders of magnitude smaller than the average scale of innovation (columns (3) and (4), bottom row). SIC divisions are ordered by normalized R&D expenditures (column (5)), highlighting the negative correlation between division-wide scale-normalized innovation and lobbying. The "Public Administration" division was excluded because it consists of government entities and other uncategorized firms.

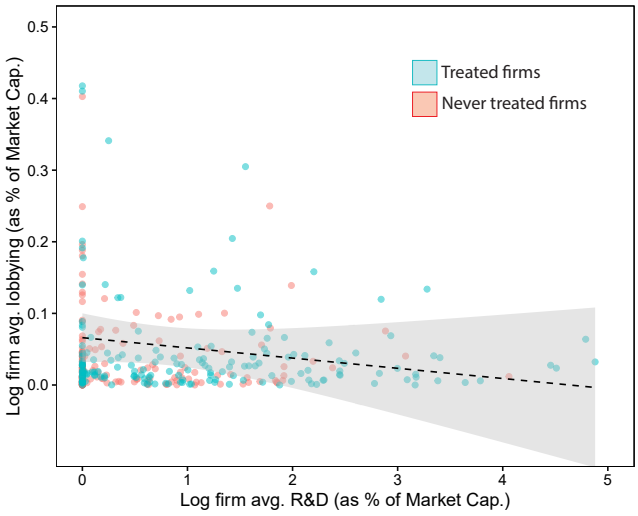


Figure 1: **Cross sectional variation in firm R&D and lobbying investments**, for a subset of firms that can be matched across the Compustat and CRP datasets (see Appendix E). Normalized R&D and lobbying are averaged by firm. The plot displays $\log(100 \times \textit{firm average value} + 1)$. A firm is "treated" (blue) if it is affected by at least one discovery of a previously-unknown chemical harm, and is "untreated" (red) otherwise. Dashed line: linear fit with 95% confidence interval.

5 Empirical specification

This section describes the main empirical specifications employed in the analysis. Because firm responses may in practice be dynamic, I begin by grounding the specifications in a simple dynamic description of firm investments. Importantly, because firms can be exposed to multiple different information shocks over time, the final model (Eq. 3) is estimated as a distributed lag model and not an event study. The distributed lag approach explicitly allows for overlapping impulse-responses to multiple treatments over time (Schmidheiny and Siegloch, 2023),[30] which is a feature of the data in this context.

5.1 Estimating the innovation and lobbying responses

Comparative statics in Section 2 describe an overall ambiguous response of firm innovation and lobbying investments to a regulatory shock (Table 1), motivating empirical estimation. This section grounds the main econometric specification in a simple description of firm innovation or lobbying investments, which may be dynamic and which scale with some measure of firm size:

$$Y_{it} = \rho Y_{it-1} + SCALE_{it} \times e_{it}.$$

Here, Y_{it} represents the level of innovation or lobbying investment by firm i in year t, $SCALE_{it}$ is a measure of firm size, and e_{it} is a disturbance term. Such a model might arise, for example, if firm managers take last year's innovation budget as an input into planning the subsequent year's budget, but respond to aggregate shocks (such as changes in macroeconomic or regulatory conditions) in a way that scales with the size of the firm.[31]

The disturbance term e_{it} can be further decomposed into a firm-specific regulatory shock r_{it} (e.g., an information shock over the possibility of future regulation), a set of sector-by-year shocks γ_{st} reflecting both stochasticity in the economy as well as sectoral regulatory trends and

[30] As noted previously, Schmidheiny and Siegloch (2023) propose a modified event study specification that does accommodate multiple treatments, but they show that their modified design is equivalent to estimating a distributed lag model.

[31] Estimation in logged scale-adjusted outcomes is typical, at least in the innovation literature where it is based in a CES production function (e.g. Bloom, Griffith, and Van Reenen, 2002). However, both the lobbying and R&D expenditures data contain many zeros at the firm level, requiring in logs that these firms be dropped or imposing other adjustments to the data. I avoid such issues by estimating in scale-adjusted levels instead.

shocks, and a noise term u_{it}. This yields the estimating equation

$$Y_{it} = \rho Y_{it-1} + SCALE_{it} \times \Big(\beta r_{it} + \gamma_{st} + u_{it} \Big).$$

In the literature on firm innovation, it has been difficult to reject a unit root in the data generating process (e.g., Okunade and Murthy, 2002; Apergis, Economidou, and Filippidis, 2008), and I cannot reject a unit root in this analysis for either firm innovation or lobbying expenditures.[32] This leads me to first-difference innovation and lobbying expenditures, to eliminate potential bias from a unit root process (Greene, 2003). Additionally, the objective is to isolate the firm response to the regulatory shock r_{it} for estimation, so I normalize the differenced outcome by firm scale. However, any measure of firm scale may itself be affected by an information shock over the firm's future regulatory exposure, changing the outcome variable (through its denominator) even if actual investments are held constant. Thus, I normalize by $\overline{SCALE}_i$, the average of $SCALE_{it}$ over the two periods directly preceding the *first* regulatory shock faced by each firm,[33] yielding

$$\frac{\Delta Y_{it}}{\overline{SCALE}_i} = \beta \Delta r_{it} + \phi_{st} + \epsilon_{it}$$

where ϕ_{st} and ϵ_{it} represent first differences in γ_{st} and u_{it}. Finally, two additional points remain. First, firm responses to Δr_{it} may evolve over multiple periods. Such a dynamic response might occur if, for example, new innovation or lobbying programs require time-consuming adjustments like developing new product offerings or influence networks. I employ a distributed lag estimator to capture the dynamics of firm responses. This estimator flexibly allows for either transitory or permanent effects of information shocks on firm innovation or lobbying investments. It additionally allows for overlapping impulse-response functions in estimation (Schmidheiny and Siegloch, 2023). Second, firms may be differentially exposed to a given information shock. I

[32] An augmented Dickey-Fuller test with drift and a time trend fails to reject a unit root in R&D expenditures for 88% of firms at $p < 0.05$, and still fails to reject a unit root at $p < 0.10$ for 81% of firms. The failure to reject a unit root increases if one or two lags in the dependent variable are added. The failure to reject a unit root is even higher for firms' lobbying expenditures (e.g. 91% of firms fail to reject a unit root at $p < 0.05$), and this holds regardless of lag specification.

[33] As discussed in Section 3.3, I scale innovation investments by the firm's two-year pre-period average market capitalization, which I choose instead of sales to account for differences in pricing power across firms and industry sectors (Jaffe and Palmer, 1997). I correspondingly scale lobbying investments by two half-year (12-month) pre-period average lobbying, which is the only information about the firm available in that dataset.

capture treatment intensity in the R&D data using Compustat data on firm net sales across SIC 4-digit industries. Treatment intensity in firm innovation responses is measured as the fraction of a firm's average net sales occurring in industries affected by a given information shock:

$$INTENSITY_{it} = \sum_{SIC4} \left(\overline{SALES}_{SIC4,i} \times TREAT_{SIC4,t}\right) / \sum_{SIC4} \overline{SALES}_{SIC4,i}.$$

This information is not available in the lobbying panel of data, so treatment intensity is set to 1 for all shocks in the lobbying response. The final empirical specification for firm i in industry sector s and period t is given as[34]

$$\Delta y_{it} = \sum_{\ell \in b} \underbrace{\beta_\ell LAG_{\ell it}}_{\text{dynamic "}\beta \Delta r_{it}\text{"}} + \phi_{st} + \epsilon_{it}, \tag{3}$$

$$\text{with } LAG_{\ell it} \equiv \sum_{\tau_{ij} \in trt\_years} \mathbb{1}\{t - \ell = \tau_{ij}\} \times INTENSITY_{it}.$$

Here, y_{it} either represents $SCALE$-normalized firm annual innovation or biannual lobbying investments, b represents the bandwidth (set of leads and lags) over which lagged effects β_ℓ are estimated, and s indexes SIC 2-digit industries in the Compustat data and sector codes in the CRP data. In the definition of $LAG_{\ell it}$, trt_years denotes the set of j treatment years (discovery years) $\{\tau_{i1}, \tau_{i2}, ...\}$ for firm i, $\mathbb{1}\{t - \ell = \tau_{ij}\}$ is an indicator for whether period t is ℓ lags from the j^{th} treatment year in the set $\{\tau_{ij}\}$, and $INTENSITY_{it}$ is treatment intensity as described above.[35] Firm fixed effects in levels are implicitly captured by first-differencing the outcome, purging the data of any time-invariant differences between firms. The term ϕ_{st} absorbs any remaining sector-wide shocks to innovation or lobbying expenditures, such as sectoral market fluctuations or sectoral shocks to the overall regulatory environment. The coefficients of interest, β_ℓ, are estimated off of plausibly exogenous variation in within-firm, within-sector-year external shocks to scientific knowledge of chemicals externalities, comparing the trajectories of treated firms to those of similar, untreated firms before and after an information shock. Standard

[34] Note that this specification is entirely consistent with the first-differenced distributed lag specification laid out by Schmidheiny and Siegloch (2023), though the notation differs. Here, LAG captures changes in treatment status by period.

[35] Typically, $\sum_{\tau_{ij} \in trt\_years} \mathbb{1}\{t - \ell = \tau_{ij}\}$ is either 0 or 1. Either the firm was newly exposed to a discovery of a low-dose harm ℓ lags ago, or it wasn't. However, this term can be greater than 1 if a given firm is exposed to more than one discovery in the same year.

errors are cluster-robust by firm, allowing for arbitrary correlations in innovation and lobbying activities at the firm level over time.[36]

Because the model is estimated in first differences, the cumulative effect Ω_k of a shock on an outcome at some time k periods after the shock occurred is given by

$$\Omega_k = \sum_{\ell=0}^{k} \beta_\ell \ . \tag{4}$$

Results in Section 6 are plotted in terms of these cumulative effects Ω_k, with uncertainties that capture the full intertemporal covariance structure of the estimated dynamic response.

5.2 Testing for innovation and lobbying complementarities

Because post-shock innovation and lobbying are simultaneously determined endogenous outcomes of firm optimization, tests for heterogeneity in the firm's innovation response over its post-period lobbying (and *vice versa*) will be uninformative of complementarities between the two actions.[37] To address this simultaneity bias, I take averages of innovation and lobbying investments before the first shock a firm is exposed to, and use these as measures of preexisting innovation and lobbying capacities. Because these measures are constructed before a firm is treated, they are not simultaneously determined with post-period treatment responses. Innovation response heterogeneity over preexisting lobbying capacity serves to test for complementarities in firm innovation actions, while a parallel specification tests for the same complementarities in firm lobbying actions (Eqs. 5 and 6).

$$\Delta y_{it}^{R\&D} = \sum_{\ell \in b} \beta_\ell^{R\&D} LAG_{\ell it} + \sum_{\ell \in b} \gamma_\ell^{R\&D} LAG_{\ell it} \times \textit{High Pre-Lobby}_i \ + \ \phi_{st} + \epsilon_{it} \tag{5}$$

$$\Delta y_{it}^{Lobby} = \sum_{\ell \in b} \beta_\ell^{Lobby} LAG_{\ell it} + \sum_{\ell \in b} \gamma_\ell^{Lobby} LAG_{\ell it} \times \textit{High Pre-R\&D}_I \ + \ \phi_{st} + \epsilon_{it} \tag{6}$$

[36] Estimating cluster-robust standard errors by industry-year results in reduced uncertainty, so I allow for firm-level clusters to be conservative.

[37] Consider a situation where the rents at stake are large. Returns to capturing a market for clean substitutes may be large, while the regulatory profit wedge may also be large. Thus, the firm may have a large first-order payoff to both innovation and lobbying, regardless of whether the two are complements or substitutes. As such, even if the two are substitutes, the firm will both innovate more and lobby more relative to other firms (facing different shocks) operating in markets with lower rents at stake.

High Pre-Lobby$_i$ and *High Pre-R&D*$_I$ are indicators for pre-period lobbying and innovation being above the median value for treated firms. Compustat reports firm sales across industries, thus *High Pre-Lobby*$_i$ in Eq. 5 is based on the firm-specific, sales-share weighted average of pre-period industry lobbying. In Eq. 6, *High Pre-R&D*$_I$ is based on the average innovation expenditure across firms in industry I. The use of industry averages of pre-period measures in Eqs. 5 and 6 is motivated by the empirical fact that trade groups explicitly coordinate lobbying activities among members (Bombardini and Trebbi, 2012, 2020), leading me to account for spillovers by default.[38] In both specifications, the pre-period is the three periods which precede the *first* discovery an industry is exposed to, mitigating against information in discoveries entering pre-period actions. Note that the pre-period interaction terms represent a cross-sectional source of variation for estimating heterogeneity in firm responses. Finally, note that Equations 5 and 6 represent two separate tests of whether innovation and lobbying are complements or substitutes in firm actions. Cumulative effects increasing in the interaction value would be consistent with innovation and lobbying as complements (e.g., increased lobbying capacity associated a larger innovation response), while the opposite finding would be consistent with the two as substitutes. In theory the two tests should agree. However, because the specifications are separate estimates across different outcomes, in practice they need not agree on sign and may empirically support opposing conclusions.

6 Results

6.1 The innovation response

Table 4 displays estimated cumulative effects (Equations 3 and 4) of the discovery of a previously-unknown externality on firm R&D expenditures. The distributed lag model employed here estimates the overall effect of a shock as well as the dynamic evolution of firm responses, and is explicitly designed for a context of multiple treatments with temporally overlapping responses for the same panel unit — as distinct from a conventional event study estimator (Schmidheiny and Siegloch, 2023). Each row in Table 4 represents the cumulative average treatment effect of a scientific discovery of a previously unknown externality on the R&D investments of affected

[38] Appendix E further examines the potential role of spillovers.

firms ℓ years after the discovery. Time $\ell = 0$ is the year of publication of the discovery (by independent epidemiological researchers, not firm R&D scientists). Cumulative effects are reported relative to the first year before the discovery, $\ell = -1$.

Table 4 demonstrates the stability of the estimates across fixed effects and bandwidth specifications. Recall that all specifications are estimated in first differences, so that all estimates implicitly include a firm fixed effect (in levels) that is differenced out. Estimates in columns 1 – 4 of Table 4 employ the preferred bandwidth of three leads and six lags, while the final column examines robustness to a wider bandwidth of 10 leads and lags.[39] Bandwidth decisions are discussed further below, and robustness to alternative specifications is given in Appendix C. All estimates require balance of treated firms over their treatment period, starting from the longest lead before the first discovery affecting the firm until the last lag after the last discovery affecting the firm.[40] Finally, all estimates are reported for a shock that affects 100% of firm sales ($INTENSITY_{it}$ evaluated at unity in Eq. 3).[41]

Estimates in column 1 of Table 4 employ no fixed effects beyond the firm fixed effect implicit in the first differences estimator. Here, firms are estimated to increase innovation (normalized by pre-period market capitalization) by a highly significant 1.82 percentage points over the six years following discovery of a novel chemical harm. This estimated increase is quite similar in magnitude to that of the preferred specification (column 4). The identifying assumption in this column 1 estimate is that scientific discoveries of chemical harms arrive randomly from the perspective of firm R&D managers, conditional on firm average expectations over the arrival of future regulations and the share of firm sales they will affect. Lead effects in column 1 are economically small and statistically zero, consistent with discovery arrival as a shock to the firm's information set.

