Innovation and lobbying in response to regulatory uncertainty

Andrew Hultgren*

January 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 regulatory expectations: 104 initial discoveries of chemical harms affecting 60 different industries over multiple decades. I find shocked firms slowly increase total innovation expenditures by 1.8% of pre-period market capitalization – the first firm-level documentation of an aggregate innovation expenditure regulatory response. However, firms also increase lobbying by 11% of pre-period levels. Suggestive evidence indicates firms may treat innovation and lobbying as substitutes, implying firms' option to lobby may reduce their innovation response.

JEL Codes: D72, D22, O31, O33, Q55, Q58, 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, Jacquelyn Pless, 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 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 regulation on firm innovation expenditures and on firm lobbying, and how does the option to lobby affect firm innovation?

Large literatures have separately studied firm innovation (Bloom et al., 2002; Popp, 2006; Aghion et al., 2016; Bloom et al., 2019; Akcigit et al., 2022; Dugoua, 2023) or lobbying (Grossman and Helpman, 1994; Gehlbach, Sonin, and Zhuravskaya, 2010; Kang, 2016; Meng and Rode, 2019; Bertrand et al., 2020) responses to regulatory and tax policy. The contribution of this paper is to recognize that firms jointly optimize over their innovation and lobbying responses, and to estimate these responses to shocks over the future regulatory state, employing for identification a newly-constructed dataset of external information shocks which mitigates against forward-looking firm anticipation effects. 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. Independent epidemiological researchers were left to test for such harms, and I scrape over 7,000 studies to create a novel dataset of *initial* scientific discoveries of harms from widely used industrial chemicals. These discoveries were made by independent epidemiologists, not firm scientists, and I exploit them as a natural experiment for identification.[1] I pair this novel data with widely used panel data on firm innovation and lobbying expenditures.[2] Matching the timing of discoveries to the industries they affect, I employ

[1]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.

[2]Firm innovation is often measured using either R&D expenditures or patenting activity (Cohen, 2010; Popp, Newell, and Jaffe, 2010). Here I focus on expenditures data from Compustat as it parallels the structure of firm lobbying data (which is also reported as firm-level expenditures) and ready estimates of its returns to the firm exists (Hall, 1999; Bloom, Schankerman, and Van Reenen, 2013). Additionally, because expenditures data aggregates across a firm's innovation activities, it is less vulnerable to within-firm reallocation of innovation effort (Dechezleprêtre and Sato, 2017).

a distributed lag model in first-differences to estimate the dynamic responses of affected firms' innovation and lobbying expenditures. Identification rests on within-firm variation in exposure to discoveries of previously unknown health harms, controlling for sector-wide temporal shocks common to both treated and untreated firms. That is, conditional on sector-wide common information and firm-average private information, I assume a given firm does not know when the next health harm discovery by independent researchers will occur, nor if that discovery will affect its own product offerings versus those of firms operating in other industries.

To fix ideas, consider a well-documented example (Benedick, 1991; Dugoua, 2023). In the 1950s and 60s 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. Critically, such lobbying also affects expected innovation payoffs. As discussed below, DuPont might lobby for or against regulation, implying innovation and lobbying may be complements or substitutes in the firm's actions.[3] 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.

The tradeoffs between innovation and lobbying faced by firms apply to a range of contexts outside of pollution externalities. For example, they apply to the field of 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. Further, the empirical approach employed here can readily be extended to other measures of innovation (e.g. patents) and other forms of regulatory influence.

A simple conceptual framework motivates the problem, in which a monopolist facing regulation can invest in clean or dirty innovation, but can also lobby for or against regulation. The firm invests in clean and dirty innovation until the respective *expected* marginal payoff equals

[3] In fact DuPont both innovated and lobbied in response to the regulatory threat, initially lobbying *against* regulation to protect its existing profits and later lobbying *for* regulation once it perceived a competitive advantage in the market for substitutes.

the marginal investment cost. Importantly, firm 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 find that in response to a new discovery of a chemical harm, firms increase their innovation 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. The use for identification first-time discoveries of chemical harms, representing external shocks to firm information, 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 innovation or lobbying investments. Importantly, while regulatory-induced increases to firm *clean* innovation are well-documented (Popp, 2019), the possibility of a net decrease in *total* innovation – driven by substitution out of other innovation activities – has been difficult to exclude (Dechezleprêtre and Sato, 2017). To my knowledge, the firm innovation response estimated here represents the first firm-level documentation of an aggregate innovation expenditure response to regulation.