One might be concerned that this column 1 identifying assumption is too restrictive, however. Discovery rates by independent scientists could vary over time in ways that are correlated with business cycles or changes in governmental regulatory regimes, or scientists could over time start to actively target for study chemicals in certain sectors of the economy more than others. Columns 2 – 4 of Table 4 progressively relax the column 1 identifying assumption, conditioning

[39] Column 5 of Table 4 is truncated for exposition, but Figure 2b shows the full estimates.

[40] Firms that are treated but not balanced over all leads and lags are excluded from all estimates. The number of balanced treated firms is reported in Table 4 below each set of estimates.

[41] Treated firms' average treatment intensity in the Compustat data is 52%, and the 75^{th} percentile is 98%.

	Firm ΔR&D as % of market capitalization (cumulative effect estimates)				
Lag (years)	(1)	(2)	(3)	(4)	(5)†
10	—	—	—	—	1.81**
⋮					(0.81)
6	1.82***	1.81***	1.86***	1.77***	2.18***
	(0.54)	(0.54)	(0.54)	(0.55)	(0.71)
5	1.75***	1.72***	1.76***	1.63***	1.69***
	(0.47)	(0.47)	(0.47)	(0.47)	(0.64)
4	1.16***	1.07***	1.10***	1.03***	1.26**
	(0.37)	(0.37)	(0.37)	(0.38)	(0.53)
3	0.73**	0.66**	0.69**	0.83**	0.91**
	(0.33)	(0.32)	(0.33)	(0.34)	(0.44)
2	0.03	0.03	0.05	0.26	-0.02
	(0.33)	(0.33)	(0.33)	(0.35)	(0.26)
1	-0.26	-0.28	-0.27	-0.19	0.05
	(0.24)	(0.24)	(0.24)	(0.25)	(0.19)
0	-0.08	-0.11	-0.10	-0.11	-0.09
	(0.20)	(0.20)	(0.21)	(0.21)	(0.17)
-1	—	—	—	—	—
-2	-0.02	0.00	-0.02	0.03	0.08
	(0.21)	(0.21)	(0.21)	(0.22)	(0.26)
-3	-0.15	-0.15	-0.17	-0.05	-0.06
⋮	(0.28)	(0.28)	(0.28)	(0.28)	(0.20)
-10	—	—	—	—	-0.44
					(0.40)
Observations	221,368	221,368	221,368	221,368	213,990
Treated firms (balanced)	428	428	428	428	225
[leads, lags]	[3, 6]	[3, 6]	[3, 6]	[3, 6]	[10, 10]
Year FE		X	X		
Sector FE			X		
Sector×year FE				X	X

Table 4: **The cumulative effect of the discovery of a previously unknown externality on firm innovation.** Point estimates indicate the cumulative effect on firm innovation of a discovery of a chemical harm at lag 0 relative to firm innovation at lag -1 (lead 1). E.g., the lag 5 point estimate of column 1 indicates that treated firms cumulatively increased R&D spending, as a percentage of pre-treatment market capitalization, by 1.75 percentage points over the years 0–5 following discovery of a chemical harm. The distributed lag estimator explicitly allows for multiple treatment effects overlapping in time when recovering the impulse response estimate, as distinct from a conventional event study estimator (Schmidheiny and Siegloch, 2023). Standard errors are cluster-robust to arbitrary correlations in innovation investments within firm over time, and capture the full intertemporal covariance structure of the estimated dynamic effect. Lead effects serve as a partial test for differential trends.

†All estimates are over three leads and six lags, except column 5 with ten leads and lags. In column (5), leads -9 through -4 and lags 7 through 9 are not displayed for brevity, but are displayed in Figure 2b.

Estimates multiplied by 100 for interpretability.

$^{***}p < 0.01$,$^{**}p < 0.05$,$^{*}p < 0.1$

on year fixed effects to control for common temporal patterns in scientific discoveries (column 2), sector fixed effects to control for any trending tendency of discoveries to focus on one sector of the economy over another (column 3), and finally joint sector-year fixed effects to allow for sector-specific temporal shocks in discovery rates that may arbitrarily change over time (preferred specification, column 4). Cumulative estimates are remarkably stable across columns 1 – 4, ranging from 1.77 to 1.86 percentage point increases, and these effects are highly significant ($p < 0.01$) in all specifications. Further, lead effects are statistical zeros in each column. Figure C.1 displays the cumulative effects plots corresponding to each of the specifications in columns 1 – 4, which may aid visual comparisons.

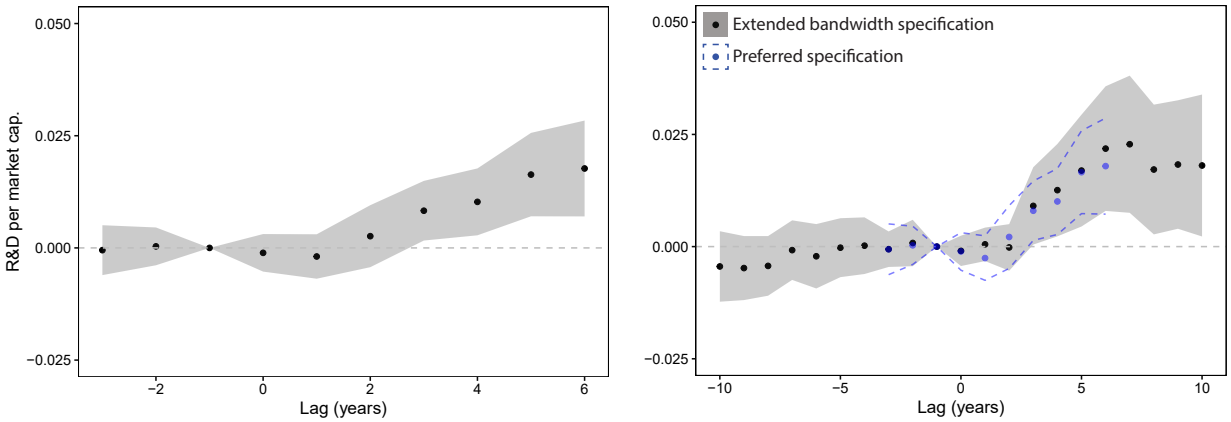


Figure 2: **The cumulative effect of the discovery of a previously unknown externality on firm innovation investments.** (a) Preferred specification (column 4 of Table 4), (b) extended bandwidth specification with 10 leads and lags (column 5 of Table 4). Time 0 is the year of publication of a discovery. The distributed lag estimator explicitly allows for multiple treatment effects overlapping in time when recovering the impulse response estimate, as distinct from a conventional event study estimator (Schmidheiny and Siegloch, 2023). 95% confidence intervals are robust to arbitrary correlations in innovation investments within firm over time, and capture the full intertemporal covariance structure of the estimated dynamic effect.

Focusing on the preferred specification (Figure 2a; column 4 of Table 4), I find that innovation investments increase by a highly significant 1.77% of a firm's pre-treatment market capitalization over the six years following a shock ($p < 0.01$). This six-year cumulative effect is economically meaningful, reflecting a 7.7% increase over the average normalized R&D flow for treated firms in my sample.[42] As discussed previously, increases in clean innovation out-

[42] A 1.77 percentage point R&D increase over six years implies an average annual flow increase in R&D investments of about 0.30% of firm pre-treatment market capitalization per year. The average normalized innovation flow is 3.83% per year for treated firms in my sample, so 0.30/3.83 = 7.7%.

puts (patents) by firms facing environmental regulation are well documented, but admit the possibility that total innovation output decreases through substitution out of other innovative activities. Indeed, evidence for changes in total patenting under environmental regulation is both more limited and mixed. Evidence regarding total innovation inputs — the key allocative decision faced by firm managers — is even more limited. The findings discussed here (Figure 2a; column 4 of Table 4) document a statistically significant and economically meaningful increase in firm total innovation inputs when managers are faced with an information shock over externalities linked to their operations. This total innovation increase is net of any within-firm substitution across innovative activities, and is representative of a broad set of industries and decades: it is estimated over treatments spanning 60 distinct industries and 44 distinct years.

The preferred specification estimates are causally identified if within-firm, within-sector-year changes in exposure to discoveries of novel chemical externalities by the epidemiological community are exogeneous shocks to firm information sets.[43] However, estimates will be biased if firms anticipate epidemiological discoveries, for example by observing trends in the epidemiological literature and the chemicals studied therein, or if epidemiologists target for study chemicals in industries with differentially changing premia to R&D investments (e.g. due to differential demand growth). In such cases, R&D investments will "respond" to a discovery before it occurs, biasing estimation of the post-shock response. A testable implication of my identifying assumption, therefore, is that firms do not react to exogenous shocks before they occur; thus lead effects provide a placebo test. Estimates in Figure 2b implement this test.

Figure 2b (column 5 of Table 4) displays a model with an extended set of 10 leads and lags. This model sacrifices power, requiring that treated firms be present in the data for 10 years before their first treatment and 10 years after their last treatment.[44] Remarkably, pre-treatment estimates are statistical and economic zeros over the entire 10 year period before a given shock. This suggests that (conditional on sector-year shocks) firms do not anticipate discoveries before they occur, nor do epidemiological researchers target chemicals or firms in industries with differential R&D growth. Note that this latter point is consistent with how epidemiologists report that they *actually do* target chemicals testing, which is based on factors

[43]Note the role of changes here: the first differenced estimator absorbs firm private time-invariant priors over the possibility of external harms from the chemicals they employ.

[44]This implies that fewer firms are estimated as "treated" in the extended model of Figure 2b relative to the preferred specification (225 versus 428 firms; Table 4).

unlikely to be correlated with changes in firm R&D investments such as potential seriousness of chemical health effects, evidence of public exposure, and costliness of testing (Section 3.1.1).

Additionally, this extended bandwidth model indicates that the preferred specification with six lags (overlaid in blue in Figure 2b for visual comparison) captures the full evolution of the dynamic response after treatment. The preferred specification cumulative estimate at lag six is nearly identical to the extended specification estimate at lag 10 (1.77% vs. 1.81%; columns 4 and 5 of Table 4).[45] This point of stabilization informs the bandwidth choice of six lags in the preferred specification, allowing it to capture the full dynamic innovation effect while conserving statistical power. Further, the temporal evolution of both specifications in Figure 2b before lag six is quite consistent, even though the preferred specification includes nearly double the number of treated firms. Finally, the cumulative effect remains significant over the extended 10-year post-treatment window ($p < 0.05$), implying the effects estimated here represent a permanent increase in total firm R&D.[46]

6.2 The lobbying response

Table 5 displays estimated cumulative effects of the discovery of a previously unknown externality on firm lobbying expenditures. These effects are estimated in a similar fashion to the innovation response previously discussed. Because the lobbying data are reported every half year, each row in Table 5 (and point in Figure 3) represents a six-month period, versus the annual periods shown in Figure 2 and Table 4. Time $\ell = 0$ is the half-year of publication of the novel externality discovery, and cumulative effects are reported relative to the first half-year before the discovery, $\ell = -1$. Estimates in columns 1 – 4 of Table 5 employ a preferred bandwidth of four leads and five lags, while the final column examines robustness to a wider bandwidth of 10 leads and lags. The number of balanced treated firms is reported in Table 5 below each set of estimates.

Estimates in column 1 of Table 5 employ no fixed effects (beyond the firm fixed effect implicit in the first differences estimator) and estimated treatment effects on firm lobbying are positive,

[45] Point estimates for lags six and seven in the extended bandwidth model are slightly larger than the lag six estimate in the preferred specification. However, differences are not statistically significant and appear to be transient noise in the extended bandwidth estimates, as the extended bandwidth cumulative effect stabilizes by lags 8 – 10 at a value nearly identical to the final estimate in the preferred specification (Figure 2b).

[46] That is, firms do not simply shift R&D expenditures forward in time, and then decrease them later on in an offsetting manner.

	Firm Δlobbying (in percentage terms) (cumulative effect estimates)				
Lag (half-years)	(1)	(2)	(3)	(4)	(5)†
10	—	—	—	—	23.01*
⋮					(12.10)
5	5.11	5.74*	7.73**	11.16**	15.91**
	(3.23)	(3.27)	(3.46)	(4.76)	(6.49)
4	2.36	4.16	7.52*	10.33*	7.88
	(4.36)	(4.35)	(4.52)	(5.40)	(5.42)
3	5.68	7.05*	11.01***	13.26***	16.60***
	(3.65)	(3.67)	(3.89)	(4.58)	(6.26)
2	2.56	4.03	6.99*	8.54**	12.42**
	(3.54)	(3.60)	(3.84)	(4.08)	(5.91)
1	6.87**	7.06**	7.97**	8.95***	11.33*
	(2.96)	(2.99)	(3.18)	(3.37)	(5.94)
0	3.33	3.94*	4.33*	4.81*	0.84
	(2.30)	(2.32)	(2.46)	(2.49)	(2.93)
-1	—	—	—	—	—
-2	-3.34	-3.42	-1.90	-2.61	-3.55
	(2.12)	(2.13)	(2.14)	(2.26)	(3.23)
-3	-2.70	-3.14	-0.19	-1.66	4.82
	(4.00)	(4.01)	(4.16)	(4.53)	(4.49)
-4	1.52	0.01	2.50	0.30	-3.13
⋮	(3.11)	(3.13)	(3.31)	(4.01)	(7.40)
-10	—	—	—	—	5.87
					(6.96)
Observations	287,670	287,670	287,670	287,670	143,689
Treated firms (balanced)	407	407	407	407	243
[leads, lags]	[4, 5]	[4, 5]	[4, 5]	[4, 5]	[10, 10]
Half-year FE		X			
Sector FE		X			
Sector×half-year FE			X	X	X
Firm FE				X	X

Table 5: **The cumulative effect of the discovery of a previously unknown externality on firm lobbying.** Point estimates indicate the cumulative effect on firm lobbying of a discovery of a chemical harm at lag 0 relative to firm lobbying at lag -1 (lead 1). E.g., the lag 1 point estimate of column 1 indicates that treated firms cumulatively increased lobbying spending, as a percentage of pre-treatment baseline lobbying, by 6.87 percentage points over the half-years 0–1 following the discovery of a chemical harm. Standard errors are cluster-robust to arbitrary correlations in lobbying investments within firm over time, and capture the full intertemporal covariance structure of the estimated dynamic effect.
†All estimates are over four leads and three lags, except column 5 with 10 leads and lags. In column 5, leads -9 through -5 and lags 6 through 9 are not displayed for brevity, but are displayed in Figure 3b.
Estimates multiplied by 100 for interpretability.
$^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$

but insignificant over most lags. As in the previous section, the column 1 identifying assumption is that discoveries by epidemiological researchers arrive randomly from the firm's perspective, conditional on firm average expectations over the arrival of future regulations affecting their output. Assuming equally random arrival of discoveries over time and across sectors of the economy may be overly restrictive, however. Columns 2 – 4 of Table 5 progressively relax this identifying assumption, conditioning on half-year fixed effects to control for common temporal patterns in scientific discoveries along with sector fixed effects to control for any tendency of discoveries to focus on one sector of the economy over another (column 2), then joint sector-half-year fixed effects to allow for sector-specific temporal shocks in discovery rates (column 3), and finally adding the firm fixed effect (column 4), which is discussed below. Cumulative estimates increase in magnitude and significance from column 1 (5.11% increase, insignificant, no fixed effects) to column 4 (11.16% increase, $p < 0.05$, firm and sector-by-half-year fixed effects), as added controls refine the comparison group in estimation. Estimated lead effects are insignificant in all specifications. Figure D.1 displays the cumulative effects plots corresponding to each of the specifications in columns 1 – 5, which may aid visual comparisons.

Focusing on the preferred specification (Figure 3a; column 4 of Table 4), I find that firms increase their lobbying expenditures over about a two year period in response to the scientific discovery of a previously unknown externality ($p < 0.05$). This increase in lobbying investments is an economically meaningful 11.16% of pre-period lobbying levels and it implies an average flow increase of 2.2% per period over baseline levels. Interestingly, firm increases in lobbying expenditures occur much more rapidly (about two years) than does the increase in R&D expenditures (about six years). As with the innovation estimates, the lobbying outcome captures aggregate expenditures across the firm, and therefore these estimates represent increases in total firm lobbying expenditures and not simply reallocation within the firm. To my knowledge, these estimates are the first evidence of firm responses to the type of knowledge shock (epidemiological findings) that is used by regulators to initiate regulatory proceedings.

As in the R&D results, I estimate an extended bandwidth model with 10 leads and lags, displayed in Figure 3b (column 5 of Table 5). Several key points from this model are worth discussing. First, as before, this extended bandwidth model sacrifices power, and estimates exhibit more noise than the preferred specification. Second, all specifications are still estimated in first-differences and so a firm fixed effect (in levels) is implicit in each set of estimates.

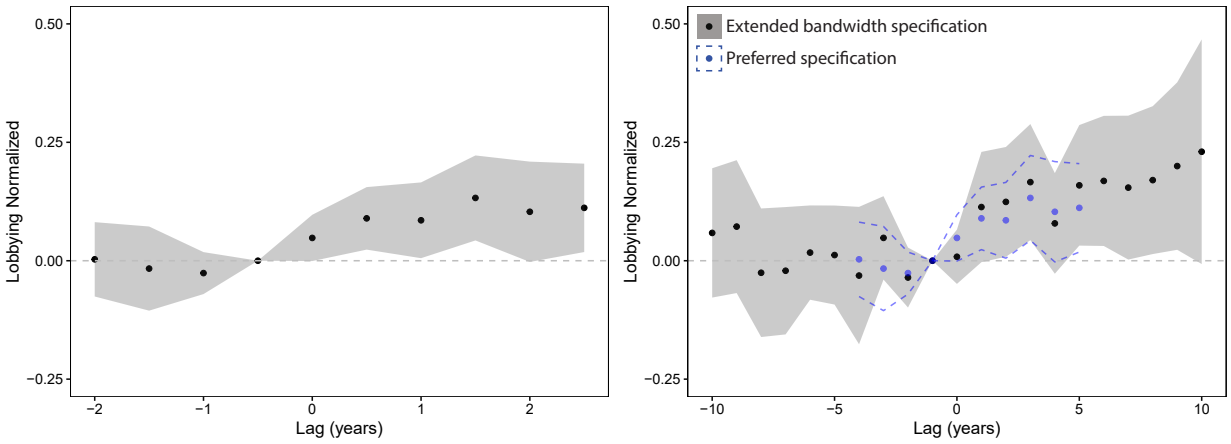


Figure 3: **The cumulative effect of the discovery of a previously unknown externality on firm lobbying investments.** (a) Preferred specification (column 4 of Table 5), (b) extended bandwidth specification with 10 leads and lags (column 5 of Table 5). Time 0 is the half-year of publication of a discovery, and the estimator includes five subsequent periods (half-years) of lagged effects. The distributed lag estimator explicitly allows for multiple treatment effects overlapping in time when recovering the impulse response estimate, as distinct from a conventional event study estimator (Schmidheiny and Siegloch, 2023). 95% confidence intervals are robust to arbitrary correlations in innovation investments within firm over time, and capture the full intertemporal covariance structure of the estimated dynamic effect.

However, distinct from the R&D estimates, the preferred specification for firm lobbying also includes a firm fixed effect in the first-differenced estimator (equivalent to a firm average time trend in levels). This firm fixed effect stabilizes some pre-trending behavior in the extended bandwidth lobbying estimates over 10 leads and lags, which motivates its inclusion in the preferred specification. When including this firm fixed effect, estimated lobbying lead effects are flat even up to 10 periods before a shock (Figure 3b).[47] As before, this suggests that (conditional on sector-year shocks and firm average time trends) firms do not anticipate discoveries before they occur, nor do epidemiological researchers target chemicals or firms in industries with differential lobbying growth. Again, this latter point is consistent with how epidemiologists report that they actually do target chemicals testing, which is based on factors unlikely to be correlated with changes in firm investments (Section 3.1.1). Third, the 10-period cumulative effect estimates stabilize around a 15% increase about 2.5 years after a discovery (lag 5). This increase in lobbying is slightly higher than the 11.16% increase in my preferred specification at the same lag, though not statistically different from it. This point of stabilization of the

[47]While lobbying lead estimates show some noise over the ninth and tenth half-years preceding a discovery, they are all statistical zeros and are reasonably flat over leads 1 – 8 (Figure 3b). A firm fixed effect is not necessary in the R&D estimates because pre-trends, even over 10 leads, are already remarkably flat (Figure 2b).

cumulative effect five periods after a discovery informs the bandwidth choice of five lags in the preferred specification, allowing the preferred specification to capture the full dynamic lobbying effect while conserving statistical power.[48] Fourth, the cumulative lobbying effect remains largely significant over this extended window ($p < 0.05$ for most lags; $p < 0.1$ in the final lag) implying the estimated effects are permanent increases in firm lobbying.

Finally, it is interesting to note that while the R&D estimates are remarkably robust to fixed effects specification, with consistent magnitudes and precision across specifications (Table 4), the lobbying results exhibit a more typical level of sensitivity to such specification choices (Table 5). Importantly, R&D reporting dates back to the 1970s, while lobbying reporting only started in 1998. So, while both sets of estimates have similar numbers of treated firms (a little over 400), lobbying estimates only employ treatments from the post-2000 period while the R&D estimates span five decades (Table 2). This suggests there may be a greater potential for realizations of idiosyncratic shocks in the shorter-run lobbying panel to correlate with the timing of post-2000 treatments. In such a case, increasing the stringency of fixed effects in estimation absorbs more sources of potentially confounding variation, thus refining the control group and relaxing the identifying assumption in the model. This indeed appears to be the case in the lobbying estimates here.

6.3 Additional robustness tests

The R&D and lobbying estimates in Sections 6.1 and 6.2 are robust to multiple variants of the main specification. Tables 4 and 5 display a range of fixed effects specifications, and the corresponding plots for each specification are given in Appendix Figures C.1 (innovation response) and D.1 (lobbying response). Robustness tests to alternative bandwidths over which leads and lags are estimated are given in Appendix Figures C.2 (R&D response) and D.2 (lobbying response). Robustness tests to dropping individual discoveries from the estimating data set (indicating that estimated average effects are not driven by any one single discovery) are given in Appendix Figures C.3 (R&D response) and D.3 (lobbying response). Finally, robustness

[48] As in the previous section, Column 5 of Table 5 omits several lead and lag estimates for brevity, but Figure 3b displays all cumulative effect estimates.

tests to normalizing control firm outcomes to an early panel year instead of the panel average[49] are given in Appendix Figures C.4 (R&D response) and D.4 (lobbying response).

6.4 Evidence of firm substitution between innovation and lobbying

Having estimated the magnitude of firm first-order R&D and lobbying responses to the discovery of a novel externality, I now turn to the question of whether R&D and lobbying are complements or substitutes in firm actions.

Two independent tests both provide suggestive evidence that R&D and lobbying are on average substitutes in the actions of treated firms, though the second is not statistically significant. The first test suggests that the option to lobby induces less R&D from firms than they otherwise might choose (Figure 4; Table 6, left panel), and the second suggests (with much more noise) that the option to innovate induces less lobbying from firms than they might otherwise choose (Figure 5; Table 6, right panel). Additional tests suggest that lobbying spillovers are an important component of firm R&D responses (Appendix Figure E.1). However, the test for innovation spillovers in firm lobbying is inconclusive (Appendix Figure E.2).

6.4.1 Innovation response heterogeneity over preexisting lobbying capacity

Testing for R&D response heterogeneity over preexisting lobbying capacity ("high lobby" or "low lobby" firms), I find that low lobby firms innovate substantially more in response to the discovery of a novel externality than do high lobby firms (Figure 4; Table 6, left panel). High lobby firms do eventually have a marginally significant cumulative increase of 0.71 percentage points in normalized R&D ($p < 0.1$) six years after a scientific discovery. However, this high lobby response is a statistical and economic zero for four years after a discovery. In contrast, low lobby firms have an R&D response that is significant in all post-discovery periods (lag 1 is marginally significant, all other lags significant at $p < 0.05$ or $p < 0.01$) and reaches a 2.29 percentage point increase in normalized innovation in the six years following a discovery —

[49]Never-treated control firms do not have any well defined "baseline" period (Section 3.3). Thus, in the main specification, control firm R&D is normalized to the firm's panel average market capitalization (and for lobbying control firms, panel average lobbying). However, one might be concerned that this normalization implies information about treatments enters the denominator for some control firms. These appendix estimates demonstrate robustness to normalizing control firms over a very early panel year. Note that treated firms are always normalized to the period before the *first* treatment they are exposed to, avoiding such concerns.

triple the high-lobby firm response.[50] The interaction effect estimated here is highly flexible: each lead and lag indicator in the estimator has a separate interaction with pre-period lobbying (Eq. 5), and nothing constrains the cumulative effects to be consistently above or below each other in every period. Yet, for every post-period in the estimate, low lobby firms exhibit a larger R&D response while high lobby firms exhibit a diminished response. These findings suggest preexisting lobbying capacity may be a substantial driver of firm R&D responses. Further, they are consistent with innovation and lobbying as substitutes in firm actions, suggesting that the option to lobby may induce firms to innovate less than they otherwise would have. Variation in preexisting lobbying capacity is cross-sectional, thus this heterogeneity should not be interpreted causally. However, to my knowledge, these estimates represent the first empirical (suggestive) evidence for a firm tradeoff between innovation and lobbying.

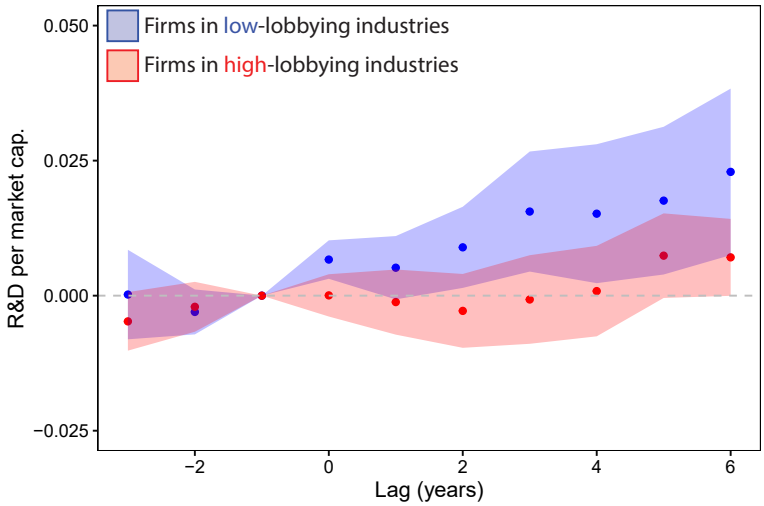


Figure 4: **Heterogeneity over preexisting lobbying capacity in the cumulative R&D response.** High lobbying capacity firms tend to innovate less in the post-period, an important source of heterogeneity in the main R&D response (Figure 2). This is consistent with innovation and lobbying as substitutes in firm actions. Note that preexisting lobbying capacity is a cross-sectional source of variation, and thus the heterogeneity in effects shown here should be considered as suggestive in nature. 95% confidence intervals are robust to arbitrary correlations in innovation investments within firm over time, and capture the full intertemporal covariance structure of the estimated dynamic effect.

6.4.2 Lobbying response heterogeneity over preexisting innovation capacity

Estimates of heterogeneity in the lobbying response exhibit much more noise than those for the R&D response, limiting the ability to draw strong conclusions. As with the R&D heterogeneity

[50] One of the lead estimates for high-lobby firms is also marginally different from zero. But, note that this regression includes 20 cumulative effect estimates.

estimates, the interaction effect estimated here is highly flexible: each lead and lag indicator has a separate interaction with preexisting innovation capacity (Eq. 6), and nothing constrains the cumulative effects to be consistently above or below each other in every period. Point estimates suggest low R&D firms tend to lobby more in response to the threat of future regulation (Figure 4; Table 6, right panel), which would be consistent with innovation and lobbying as substitutes in firm actions. Note, though, that differences in magnitude over most post-periods are not large, except the last period. In that period, the estimated lobbying magnitude for low R&D firms is double that of high R&D firms (20.48% significant increase versus 9.46% insignificant increase). However, estimated magnitudes for the two types of firms are generally more similar for other lags (excepting lag 3). After lag 1, low R&D estimates are generally significant while high R&D estimates are rarely different from zero; however, instability in the estimates and wide standard errors suggests on balance that limited conclusions can be drawn from these estimates.

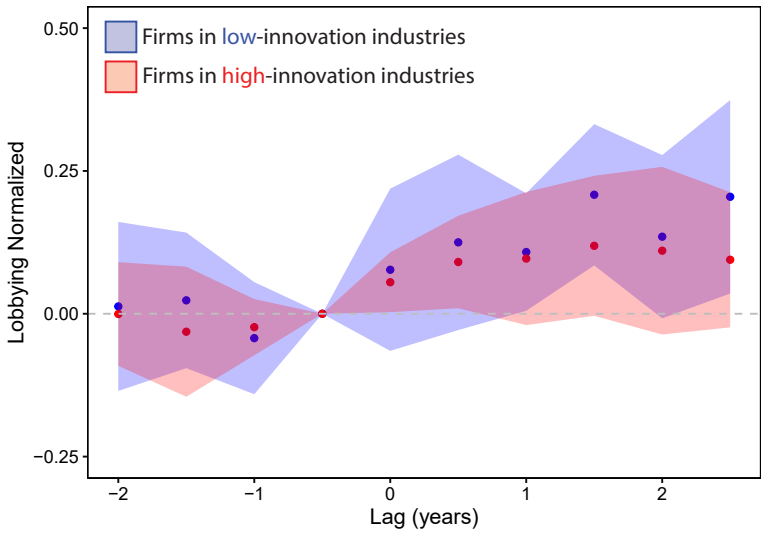


Figure 5: **Heterogeneity over preexisting innovation capacity in the cumulative lobbing response.** High R&D capacity firms tend to lobby less in the post-period, though estimates are too noisy to draw clear conclusions. Preexisting innovation capacity is a cross-sectional source of variation, and thus the heterogeneity in effects shown here should be considered as suggestive in nature. 95% confidence intervals are robust to arbitrary correlations in innovation investments within firm over time, and capture the full intertemporal covariance structure of the estimated dynamic effect.