Testing for heterogeneity in the dynamics of firm responses, I find suggestive evidence that both industry concentration and lobbying capacity may explain heterogeneity in firm innovation responses. Lobbying capacity appears to be a particularly important descriptor of heterogeneity in the innovation response, with the majority of the average response driven by low lobbying-capacity firms. Innovation capacity may drive some variation in firm lobbying responses, but heterogeneity in the lobbying response is not statistically detectable. I also find suggestive evidence that within-industry lobbying spillovers may play an important role in firm innovation responses. Though not the main focus of this analysis, the evidence I find for rival firm lobbying spillovers to focal firm innovation complements and extends the literature on knowledge spillovers in firm innovation decisions (e.g., Bloom, Schankerman, and Van Reenen, 2013; Arora, Belenzon, and Sheer, 2021), documenting for future research a new channel through which firm innovation may be influenced by spillovers from rival actions.

Across all tests employed, results are consistent with innovation and lobbying as substitutes within the firm, 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. This finding is consistent in spirit with that of Akcigit, Baslandze, and Lotti (2023), who descriptively find that market-leading Italian firms are more likely to hire local politicians and also tend to patent less, relative to their competitors.[4] 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.[5]

The remainder of the paper proceeds as follows: Section 2 further contextualizes this study in the literature, Section 3 motivates the problem with a simple conceptual framework, Section 4 describes the construction of the novel dataset of epidemiological discoveries of unknown chemical harms and gives details on the innovation and lobbying expenditure data, Section 5 presents a descriptive analysis, Sections 6 and 7 describe the empirical models and results, Section 8 gives a welfare interpretation, and Section 9 concludes.

2 Firm Innovation and Lobbying Under Regulation

A large literature has examined the effect of environmental regulation on firm innovation (see Popp, Newell, and Jaffe (2010) and Popp (2019) for recent surveys). This literature often studies the effect of the introduction of a given regulation on firm patenting (e.g. Lanjouw and Mody, 1996; Popp, 2006; Calel and Dechezleprêtre, 2016; Brunel, 2019; Dugoua, 2023), though innovation expenditures and new product development have also been studied (e.g. Jaffe and Palmer, 1997; Newell, Jaffe, and Stavins, 1999). Findings of increased patenting in response to new environmental regulations appear to be robust – especially when focusing on patents directly related to the production activity being regulated (Popp, Newell, and Jaffe, 2010).

The theoretical literature on the political economy of lobbying goes back to Grossman and Helpman (1994), in which lobby groups pay off the planner to obtain trade protections. Empirical work in the area has ranged from estimating the effect of corporate lobbying activities 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

[4]See Section 2 for further discussion of Akcigit, Baslandze, and Lotti (2023), whose causal estimates are distinct from the outcomes studied here.

[5]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.

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).

The closest work in spirit to this analysis is that of Akcigit, Baslandze, and Lotti (2023), who study innovation implications of political connections (hiring of local politicians) among Italian firms.[6] They descriptively find market-leading firms are more politically connected but less likely to patent overall.[7] Distinct from this study, Akcigit, Baslandze, and Lotti (2023) do not study firm lobbying expenditures, innovation expenditures, or responses to regulatory shocks. Even so, their documentation of reduced innovation among politically connected firms overall is directionally consistent with evidence in this study suggesting innovation and lobbying are substitutes in firm responses to the prospect of future regulation.

Importantly, none of these works in either literature considers that innovation and lobbying may be distinct margins along which profit-maximizing firms simultaneously optimize in response to the prospect of future regulation. Further, the literature at times studies events (such as new regulations) that are announced years in advance and to which forward-looking firms may have already begun to respond. This anticipation effect implies event-study style estimates may understate the true causal effects (Rittenhouse and Zaragoza-Watkins, 2018). Compiling a dataset of initial scientific discoveries by independent scientists of previously-unknown chemical harms, I use for estimation external shocks of novel information over the possibility of future regulation, and thus seek to mitigate anticipation effects in estimation.