Firm ΔR&D as % of market capitalization (cumulative effect estimates)			Firm Δlobbying in percentage terms (cumulative effect estimates)		
	(1)			(2)	
Lag (years)	R&D response for: High lobby	Low lobby	Lag (half-years)	Lobby response for: High R&D	Low R&D
6	0.71*	2.29***	5	9.46	20.48**
	(0.36)	(0.79)		(6.04)	(8.64)
5	0.74*	1.76**	4	11.04	13.49*
	(0.40)	(0.70)		(7.49)	(7.29)
4	0.08	1.52**	3	11.89*	20.83***
	(0.43)	(0.66)		(6.25)	(6.31)
3	-0.07	1.56***	2	9.67	10.81**
	(0.42)	(0.57)		(5.94)	(5.25)
2	-0.28	0.89**	1	9.06**	12.50
	(0.35)	(0.38)		(4.13)	(7.83)
1	-0.12	0.52*	0	5.52**	7.70
	(0.31)	(0.30)		(2.67)	(7.24)
0	0.01	0.67***	-1	—	—
	(0.20)	(0.18)			
-1	—	—	-2	-2.34	-4.27
				(2.51)	(5.01)
-2	-0.21	-0.30	-3	-3.14	2.35
	(0.24)	(0.21)		(5.81)	(6.05)
-3	-0.48*	0.02	-4	-0.04	1.28
	(0.28)	(0.42)		(4.63)	(7.55)
Observations	253,664			287,670	
Treated firms (balanced)	371			407	
[leads, lags]	[3, 6]			[4, 3]	
Fixed effects	Sector-year			Sector-year, firm	

Table 6: **Heterogeneity in the innovation response by lobbying capacity, and in the lobbying response by innovation capacity.** Columns of the left panel (1) correspond to Figure 4; columns of the right panel (2) correspond to Figure 5. Left panel (1): Cumulative innovation response for firms in industries with high preexisting lobbying capacity ("High lobby" column), and low preexisting lobbying capacity ("Low lobby" column). E.g., six years after a discovery, "high lobby" firms increased innovation by a marginally significant 0.71 percent of market capitalization. In contrast, "low lobby" firms increased innovation by a highly significant 2.29 percentage points. Right panel (2): Cumulative lobbying response for firms in industries with high preexisting innovation capacity ("High R&D" column), and low preexisting innovation capacity ("Low R&D" column). E.g., five half-years after a discovery, "high R&D" firms increased lobbying by an insignificant 9.46 percentage points. In contrast, "low R&D" firms increased lobbying by a significant 20.48 percentage points. Estimates in each panel are from a single regression, and standard errors are robust to arbitrary correlations in firm outcomes within firm over time, and capture the full intertemporal covariance structure of the estimated dynamic effect. $^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$

7 Welfare implications of the option to lobby

Considering the welfare implications of the firm's option to lobby requires caution for two reasons. First, lobbying over regulation may itself have positive or negative welfare implications (Bombardini and Trebbi, 2020). If the regulatory instrument considered requires private firm information to determine optimal regulation (e.g. a performance standard or command and control regulation), then it is entirely possible that lobbying by firms communicates information necessary for optimal regulatory design. In such a case, lobbying would have positive welfare implications. On the other hand, if the regulatory instrument considered does not rely on private information (e.g. a Pigouvian tax[51]), then lobbying by firms may be nothing more than an effort to protect or appropriate rents. However, even in the case of Pigouvian taxation — where firms might not have private information over the optimal tax — the welfare implications of lobbying in a second-best policy context may not be obvious. For example, the regulator may be concerned about leakage when regulating a trade-exposed sector, in which case firm private information regarding supply in the regulated and unregulated counterfactuals, as well as demand for output, may be necessary to estimate leakage effects (Baylis, Fullerton, and Karney, 2014). To be conservative, I do not take a stand here on the direct social welfare value of firm lobbying, which may be positive or negative.

Second, the heterogeneity in firm R&D responses that I estimate is over cross-sectional variation in preexisting lobbying capacity (Equation 5). It may be the case that firms with stronger R&D responses simply happen to have sales in industry sectors where lobbying is less prevalent, and vice versa (Figure 4). The same logic applies to my estimates of heterogeneity in the lobbying response (Figure 5). While my findings are consistent with innovation and lobbying as substitutes in firm actions, they are not conclusive evidence.

With these two caveats in mind, I conduct a back-of-the-envelope evaluation of the welfare implications of my findings. I estimate an average reduction in firms' six-year cumulative R&D responses from 2.3% to 0.7% of firm market capitalization for firms in industries with below median versus above median preexisting lobbying capacity (Table 6, left panel). Taking these estimates at face value, going from below median to above median lobbying capacity is associated with a 1.5 percentage point reduction in normalized R&D. The total pre-period

[51]The Pigouvian tax to correct an externality is set at the level of marginal damages, which is not something that firms generally have private information about.

market capitalization of treated firms in sample is $2.4 trillion, which implies a reduction in R&D investments of $36 billion. A range of estimates exist to value this reduction in innovation Hall (1999) estimates private benefits to innovation at a factor of 5x the cost. Bloom, Schankerman, and Van Reenen (2013) estimate marginal social returns at 55%, corresponding to present value marginal benefits at roughly 9x the cost (Jones and Summers, 2020). Jones and Summers (2020) parameterize a wide range of possible innovation returns and have difficulty finding estimates for average or marginal benefits below 4x the cost, with many estimates exceeding 10x. Taking the most conservative estimate available ($4 of marginal social benefit per $1 spent) implies the social value of foregone R&D estimated here is about $144 billion ($36 billion $\times$ 4).

8 Conclusions

This paper assembles novel data on shocks to scientific knowledge regarding original discoveries of pollution externalities associated with commonly used industrial chemicals, then empirically examines the innovation and lobbying responses of affected firms. I focus on new discoveries of low dose, chronic exposure chemical harms, as these harms are difficult to observe and U.S. regulators were not permitted to require testing for them. Reviewing over 7,000 publications from the epidemiological literature, I compile 104 discoveries of previously unknown pollution externalities that affect firms spanning 60 different industries and 44 distinct years.

I find that firms exposed to the discovery of a novel pollution externality increase total R&D expenditures by an economically important margin, an increase that is net of any within-firm substitution in innovation activities. I additionally find that shocked firms increase their lobbying expenditures, a finding which indicates that the knowledge shocks I study do indeed shift firm regulatory expectations. More importantly, this finding highlights that the same knowledge shocks that drive firm R&D also drive firms to lobby the regulator.

This latter statement motivates the observation, formalized in a simple conceptual framework, that a firm's option to lobby the regulator distorts its private innovation incentives, but the sign of the distortion is ambiguous. Empirically, I find that preexisting lobbying capacity is linked to substantial heterogeneity in firm R&D treatment effects. Firms with low preexisting lobbying capacity make additional R&D investments at *triple* the rate of their high lobbying

capacity counterparts. A symmetric test for heterogeneity in firm lobbying effects yields consistent findings, though with much more noise. Additional tests suggest that lobbying spillovers may be an important driver of firm R&D responses. Importantly, while all suggestive, each of these tests is consistent with innovation and lobbying as substitutes in firm actions, suggesting that the option to lobby induces shocked firms to innovate less than they otherwise would have.

These findings have multiple broader implications. First, innovation is an important driver of welfare, so whether or not regulation drives changes in total innovation has important implications for the economic value of regulatory action. Findings of regulatory-induced clean patenting generally do not account for potential substitution out of dirty innovation, and existing evidence on firm total innovation net of such substitution is both limited and mixed. The R&D increase I document provides new and comprehensive evidence on total innovation efforts of firms facing environmental regulation, evidence spanning over one hundred externalities, many dozens of industries, and four decades. Second, while lobbying itself may or may not be beneficial for optimal regulatory design, evidence here suggests it does distort firm innovation incentives with large welfare implications. Third, this distortionary effect of firm lobbying can affect the direction of technical change in a dynamic economy, where welfare implications may be even more substantial than those estimated here. This represents a promising avenue for future work.

References

Acemoglu, Daron, Philippe Aghion, Leonardo Bursztyn, and David Hemous. 2012. "The environment and directed technical change." *American Economic Review* 102 (1):131–66.

Aghion, Philippe, Roland Bénabou, Ralf Martin, and Alexandra Roulet. 2023. "Environmental preferences and technological choices: Is market competition clean or dirty?" *American Economic Review: Insights* 5 (1):1–19.

Aghion, Philippe, Antonin Bergeaud, and John Van Reenen. 2023. "The impact of regulation on innovation." *American Economic Review* 113 (11):2894–2936.

Aghion, Philippe, Nick Bloom, Richard Blundell, Rachel Griffith, and Peter Howitt. 2005. "Competition and innovation: An inverted-U relationship." *The Quarterly Journal of Economics* 120 (2):701–728.

Aghion, Philippe, Antoine Dechezleprêtre, David Hemous, Ralf Martin, and John Van Reenen. 2016. "Carbon taxes, path dependency, and directed technical change: Evidence from the auto industry." *Journal of Political Economy* 124 (1):1–51.

Akcigit, Ufuk, Salomé Baslandze, and Francesca Lotti. 2023. "Connecting to power: political connections, innovation, and firm dynamics." *Econometrica* 91 (2):529–564.

Ambec, Stefan, Mark A Cohen, Stewart Elgie, and Paul Lanoie. 2013. "The Porter hypothesis at 20: can environmental regulation enhance innovation and competitiveness?" *Review of environmental economics and policy* .

Ansolabehere, Stephen, John M De Figueiredo, and James M Snyder Jr. 2003. "Why is there so little money in US politics?" *Journal of Economic Perspectives* 17 (1):105–130.

Apergis, Nicholas, Claire Economidou, and Ioannis Filippidis. 2008. "Innovation, technology transfer and labor productivity linkages: evidence from a panel of manufacturing industries." *Review of World Economics* 144 (3):491–508.

Arora, Ashish, Sharon Belenzon, and Lia Sheer. 2021. "Knowledge spillovers and corporate investment in scientific research." *American Economic Review* 111 (3):871–898.

Arrow, Kenneth. 1962. "Economic Welfare and the Allocation of Resources for Invention (pp. 609-626)." *The Rate and Direction of Inventive Activity: Economic and Social Factors (ed. R. Nelson)* .

Baylis, Kathy, Don Fullerton, and Daniel H Karney. 2014. "Negative leakage." *Journal of the Association of Environmental and Resource Economists* 1 (1/2):51–73.

Benedick, Richard Elliot. 1991. *Ozone diplomacy: new directions in safeguarding the planet*. Harvard University Press.

Beraja, Martin, David Y Yang, and Noam Yuchtman. 2023. "Data-intensive innovation and the state: Evidence from AI firms in China." *The Review of Economic Studies* 90 (4):1701–1723.

Bertrand, Marianne, Matilde Bombardini, Raymond Fisman, and Francesco Trebbi. 2020. "Tax-exempt lobbying: Corporate philanthropy as a tool for political influence." *American Economic Review* 110 (7):2065–2102.

Bertrand, Marianne, Matilde Bombardini, and Francesco Trebbi. 2014. "Is it whom you know or what you know? An empirical assessment of the lobbying process." *American Economic Review* 104 (12):3885–3920.

Bloom, Nicholas, Mark Schankerman, and John Van Reenen. 2013. "Identifying technology spillovers and product market rivalry." *Econometrica* 81 (4):1347–1393.

Bloom, Nick, Rachel Griffith, and John Van Reenen. 2002. "Do R&D tax credits work? Evidence from a panel of countries 1979–1997." *Journal of Public Economics* 85 (1):1–31.

Blundell, Richard, Rachel Griffith, and John Van Reenen. 1999. "Market share, market value and innovation in a panel of British manufacturing firms." *The Review of Economic Studies* 66 (3):529–554.

Bombardini, Matilde and Francesco Trebbi. 2012. "Competition and political organization: Together or alone in lobbying for trade policy?" *Journal of International Economics* 87 (1):18–26.

———. 2020. "Empirical models of lobbying." *Annual Review of Economics* 12:391–413.

Brown, James R, Gustav Martinsson, and Christian Thomann. 2022. "Can environmental policy encourage technical change? Emissions taxes and R&D investment in polluting firms." *The Review of financial studies* 35 (10):4518–4560.

Brulle, Robert J. 2018. "The climate lobby: a sectoral analysis of lobbying spending on climate change in the USA, 2000 to 2016." *Climatic Change* 149 (3):289–303.

Brunel, Claire. 2019. "Green innovation and green Imports: Links between environmental policies, innovation, and production." *Journal of Environmental Management* 248:109290.

Calel, Raphael. 2020. "Adopt or innovate: Understanding technological responses to cap-and-trade." *American Economic Journal: Economic Policy* 12 (3):170–201.

Calel, Raphael and Antoine Dechezleprêtre. 2016. "Environmental policy and directed technological change: evidence from the European carbon market." *Review of Economics and Statistics* 98 (1):173–191.

CDC. 2022. "National Report on Human Exposure to Environmental Chemicals." *National Center for Environmental Health, U.S. Department of Health and Human Services, Centers for Disease Control and Prevention* URL https://www.cdc.gov/environmental-exposure-report/about-the-chemicals/index.html. Updated December 2024. [Accessed July 6, 2026].

Chan, Louis KC, Josef Lakonishok, and Theodore Sougiannis. 2001. "The stock market valuation of research and development expenditures." *The Journal of Finance* 56 (6):2431–2456.

Cohen, Wesley M. 2010. "Fifty years of empirical studies of innovative activity and performance." *Handbook of the Economics of Innovation* 1:129–213.

Dechezleprêtre, Antoine and Misato Sato. 2017. "The impacts of environmental regulations on competitiveness." *Review of environmental economics and policy* .

Denison, Richard. 2017. "A primer on the new Toxic Substances Control Act (TSCA) and what led to it." *Whitepaper: Environmental Defense Fund* .

Dugoua, Eugenie. 2023. "Induced innovation and international environmental agreements: evidence from the ozone regime." *Review of Economics and Statistics* :1–45.

Dugoua, Eugenie and Todd D Gerarden. 2025. "Induced innovation, inventors, and the energy transition." *American Economic Review: Insights* 7 (1):90–106.

EPA, US. 2021. "TRI-Listed Chemicals." https://www.epa.gov/toxics-release-inventory-tri-program/tri-listed-chemicals. [Accessed October 29, 2021].

Gehlbach, Scott, Konstantin Sonin, and Ekaterina Zhuravskaya. 2010. "Businessman candidates." *American Journal of Political Science* 54 (3):718–736.

Gerlach, C. 2016. "New Toxic Substances Control Act: An End to the Wild West for Chemical Safety." https://sitn.hms.harvard.edu/flash/2016/new-toxic-substances-control-act-end-wild-west-chemical-safety/. [Accessed October 26, 2021].

Gowrisankaran, Gautam, Ashley Langer, and Wendan Zhang. 2022. "Policy uncertainty in the market for coal electricity: The case of air toxics standards." *Working Paper. National Bureau of Economic Research* .

Greene, William H. 2003. *Econometric Analysis*. Pearson Education India.

Hall, Bronwyn H. 1999. "Innovation and market value."

Hall, Bronwyn H and Dietmar Harhoff. 2012. "Recent research on the economics of patents." *Annu. Rev. Econ.* 4 (1):541–565.

Harvey, Philip W. 2005. "Human relevance of rodent prolactin-induced non-genotoxic mammary carcinogenesis: prolactin involvement in human breast cancer and significance for toxicology risk assessments." *Journal of Applied Toxicology: An International Journal* 25 (3):179–183.