Finally, there is a literature concerned with rent-seeking as a substitute for entrepreneurship (Murphy, Shleifer, and Vishny, 1991, 1993). This theoretical literature is similar in spirit to the types of tradeoffs I consider, and is concerned with appropriative activities that might have a higher payoff than entrepreneurship, limiting economic growth in favor of outright corruption. Correlational work by Lenway, Morck, and Yeung (1996) is the closest empirical application to

[6]Similarly motivated work studies regulatory capture and firm growth (Slinko, Yakovlev, and Zhuravskaya, 2005) and business leaders running for political office as a channel for firm influence (Gehlbach, Sonin, and Zhuravskaya, 2010).

[7]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 innovation and lobbying trade-off that I estimate. There, the authors study U.S. steel mills' stock price changes over four separate announcements of possible trade protection from 1977 to 1984. They find that steel mills that lobbied more on average had lower average expenditures on innovation but higher stock returns to announcements of potential protection, consistent with lobbying as rent-seeking that acts to decrease the returns to innovation in their context.

3 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.[8] 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

[8] 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).

action: $q_r^* = \text{argmax}_{q_r}\, \pi(q_r, \rho_r)$, which implies the payoff $\pi_r^* = \pi_r^*(\rho_r)$.[9] The stage 1 problem 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"[10] ($\Delta\pi > 0$) 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

[9]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).

[10]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).

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 $\eta = \eta(\lambda, \eta_0)$. Shocks 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.[11] 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[12] 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

[11] 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.

[12] 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 $\frac{\eta'}{\eta}\frac{\pi_c^{*\prime}}{\pi_c^{*\prime\prime}} < \frac{\eta'}{1-\eta}\frac{\pi_d^{*\prime}}{\pi_d^{*\prime\prime}} < 0$.

square brackets in Eq. 2 is negative.[13] However, if the shock to η_0 has a large relative impact on 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 (Section 7.4).[14]

	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	Fig. 6
R&D spillovers to lobbying	n/a	n/a	n/a	Fig. E.1

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.*

[13] That is, if the "curvature-like" effect of the shock to η_0 is larger in magnitude than the sum of the clean and dirty innovation payoff curvatures, weighted by the squares of the respective marginal innovation returns as a fraction of the profit wedge.

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

4 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. 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.

4.1 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)).[15] 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 — who have no private incentive to produce it.[16] Thus, there was often insufficient data to prove

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

[16]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.

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.[17]

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.

4.2 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 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

[17] 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.

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,[18] though many studies last for one to two years.

This process resulted in 93 scientific discoveries of 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.[19] So, cyanazine in my data has two discoveries associated 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.

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

[19] 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 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..."

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.[20] The chemicals database PubChem[21] 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.

4.3 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.

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.[22] The reported financial data are wide-ranging and include innovation expenditures (XRD), market capitalization ($MKVALT$)[23], 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, Lakon-

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

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

[22]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.

[23]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.

ishok, and Sougiannis (2001).[24] 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 7.3 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., starting in the year 1998 any lobbying activities in excess of $200 have been legally required to be reported semi-annually to the Senate Office of Public Records (SOPR). CRP collects this data from SOPR and collates it in a more accessible format. I primarily use the CRP "Lobbying" data file in this analysis. These data specify the lobbyist ($REGISTRANT$), the client firm paying for the lobbying ($CLIENT$), the amount spent on lobbying ($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

[24] 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.

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 7.3 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 Section 7.4 of this study are constructed to account for such information spillovers.

5 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 (Figure F.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 electonics 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.[25]

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 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

[25]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 July 1998, so no treated firms enter the data before the year 2000 as none have the necessary number of leads for estimation (see Section 7).

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 Section 7.4.3). 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.