Higham, Kyle, Gaétan De Rassenfosse, and Adam B Jaffe. 2021. "Patent quality: Towards a systematic framework for analysis and measurement." *Research Policy* 50 (4):104215.

i Vidal, Jordi Blanes, Mirko Draca, and Christian Fons-Rosen. 2012. "Revolving door lobbyists." *The American Economic Review* 102 (7):3731.

Jaffe, Adam B and Karen Palmer. 1997. "Environmental regulation and innovation: a panel data study." *Review of Economics and Statistics* 79 (4):610–619.

Jones, Benjamin F and Lawrence H Summers. 2020. *A calculation of the social returns to innovation*, vol. 27863. National Bureau of Economic Research.

Kang, Karam. 2016. "Policy influence and private returns from lobbying in the energy sector." *The Review of Economic Studies* 83 (1):269–305.

Kim, Sung Eun, Johannes Urpelainen, and Joonseok Yang. 2016. "Electric utilities and American climate policy: lobbying by expected winners and losers." *Journal of Public Policy* 36 (2):251–275.

Lanjouw, Jean Olson and Ashoka Mody. 1996. "Innovation and the international diffusion of environmentally responsive technology." *Research Policy* 25 (4):549–571.

Lu, Yangsiyu and Jacquelyn Pless. 2026. "Greening to grow: Evidence from environmental regulation and industrial firm productivity in China." *Journal of Public Economics* 256:105596.

Meng, Kyle C and Ashwin Rode. 2019. "The social cost of lobbying over climate policy." *Nature Climate Change* 9 (6):472–476.

Okunade, Albert A and Vasudeva NR Murthy. 2002. "Technology as a "major driver" of health care costs: a cointegration analysis of the Newhouse conjecture." *Journal of Health Economics* 21 (1):147–159.

Popp, David. 2006. "International innovation and diffusion of air pollution control technologies: the effects of NOX and SO2 regulation in the US, Japan, and Germany." *Journal of Environmental Economics and Management* 51 (1):46–71.

———. 2019. "Environmental policy and innovation: a decade of research." *Working Paper. National Bureau of Economic Research* .

Popp, David, Richard G Newell, and Adam B Jaffe. 2010. "Energy, the environment, and technological change." *Handbook of the Economics of Innovation* 2:873–937.

Porter, Michael. 1996. "America's green strategy." *Business and the Environment: A Reader* 33:1072.

Rittenhouse, Katherine and Matthew Zaragoza-Watkins. 2018. "Anticipation and environmental regulation." *Journal of Environmental Economics and Management* 89:255–277.

Romer, Paul M. 1990. "Endogenous technological change." *Journal of Political Economy* 98 (5, Part 2):S71–S102.

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

Schmidheiny, Kurt and Sebastian Siegloch. 2023. "On event studies and distributed-lags in two-way fixed effects models: Identification, equivalence, and generalization." *Journal of Applied Econometrics* 38 (5):695–713.

Schmidt, Charles W. 2016. "TSCA 2.0: A new era in chemical risk management." *Environmental Health Perspectives* 124 (10).

Schumpeter, Joseph. 1942. "Creative destruction. Capitalism." *Socialism and Democracy* 825 (82-85).

Slinko, Irina, Evgeny Yakovlev, and Ekaterina Zhuravskaya. 2005. "Laws for sale: evidence from Russia." *American law and economics review* 7 (1):284–318.

Woodruff, Tracey J, Amy D Kyle, and Frédéric Y Bois. 1994. "Evaluating health risks from occupational exposure to pesticides and the regulatory response." *Environmental Health Perspectives* 102 (12):1088–1096.

Online Appendix

A Conceptual framework details

This appendix presents details of the conceptual framework summarized in Section 2. First order conditions are as given in Section 2:

$$\eta \pi_c^{*\prime} = 1 \tag{A.1}$$

$$(1-\eta)\pi_d^{*\prime} = 1 \tag{A.2}$$

$$\Delta\pi \partial\eta/\partial\lambda = 1 \tag{A.3}$$

where $\pi_c^{*\prime} = \partial\pi_c^*/\partial\rho_c$, $\pi_d^{*\prime} = \partial\pi_d^*/\partial\rho_d$, and $\partial\eta/\partial\lambda = \partial\eta/\partial\lambda_f$ if $\Delta\pi > 0$ and $\partial\eta/\partial\lambda_a$ otherwise.

The Jacobian of the FOCs and its determinant are:

$$J = \begin{bmatrix} \eta''\Delta\pi & \eta'\pi_c^{*\prime} & -\eta'\pi_d^{*\prime} \\ \eta'\pi_c^{*\prime} & \eta\pi_c^{*\prime\prime} & 0 \\ -\eta'\pi_d^{*\prime} & 0 & (1-\eta)\pi_d^{*\prime\prime} \end{bmatrix}$$

$$det(J) = \eta''\Delta\pi\eta(1-\eta)\pi_c^{*\prime\prime}\pi_d^{*\prime\prime} - (1-\eta)\left(\eta'\pi_c^{*\prime}\right)^2\pi_d^{*\prime\prime} - \eta\left(\eta'\pi_d^{*\prime}\right)^2\pi_c^{*\prime\prime} < 0$$

where the inequality holds by the convexity of the problem (assuming an interior solution).

Applying the implicit function theorem requires J^{-1}:

$$J^{-1} = \frac{1}{det(J)} \begin{bmatrix} \eta(1-\eta)\pi_c^{*\prime\prime}\pi_d^{*\prime\prime} & -(1-\eta)\pi_d^{*\prime\prime}\eta'\pi_c^{*\prime} & \eta\pi_c^{*\prime\prime}\eta'\pi_d^{*\prime} \\ -(1-\eta)\pi_d^{*\prime\prime}\eta'\pi_c^{*\prime} & (1-\eta)\Delta\pi\eta''\pi_d^{*\prime\prime} - \left(\eta'\pi_d^{*\prime}\right)^2 & -(\eta')^2\pi_c^{*\prime}\pi_d^{*\prime} \\ \eta\pi_c^{*\prime\prime}\eta'\pi_d^{*\prime} & -(\eta')^2\pi_c^{*\prime}\pi_d^{*\prime} & \eta\Delta\pi\eta''\pi_c^{*\prime\prime} - \left(\eta'\pi_c^{*\prime}\right)^2 \end{bmatrix}$$

Now, application of the implicit function theorem yields the following comparative statics:

$$\frac{\partial \rho_c^*}{\partial \eta_0} = -\frac{1}{det(J)}(1-\eta)\eta'\Delta\pi\pi_c^{*\prime}\pi_d^{*\prime\prime}\left(\frac{\eta''}{\eta'} - \frac{\partial^2\eta/\partial\lambda\partial\eta_0}{\partial\eta/\partial\eta_0}\right)\frac{\partial\eta}{\partial\eta_0} \tag{A.4}$$

$$\frac{\partial \rho_d^*}{\partial \eta_0} = \frac{1}{det(J)}\eta\eta'\Delta\pi\pi_d^{*\prime}\pi_c^{*\prime\prime}\left(\frac{\eta''}{\eta'} - \frac{\partial^2\eta/\partial\lambda\partial\eta_0}{\partial\eta/\partial\eta_0}\right)\frac{\partial\eta}{\partial\eta_0} \tag{A.5}$$

$$\frac{\partial \lambda^*}{\partial \eta_0} = -\frac{1}{det(J)}\eta(1-\eta)\pi_c^{*\prime\prime}\pi_d^{*\prime\prime}\left[\frac{\partial^2\eta/\partial\lambda\partial\eta_0}{\partial\eta/\partial\eta_0} - \left(\frac{\eta'}{\eta}\frac{\pi_c^{*\prime}}{\Delta\pi}\frac{\pi_c^{*\prime}}{\pi_c^{*\prime\prime}} + \frac{\eta'}{1-\eta}\frac{\pi_d^{*\prime}}{\Delta\pi}\frac{\pi_d^{*\prime}}{\pi_d^{*\prime\prime}}\right)\right]\Delta\pi\frac{\partial\eta}{\partial\eta_0} \tag{A.6}$$

In Eq. A.6, use the equilibrium FOCs to substitute out the η, $(1-\eta)$ and η' terms where they appear in the square brackets. Then, defining $A \equiv -\frac{1}{det(J)}\frac{\pi_c^{*\prime\prime}\pi_d^{*\prime\prime}}{\pi_c^{*\prime}\pi_d^{*\prime}}$ and K as given in the main text and rearranging yields the expressions given in Section 2. To obtain the signs discussed in Section 2, note that $\Delta\pi$ and η' are always the same sign, so that their product is always positive.

B Time series variation in the data

Time series variation in innovation and lobbying In the time series of firm investments, both R&D (Figure B.2a) and lobbying (Figure B.2d) expenditures are increasing in real terms for both treated and untreated firms. Additionally, levels of R&D and lobbying are both higher on average for treated firms than for untreated firms. As discussed in the main text, this points to the classic concern that, at least for outcomes in levels, selection into treatment is present both within the cross section as well as over time. Because the scientific discoveries of previously-unknown chemical harms that I compile only affect treated firms and are increasing in frequency over time (Figure 1b), both the cross sectional and temporal patterns exhibited here would contribute to an upward bias in the estimated effect of a scientific discovery on both innovation and lobbying investments by firms. I address these concerns in my empirical specification (Section 5).

As discussed in the main text, to account for scale effects, I normalize R&D investments by the firm's market capitalization (Figure B.2b) and lobbying investments by the firm's average lobbying expenditure[52] (Figure B.2e). These normalizations largely eliminate average cross sectional differences between treated and untreated firms. Temporal trends are reduced, but still present, and are not always parallel across treated and untreated firms (see e.g. the end of both panels in Figure B.2b and B.2e).

Because R&D expenditures often exhibit a unit root (which I also cannot reject, see Section 5), I use first differences in normalized outcomes in my empirical specification. The first-differenced data show no average cross-sectional differences over treated and untreated firms, and no differences in time trends (Figure B.2c and B.2f).

One additional detail in the data is worth pointing out. The first-differenced, normalized R&D data (Figure B.2c) exhibit a large positive spike for both treated and untreated firms in 2008 and a subsequent negative spike in 2009. This appears to be due to the Great Recession, which began in late 2008. Firm R&D budgets were largely set or perhaps largely spent, but their market values collapsed, leading to very high levels of innovation normalized by market capitalization in that year and a large first difference with the previous year. In the following

[52] As noted previously, the lobbying data does not contain any firm covariate information that can be used as a measure of firm scale.

year R&D expenditures fell and firm market values partially recovered, leading to low levels of normalized R&D and a large (negative) first difference. Interannual shocks such as these are controlled for by sector-year fixed effects in my empirical specification.

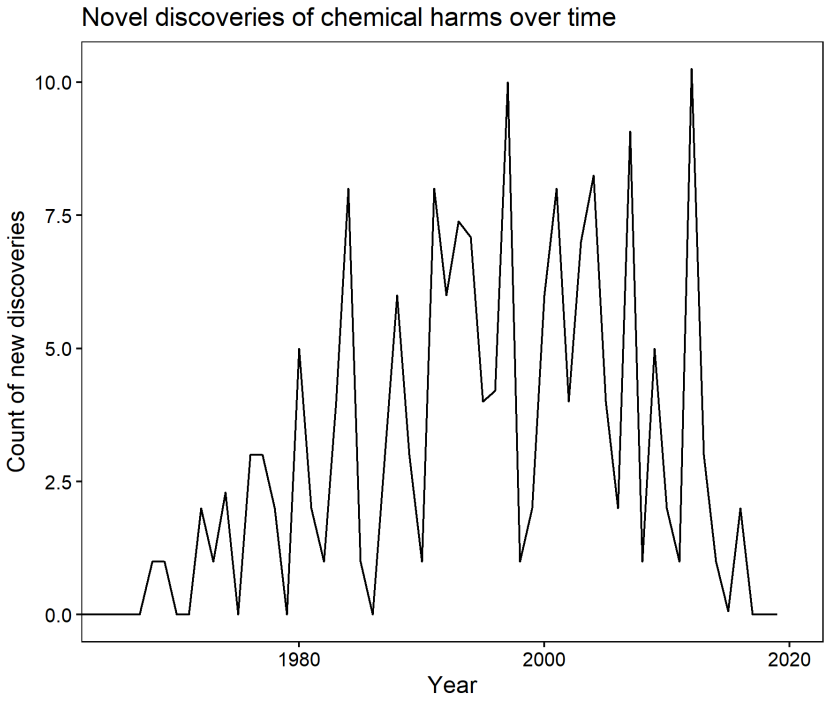


Figure B.1: **Time series of the count of discoveries of previously-unknown chemical harms by year.** Discoveries of harms are by independent epidemiological researchers, not R&D scientists employed at the sample firms. While the rate of new discoveries of chemical harms increases over time, discoveries are not themselves highly clustered in any one period and exhibit substantial interannual variation over all periods in the data.

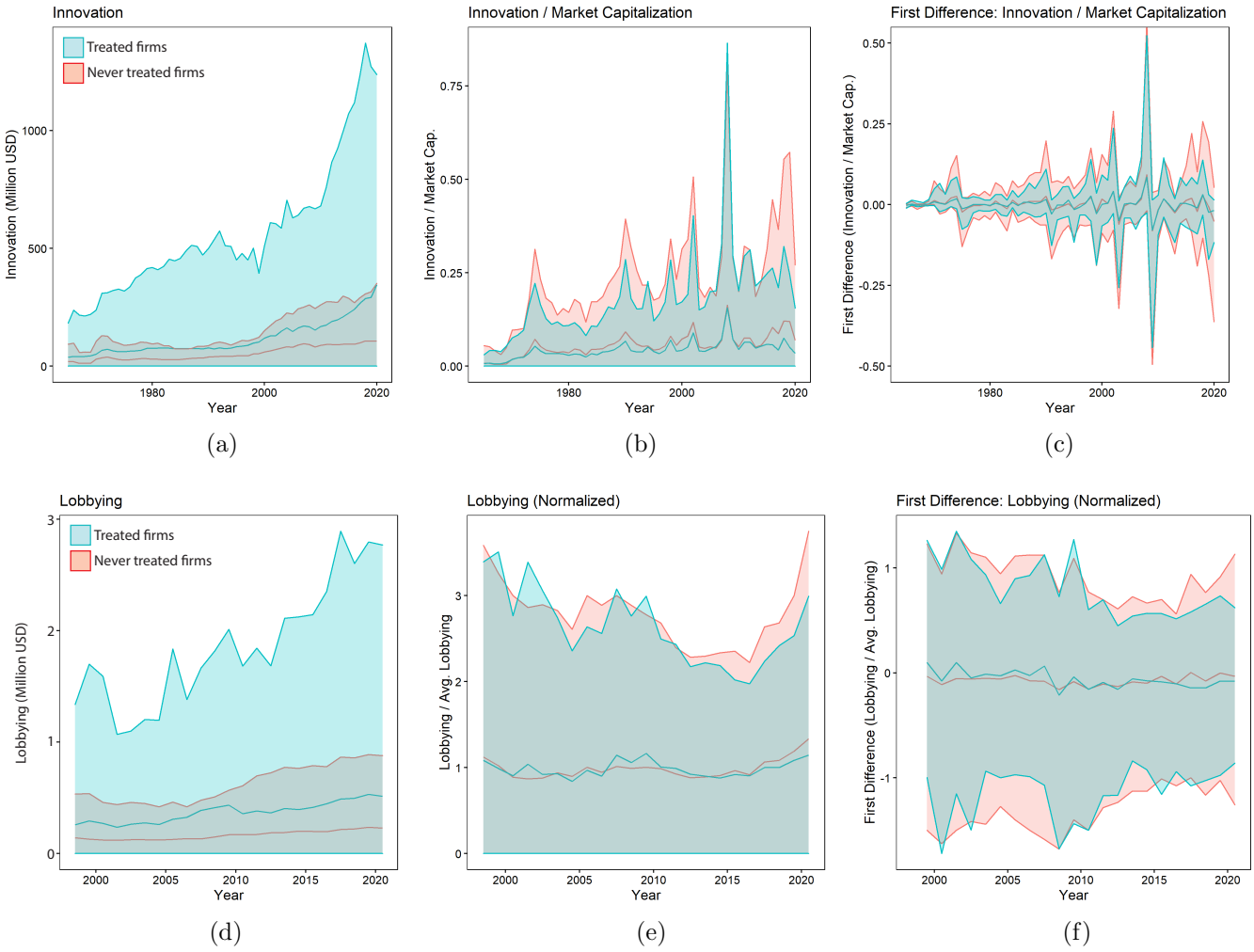
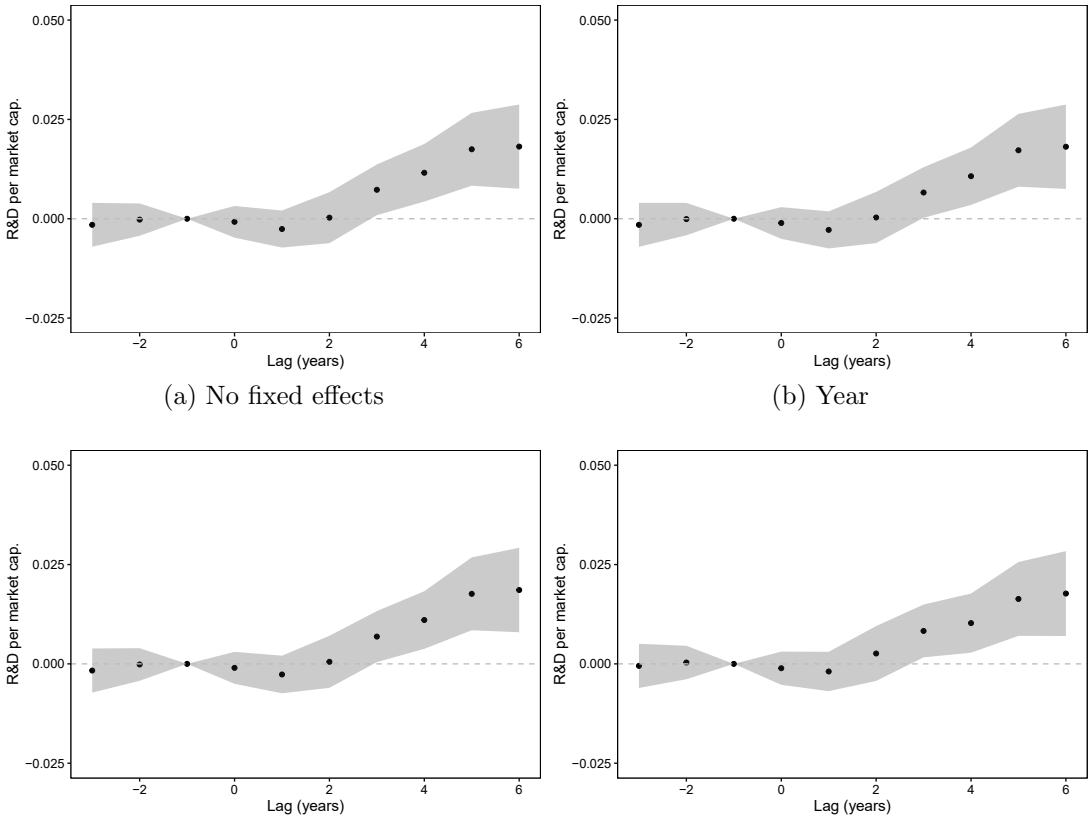


Figure B.2: **Time series variation in R&D and lobbying.** Time series variation in R&D (a-c) and lobbying (d-f) across treated firms (blue) and untreated firms (red). R&D and lobbying expenditures are shown in real terms. Levels of firm R&D and lobbying investments exhibit average differences in the cross section and over time (a and d). Normalizing for scale effects (b and e) largely eliminates average cross-sectional differences, but temporal trends persist and are not always parallel even across aggregate treated and untreated categories. First differencing the normalized outcomes largely eliminates temporal trends (c and f). Shaded areas reflect the 5^{th} and 95^{th} quantiles of each year's distribution over treated or never-treated firms.

C Robustness of the R&D response

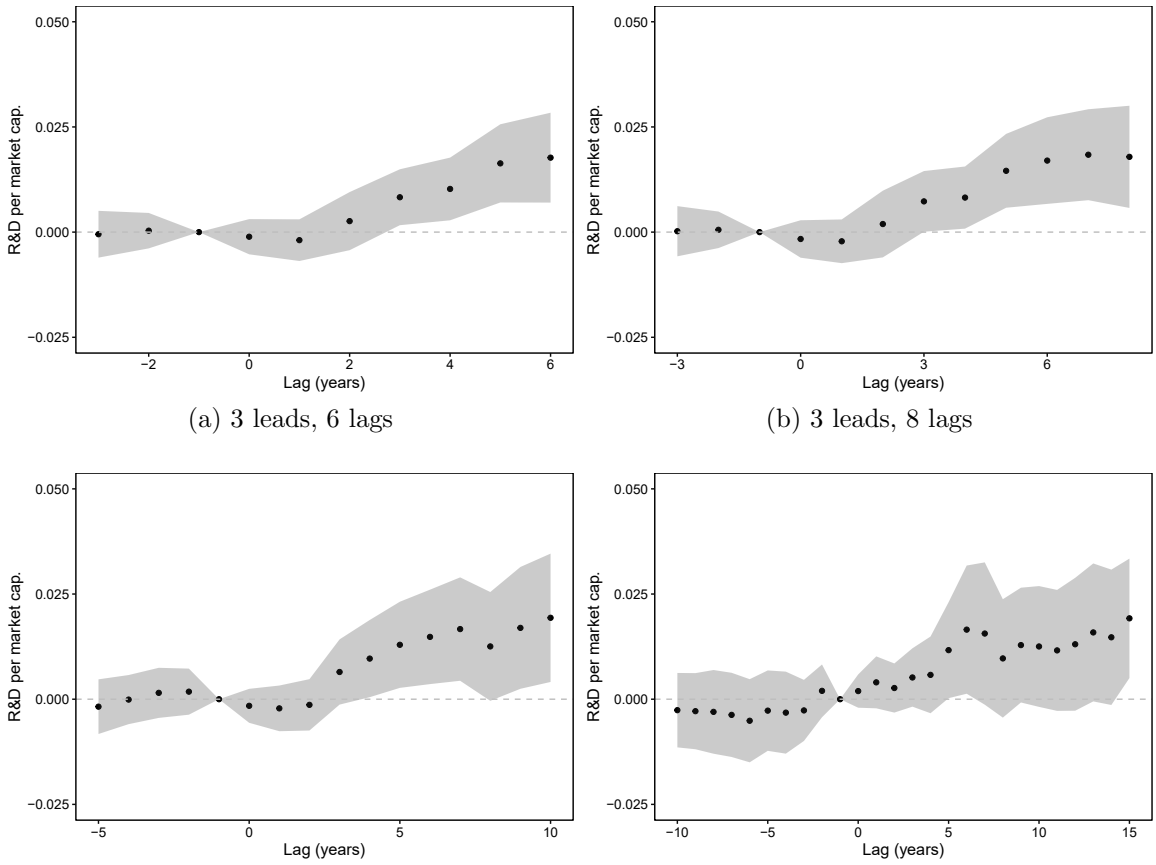
This section demonstrates robustness of the R&D response to alternative fixed effects specifications (Figure C.1) and lag specifications (Figure C.2), and to dropping individual discoveries (Figure C.3, only the most influential dropped discoveries shown).



(a) No fixed effects (b) Year

(c) Sector + year (d) Sector×year

Figure C.1: **Robustness of the R&D response to alternative fixed effects specifications.** Panel (d) is the preferred specification presented in the main text, included here for comparison. 95% confidence intervals are robust to arbitrary correlations in innovation investments within firm over time, and capture the full intertemporal covariance structure of the estimated dynamic effect.



(a) 3 leads, 6 lags

(b) 3 leads, 8 lags

(c) 5 leads, 10 lags

(d) 10 leads, 15 lags

Figure C.2: **Robustness of the R&D response to alternative lag specifications.** Panel (a) is the preferred specification presented in the main text, included here for comparison. All estimates are on a balanced panel, so the number of firms entering each estimate varies as the number of estimated leads and lags varies. 95% confidence intervals are robust to arbitrary correlations in innovation investments within firm over time, and capture the full intertemporal covariance structure of the estimated dynamic effect.

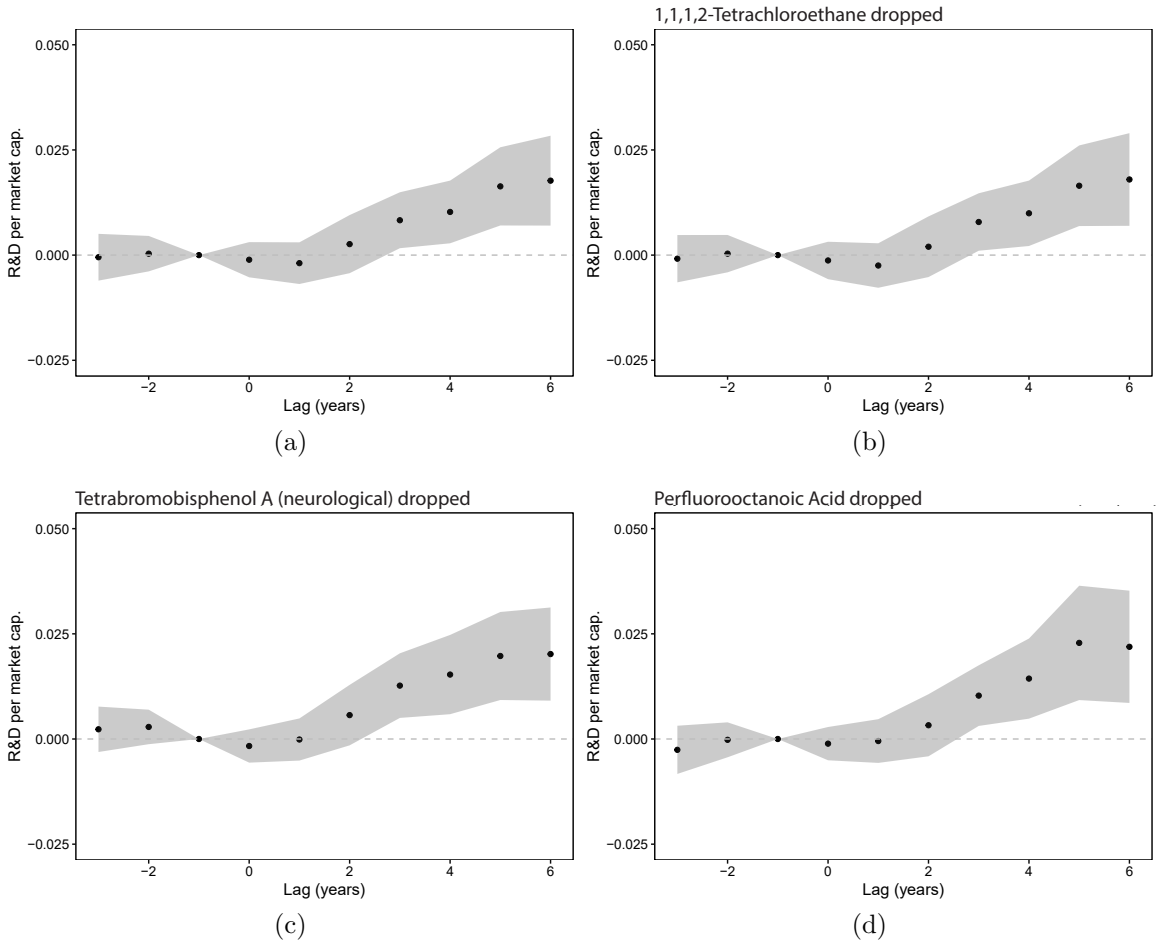


Figure C.3: **Robustness of the R&D response to dropping individual discoveries.** Panel (a) is the preferred specification presented in the main text, included here for comparison. Panel (b) is representative of dropping a random discovery. Panels (c) and (d) are the two discoveries that *most* affect the estimates when they are dropped. 95% confidence intervals are robust to arbitrary correlations in innovation investments within firm over time, and capture the full intertemporal covariance structure of the estimated dynamic effect.

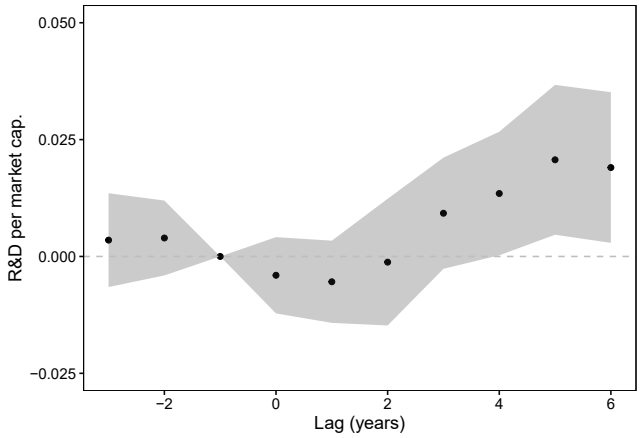
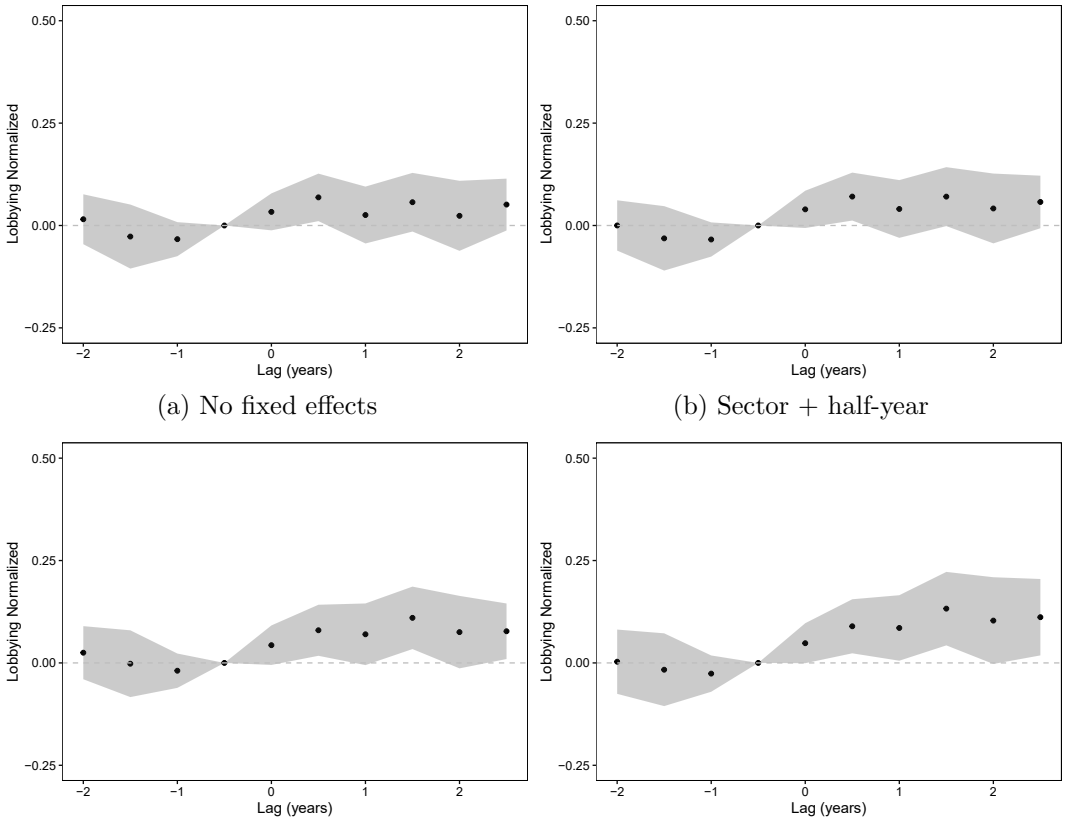


Figure C.4: **Robustness of the R&D response to normalizing controls at a common year.** Estimates follow the main specification, but normalize control firm innovation expenditures to market capitalization centered on a single common year (1980). In the main specification, control firm innovation is normalized to the firm's panel average market capitalization (this is done because control firms do not have any well defined "baseline" period). However, one might be concerned that such a normalization implies information about treatment shocks enters the denominator for some control firms. Estimates here demonstrate robustness to normalizing control firms over a very early panel year. Note that treated firms are always normalized to the period before the *first* treatment they are exposed to, avoiding such denominator concerns. Confidence intervals as in Figure 2.