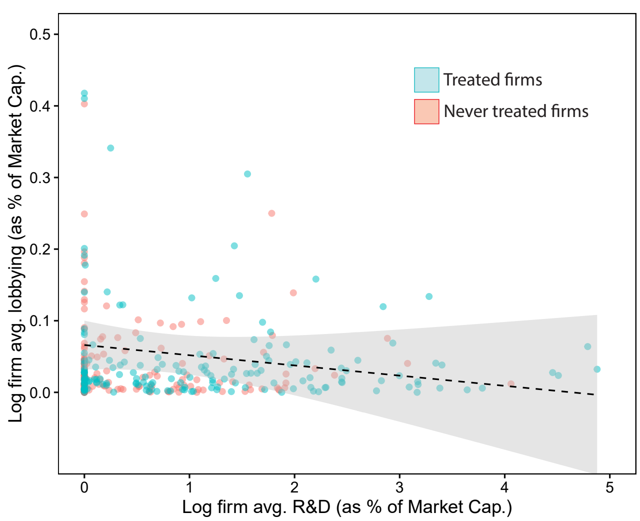


Figure 1: **Cross sectional variation in firm innovation and lobbying investments**, for a subset of firms that can be matched across the Compustat and CRP datasets (see Section 7.4.3). Normalized innovation and lobbying are averaged by firm. The plot displays log(100 × *firm average value* + 1). A firm 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.

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.

6 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.

6.1 Estimating the innovation and lobbying responses

Comparative statics in Section 3 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.[26]

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 ϕ_{st} reflecting both stochasticity in the economy as well as sectoral regulatory trends and shocks, and a noise term ϵ_{it}. This yields the estimating equation

$$Y_{it} = \rho Y_{it-1} + SCALE_{it} \times \bigg(\beta r_{it} + \phi_{st} + \epsilon_{it}\bigg).$$

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

[26] 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.

expenditures.[27] This leads me to first-difference innovation and lobbying expenditures, to eliminate potential bias from a unit root process (Greene, 2003). 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, giving

$$\frac{\Delta Y_{it}}{SCALE_{it}} = \beta r_{it} + \phi_{st} + \epsilon_{it}.$$

Finally, three additional points remain. First, 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.[28] Second, firm responses r_{it} may evolve over multiple periods. Such a dynamic response might occur if the perceived probability of regulation evolves after the initial shock (as new evidence comes to light or due to strategic actions by the players), or if 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. Third, firms may be differentially exposed to a given information shock. I 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. That is,

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

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

[27] 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.

[28] As discussed in Section 4.3, 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.

sector s and period t is given as

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

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

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.[29] 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 within-firm, within-sector-year deviations in innovation and lobbying investments, comparing the within-firm trajectories of treated firms to those of similar, untreated firms before and after an information shock. Standard errors are cluster-robust by firm, allowing for arbitrary correlations in innovation and lobbying activities at the firm level over time.[30]

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

[29] Typically, $\sum_{\tau_{ij} \in trt\_years} \mathbb{1}\{t - \ell = \tau_{ij}\}$ is either 0 or 1. Either the firm was 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.

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

6.2 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.[31] To address this simultaneity bias, I take averages of pre-period innovation and lobbying investments as measures of pre-existing 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 pre-existing 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).

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

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

$\textit{High Pre-Lobby}_i$ and $\textit{High Pre-R\&D}_I$ are indicators for pre-period lobbying and innovation being above the median value for treated firms. Compustat reports firm sales across industries, thus $\textit{High Pre-Lobby}_i$ in Eq. 5 is based on the firm-specific, sales-share weighted average of pre-period industry lobbying. In Eq. 6, $\textit{High Pre-R\&D}_I$ 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.[32] 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

[31] 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.

[32] Section 7.4.3 further examines the potential role of spillovers.

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. Because the specifications are separate estimates across different outcomes, they need not agree on sign and may empirically support opposing conclusions.

7 Results

7.1 The innovation response

Estimating cumulative innovation effects from Equation 3, I find that firms slowly increase total innovation investments in response to the scientific discovery of a previously unknown externality (Figure 2, Table 4). In the preferred specification, innovation investments increase by a highly significant ($p < 0.01$) and economically meaningful 1.77% of a firm's pre-period market capitalization (column 5 of Table 4) over the six years following a shock.[33] This six-year cumulative effect reflects an average annual flow increase in innovation investments of about 0.30% of firm market capitalization per year, a 7.7% increase over the average normalized innovation flow of 3.83% per year for treated firms in my sample.