D Robustness of the lobbying response

This section demonstrates robustness of the lobbying response to alternative fixed effects specifications (Figure D.1) and lag specifications (Figure D.2), and to dropping individual discoveries (Figure D.3, only the most influential dropped discoveries shown).



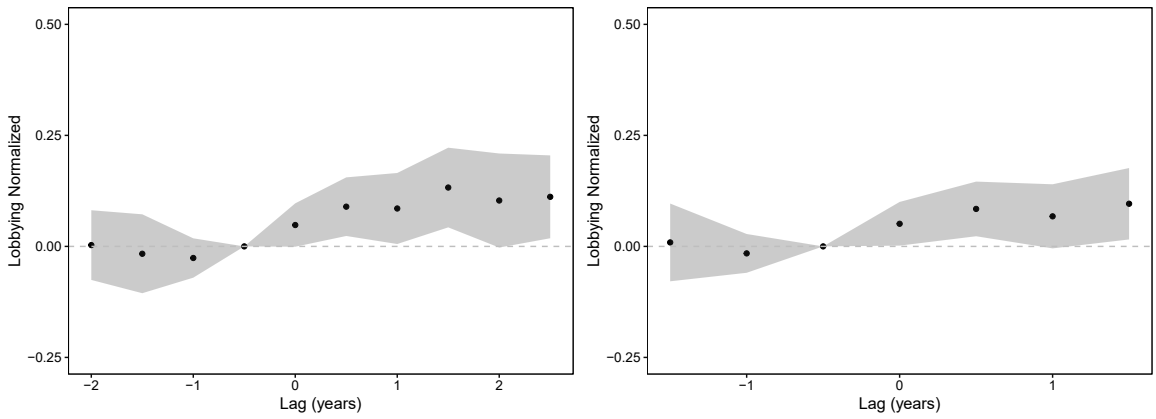
(a) No fixed effects

(b) Sector + half-year

(c) Sector×half-year

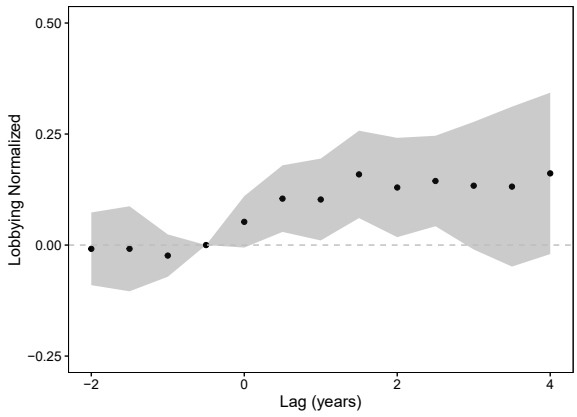
(d) Firm + Sector×half-year

Figure D.1: **Robustness of the lobbying response to alternative fixed effects specifications.** Panel (d) is the preferred specification presented in the main text, included here for comparison. 95% confidence intervals are robust to arbitrary correlations in lobbying investments within firm over time, and capture the full intertemporal covariance structure of the estimated dynamic effect.

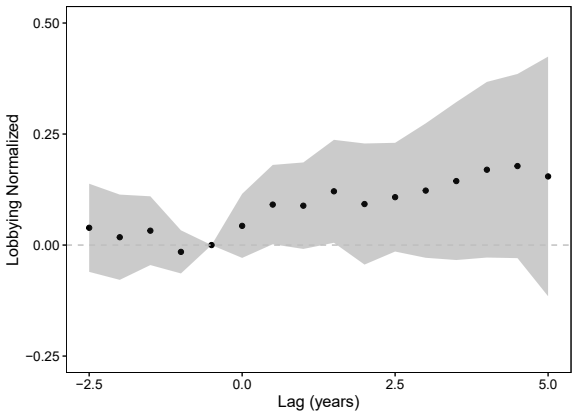


(a) 4 leads, 5 lags

(b) 3 leads, 3 lags



(c) 4 leads, 8 lags



(d) 5 leads, 10 lags

Figure D.2: **Robustness of the lobbying response to alternative lag specifications.** Panel (a) is the preferred specification presented in the main text, included here for comparison. All estimates are on a balanced panel, so the number of firms entering each estimate varies as the number of estimated leads and lags varies. 95% confidence intervals are robust to arbitrary correlations in lobbying investments within firm over time, and capture the full intertemporal covariance structure of the estimated dynamic effect.

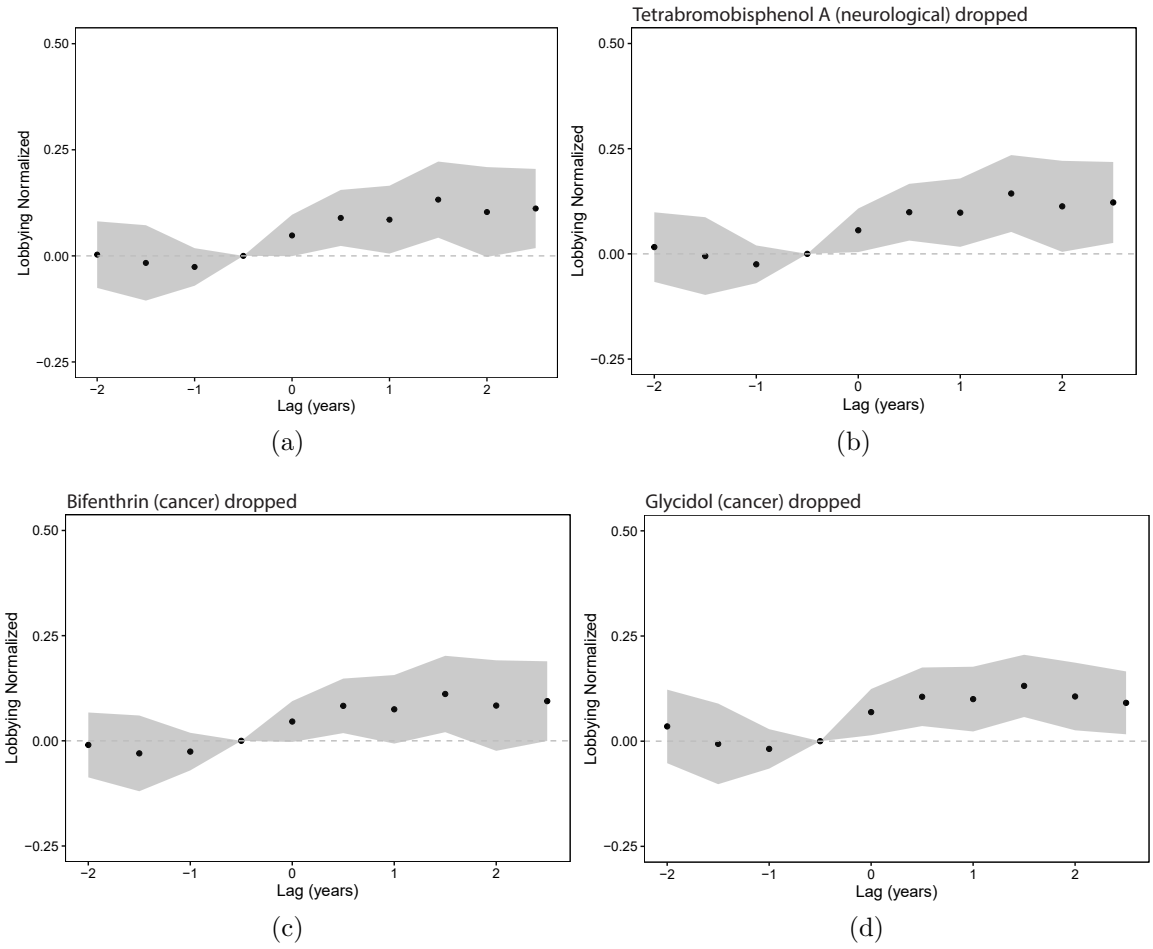


Figure D.3: **Robustness of the lobbying response to dropping individual discoveries.** Panel (a) is the preferred specification presented in the main text, included here for comparison. Panel (b) is representative of dropping a random discovery. Panels (c) and (d) are the two discoveries that *most* affect the estimates when they are dropped. 95% confidence intervals are robust to arbitrary correlations in lobbying investments within firm over time, and capture the full intertemporal covariance structure of the estimated dynamic effect.

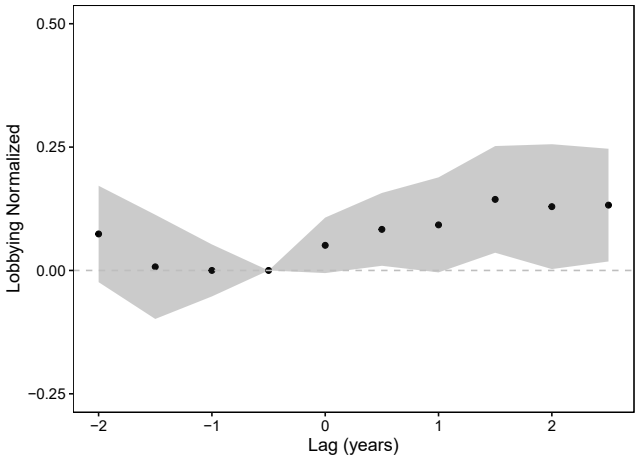


Figure D.4: **Robustness of the lobbying response to normalizing controls at a common year.** Estimates follow the main specification, but normalize control firm lobbying expenditures to a baseline centered on a single common year (2000). In the main specification, control firm lobbying is normalized to the firm's panel average lobbying (this is done because control firms do not have any well defined "baseline" period). However, one might be concerned that such a normalization implies information about treatment shocks enters the denominator for some control firms. Estimates here demonstrate robustness to normalizing control firms over a very early panel year. Note that treated firms are always normalized to the period before the *first* treatment they are exposed to, avoiding such denominator concerns. Confidence intervals as in Figure 3.

E The role of spillovers

As previously discussed, lobbying by outside actors shifts a common probability of regulation and thus should affect a given firm's innovation actions (Section 2), and indeed industry trade groups in the U.S. facilitate lobbying coordination (Section 3.3; (Bombardini and Trebbi, 2012, 2020)). In contrast, the role of innovation spillovers to firm lobbying is less clear. This section presents empirical tests for such spillovers, which requires matching firms across the Compustat and CRP datasets. Only a limited number of treated and control firms are matched across the datasets, which presents a power limitation in these spillover results.[53]

E.1 Tests for lobbying spillovers to firm R&D

The left panel of Figure E.1 replicates Figure 4, except for a subset of firms that could be directly matched across the Compustat and CRP datasets. In this panel, the red response includes an indicator for above median pre-period average lobbying within an industry. In this model, lobbying by all firms within an industry affects the innovation response of any given firm (as in Figure 4). This left-panel result closely mirrors the corresponding estimate using the full dataset (Figure 4), both in the magnitudes of the estimates and in the temporal structure. Confidence intervals are wider due to the limited data used in estimation. Notably, results here are still consistent with innovation and lobbying as substitutes in firm actions.

In the right panel of Figure E.1, the pre-period lobbying interaction is defined using a firm's own lobbying only, not the lobbying of other firms in the industry. The same firm-matched dataset is used in this panel as was used in the left panel. In this right panel, a firm's own lobbying has almost no measurable effect on its own innovation response. That is, the innovation response point estimates for both high-lobby and low-lobbying firms are quite similar in all periods, and frequently switch ordering.

Thus, for the same set of estimating data, the innovation response exhibits economically meaningful heterogeneity when industry-wide pre-period lobbying is used to model a differential

[53]Firm names are not standardized in the lobbying data, so matching firms across the innovation and lobbying datasets was not straightforward. To overcome this issue, I automatically queried Google for each firm name in each dataset. I then allowed the degree to which Google returned the same set of web addresses for any given pair of firm names to determine the degree to which those two firms were a match across the two datasets. Across the datasets, 2,894 individual firms were successfully matched (out of 20,688 firms in the innovation data and 41,080 firms in the lobbying data).

innovation response. But, this heterogeneity disappears when only firm-specific pre-period lobbying is used. This set of results suggests that spillover effects from industry-wide lobbying empirically dominate the effects of a firm's own lobbying on its own innovation.[54]

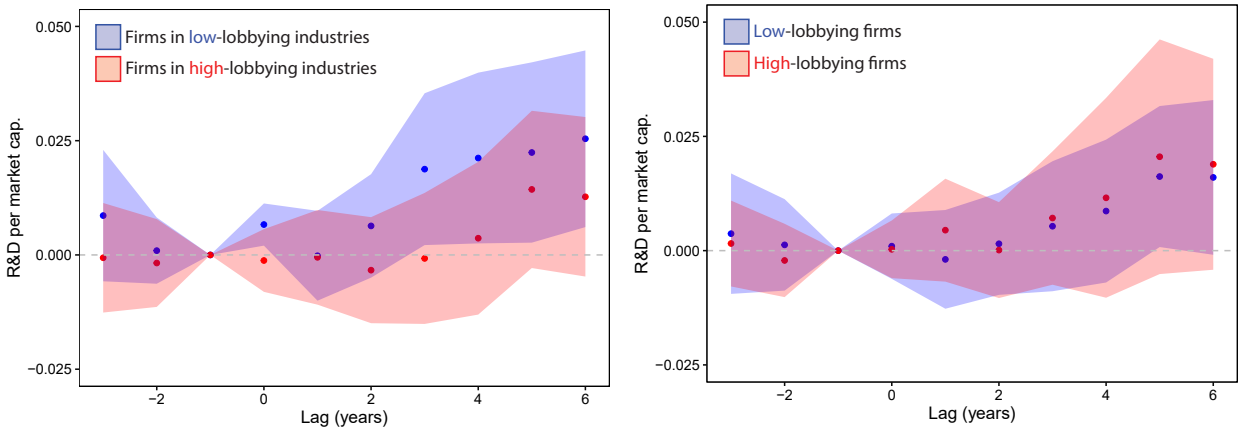


Figure E.1: **Lobbying spillovers to firm innovation.** Left: Analogous to Figure 4, with the pre-period lobbying defined as the average lobbying within a 4-digit SIC code, but for the firm-matched subset of the data. Right: Same as left, except pre-period lobbying is defined using a firm's own lobbying only. Heterogeneity in firm innovation responses is apparent in industry-wide lobbying (left) but not firm-specific lobbying (right), suggesting that lobbying spillovers are important drivers of firm innovation responses. In both cases, pre-period lobbying is a cross-sectional interaction term, thus the heterogeneity in effects shown here should be considered suggestive rather than causal. 95% confidence intervals are robust to arbitrary correlations in innovation investments within firm over time, and capture the full intertemporal covariance structure of the estimated dynamic effect.

E.2 Tests for R&D spillovers to firm lobbying

This section presents tests for spillover effects of firm innovation on other firms' lobbying investments, and is analogous to the lobbying spillover tests on firm innovation in Section E. However, results here were too unstable for inclusion in the main text.

esults here test for the presence of innovation spillovers in firm lobbying responses to a regulatory shock, using a limited dataset of firm-specific matches across the Compustat and Center for Responsive Politics datasets (see Section E). The left and right panels of Figure E.2 are directly analogous to the spillover test in Figure E.1. That is, the left panel displays an estimate of main specification for firm lobbying heterogeneity over pre-period sectoral innovation

[54]Appendix Figure E.2 presents an analogous test for spillovers from industry innovation to firm lobbying. Unfortunately, the limited firm-matched data lead to concerns regarding the stability of the estimated lobbying response, with large point estimates and standard errors. Taken at face value, the results in Figure E.2 indicate a possible role for innovation spillovers to firm lobbying, but these results are primarily included for completeness.

(Figure 5), but using the more limited firm-matched dataset. In this limited dataset, firms in low-innovation sectors (blue) appear to lobby more in response to a regulatory threat than do firms in high-innovation sectors (red), but results estimates show more noise than those in the full dataset. In the right panel of Figure E.2, heterogeneity is estimated over a firm's own pre-period innovation (rather than sectoral innovation in the left panel). Here, noise in the estimates limits the ability to interpret heterogeneity in firm responses. Not only are standard errors large, but point estimates also are large even in the pre-period, reaching (statistically insignificant) values of roughly 20%. Note that the y-axis of Figure E.2 is expanded relative to those in the main text to accommodate uncertainty in these estimates.

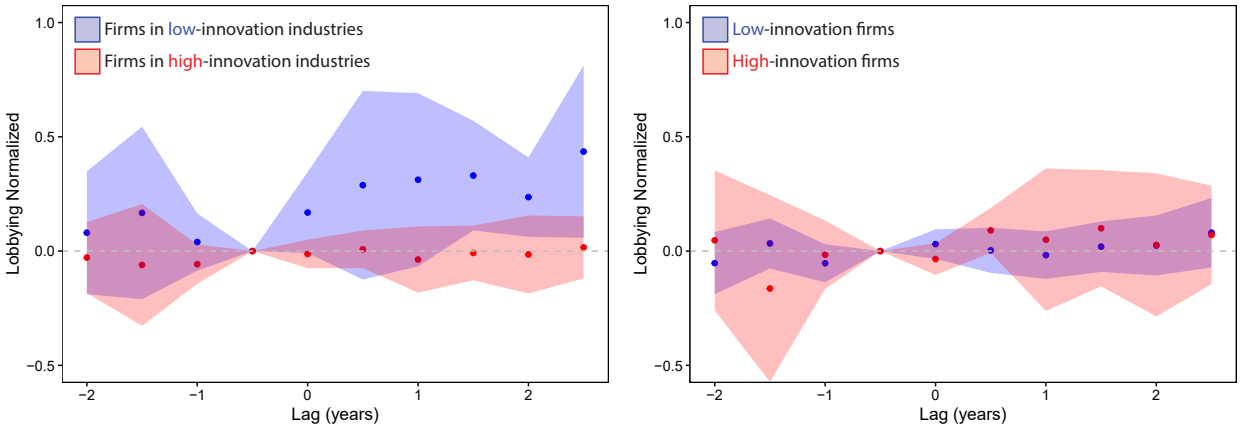


Figure E.2: **Innovation spillovers to rival firm lobbying are noisy.** Left: The main specification for firm lobbying heterogeneity over pre-period sectoral innovation (as displayed in Figure 5 with modified y-axis), but using a more limited dataset composed of individual firm matches across the Compustat and Center for Responsive Politics datasets. See Appendix E for the matching procedure. Firms in low-innovation sectors (blue) appear to lobby more in response to a regulatory threat than do firms in high-innovation sectors (red), but estimates in this limited dataset are subject to more noise — note the modified y-axis relative to other lobbying response figures. Right: Identical to left panel, except heterogeneity is estimated over a firm's own pre-period innovation (rather than sectoral average innovation). In this limited dataset, results are too noisy to distinguish any meaningful heterogeneity. In both panels, pre-period innovation is a cross-sectional interaction term, thus the heterogeneity in effects shown here should be considered suggestive. 95% confidence intervals are robust to arbitrary correlations in lobbying investments within firm over time, and capture the full intertemporal covariance structure of the estimated dynamic effect.

F Additional heterogeneity in R&D and lobbying responses

This appendix presents additional tests for heterogeneity in firm R&D and lobbying responses to the discovery of a novel externality.

F.1 Market concentration as a factor in firm R&D

A large literature exists studying the role of market concentration or market power in overall innovation throughout the economy (e.g., Arrow, 1962; Blundell, Griffith, and Van Reenen, 1999; Aghion et al., 2005; Cohen, 2010). Here I ask how market concentration, as measured using the Herfindahl-Hirschman Index (HHI), might mediate firm R&D and lobbying investments under a regulatory shock. Economic intuition suggests effects could go in either direction. For example, if the firm wins under regulation and has used regulation to raise barriers to entry in the form of new entrant marginal costs (e.g., Salop and Scheffman, 1983), then it may stand to lose less from new entry in the regulated state versus the unregulated state, and so the marginal payoff to innovation increases with regulation. However, if the firm wins under regulation but faces new entrants who are disruptive innovators (e.g., Blundell, Griffith, and Van Reenen, 1999) then it may stand to lose more from new entry in the regulated state and so the marginal payoff to innovation decreases with regulation. The effect on firm lobbying strategies is likewise ambiguous, depending on how the regulator weights concentrated lobbying efforts from a few firms versus diffuse lobbying efforts from many firms.

As is standard, I calculate the HHI each year at the 3-digit SIC code level using inflation-adjusted Compustat sales data, with negative values set to zero. *"High HHI"* is defined as an indicator for HHI values greater than 0.18, at which point the U.S. Department of Justice Antitrust Division defines a market as "highly concentrated."[55] The following regression examines heterogeneity over market concentration in firm R&D responses.

$$\Delta y_{it}^{R\&D} = \sum_{\ell \in b} \beta_{\ell}^{R\&D} LAG_{\ell it} + \sum_{\ell \in b} \xi_{\ell}^{R\&D} LAG_{\ell it} \times High\ HHI_{s3,t} \ + \ \phi_{st} + \epsilon_{it}$$

Estimated heterogeneity in the R&D response is noisy, but results (Figure F.1) suggest that

[55] See https://www.justice.gov/atr/herfindahl-hirschman-index.

firms operating in low-concentration markets with relatively less market power tend to innovate more in response to a discovery of a new externality than do firms in high concentration markets. In Figure F.1, low-concentration firms cumulatively increase R&D by 2.3% ($p < 0.05$) of pre-period market capitalization, but high-concentration firms only increase R&D by 1.1% ($p < 0.05$) of pre-period market capitalization. The R&D effect for low-concentration firms is roughly double that of high-concentration firms, consistent with theories and empirical findings of "creative destruction" associated with the innovative threat coming from a competitive fringe (Arrow, 1962; Blundell, Griffith, and Van Reenen, 1999; Cohen, 2010; Aghion et al., 2023).[56]

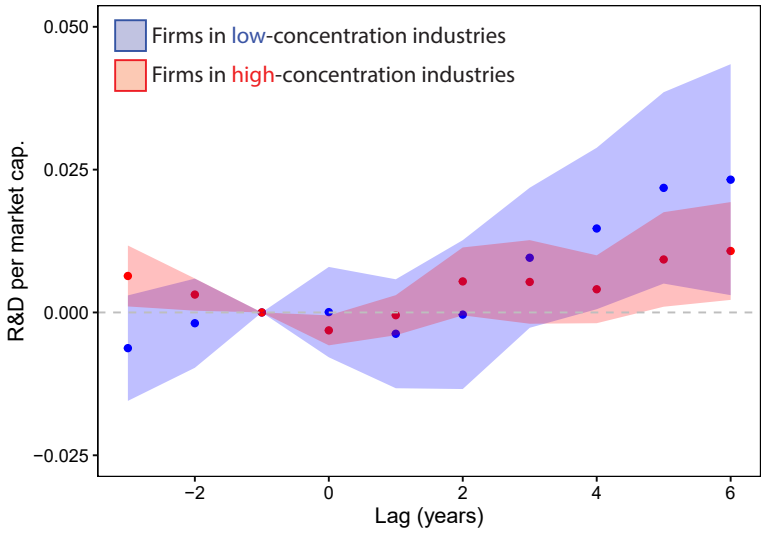


Figure F.1: **Heterogeneity over market concentration (HHI) in the cumulative effect of the discovery of a previously unknown externality on firm innovation.** "High concentration" is an indicator for HHI values of 0.18 or greater (see text for details). Results here suggest that firms operating in low-concentration markets tend to innovate more in response to the regulatory shock than do firms in high concentration markets. 95% confidence intervals are robust to arbitrary correlations in innovation investments within firm over time, and capture the full intertemporal covariance structure of the estimated dynamic effect.

As discussed in the next appendix section, analogous estimates for heterogeneity in firm lobbying over market concentration yield little meaningful insight, with both low- and high-HHI industries exhibiting similar patterns in their lobbying investments following an information

[56]Given that pre-period lobbying capacity was also found to be a driver of heterogeneity in the R&D response, one might be concerned about omitted variables in the interaction estimates. Testing for heterogeneity in R&D responses over pre-period lobbying while simultaneously controlling for "High-HHI" interactions yields results that are qualitatively similar to those of Figure 4. Note that this estimator pushes the data harder, with multiple interactions on each lagged estimate, and estimates correspondingly exhibit more noise. Similarity of the results suggests that market concentration is not an important omitted variable in the estimates of R&D response heterogeneity over pre-period lobbying capacity. Results available on request.

shock (Figure F.2). This includes both overlap in 95% confidence intervals as well as the magnitudes (and ordering) of the point estimates.

F.2 Heterogeneity over market concentration in the lobbying response

This section displays heterogeneity over market concentration (industry sector HHI) in firm lobbying responses to the discovery of a new externality. This estimate is analogous to the R&D response estimate of Figure F.1, but with firm lobbying as the outcome, and including a firm fixed effect in the first-differenced estimator. As can be seen, there is little meaningful heterogeneity to be estimated, and the result is reported here for completeness.

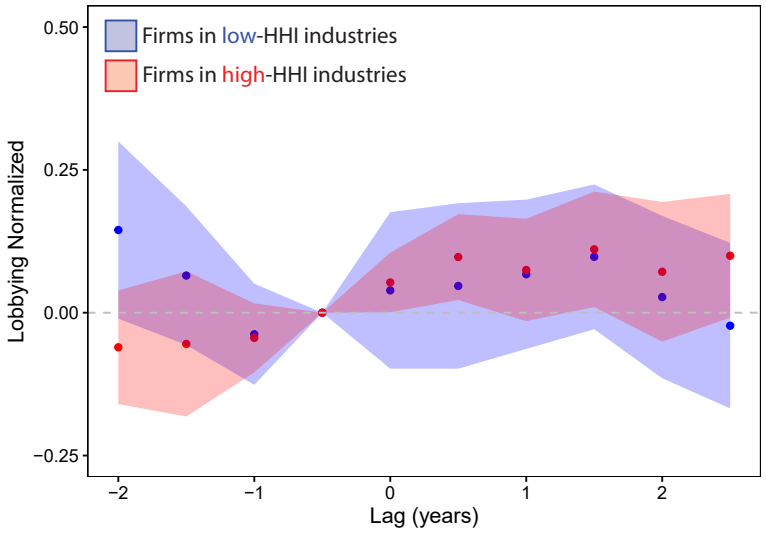


Figure F.2: **Heterogeneity over market concentration (HHI) in the cumulative effect of the discovery of a previously unknown potential externality on the lobbying investments of affected firms.** "High-HHI" defined as in Figure F.1. Little meaningful heterogeneity in the lobbying response over HHI is resolved. 95% confidence intervals are robust to arbitrary correlations in lobbying investments within firm over time, and capture the full intertemporal covariance structure of the estimated dynamic effect.

G Non-scraped discoveries

Early work involved a small number of discoveries chemical harms that were manually identified, before the process for this analysis of scraping the literature was developed. Manual identification of discoveries involved reading literature surveys and media reports to identify the earliest publications associated with previously unknown, low-dose, chronic health effects. Those discoveries are documented here.

Chemical	Discovery date	Source	Description
Chlorofluorocarbons (CFCs)	6/28/1974	Molina and Rowland (1974)	Aerosol propellant, coolant, electronics cleaning
Bisphenol A (BPA)	6/1/1993	Krishnan et al (1993)	Accidental discovery; widely use plasticizer
Parabens	11/1/1998	Routledge et al (1998)	Preservative used in cosmetics, toiletries, and some foods
GenX	6/30/2015	Rae et al (2015)	Used in production of Teflon, other fluorinated polymers
Perfluorooctane sulfonate (PFOS)	9/2/1997	Anderson and Mulvana (1997)	Fabric stain repellent, cleaning products, firefighting, metal plating and semiconductor cleaning
Perfluorooctanoic acid (PFOA)	9/1/1993	Gilliland and Mandel (1993)	Similar uses as PFOS; voluntarily phased out (announced in May 2000)
Methyl tert-butyl ether (MTBE)	9/30/1994	Moolenaar et al (1994)	Blending agent in gasoline; solvent in plastics and pharaceuticals
Fracking – Water contamination	12/31/2011	EPA (2011)	EPA draft study investigating water quality complaints
Fracking – Earthquakes	8/31/2012	BCOGC (2012)	British Columbia Oil and Gas Commission investigation
Glyphosate – Weed resistance	7/10/1996	Pratley et al (1996)	Conference report reporting weed resistance
Glyphosate – Human exposure	11/2/2004	Benedetti et al (2004)	Low dose chronic exposure

Table G.1: **Discoveries that were manually identified.** This table lists discoveries that were manually identified, before the process for this analysis of scraping the literature was developed. Manual identification of discoveries involved reading literature surveys and media reports to identify the earliest publications associated with previously unknown, low-dose, chronic health effects.