Notably, while regulatory-induced increases to firm clean innovation (particularly patenting) are well-documented (Popp, 2019), the possibility that substitution out of other innovation activities drives a net decrease in total innovation has been difficult to rule out (Dechezleprêtre and Sato, 2017). To my knowledge, the estimates here represent the first firm-level documentation of an aggregate increase in regulatory-induced innovation expenditures.

The distributed lag model employed here estimates the overall effect of a shock to the prospect of future regulation as well as the dynamic evolution of firm responses. Each point in Figure 2 (and row in Table 4) represents the cumulative average treatment effect of a scientific discovery of a previously unknown externality on the innovation investments of affected firms ℓ years after the discovery. Time $\ell = 0$ is the year of publication of the discovery (by independent

[33] Recall that this pre-period is defined for treated firms in all estimates as the period before the *first* shock a given firm is exposed to, not the period before a given shock (Section 4.3).

	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. 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 B.1.

Estimates multiplied by 100 for interpretability.

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

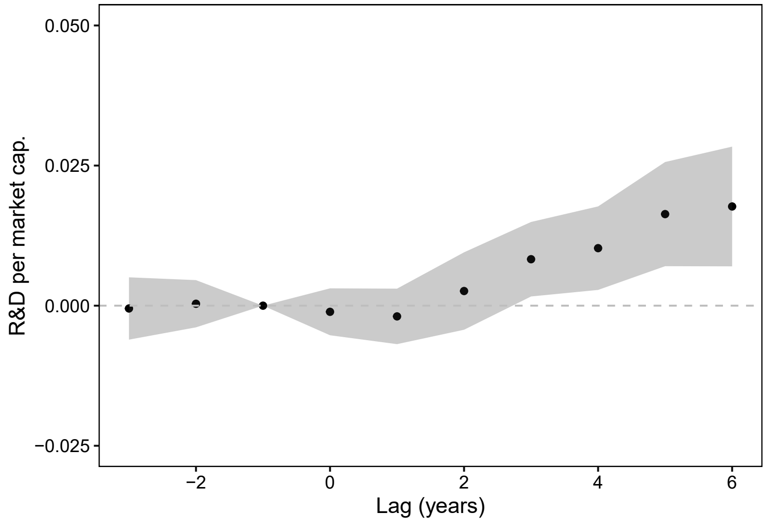


Figure 2: **The cumulative effect of the discovery of a previously unknown externality on firm innovation investments.** 95% confidence intervals are robust to arbitrary correlations in innovation investments within firm over time, and capture the full intertemporal covariance structure of the estimated dynamic effect. Time 0 is the year of publication of a discovery. Estimates represent changes in total firm innovation expenditures and not simply within-firm reallocation. Estimates correspond to column 4 of Table 4.

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 a preferred bandwidth of three leads and six lags, while the final column examines robustness to a wider bandwidth of ten leads and lags.[34] Bandwidth decisions are discussed further below. 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.[35] The number of balanced treated firms is reported in Table 4 below each set of estimates. Finally, all estimates are reported for a shock that affects 100% of firm sales ($INTENSITY_{it}$ evaluated at unity in Eq. 3).[36]

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

[34]Column 5 of Table 4 is truncated for exposition; see Appendix B for full estimates.
[35]Firms that are treated but not balanced over all leads and lags are excluded from all estimates.
[36]Treated firms' average treatment intensity in the Compustat data is 52%, and the 75^{th} percentile is 98%.

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. Further, firms might be aware of such trends and preemptively make defensive innovation investments in response. Columns 2 – 4 of Table 4 progressively relax this identifying assumption, conditioning on year fixed effects to control for common temporal patterns in scientific discoveries, sector fixed effects to control for any trending tendency of discoveries to focus on one sector of the economy over another, and finally joint sector-year fixed effects (preferred specification, column 4) to allow for sector-specific temporal shocks in discovery rates that may arbitrarily change over time. Cumulative estimates are remarkably stable across columns 1 – 4 and highly significant 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.

Estimates in column 5 of Table 4 push the data harder, estimating the model over a wide bandwidth of ten years of leads and lags (as previously mentioned, see Figure B.1 for full estimates). Estimating this wide bandwidth sacrifices power: apart from the obvious fact that more coefficients are estimated, it is also the case that fewer treated firms are balanced over this wider bandwidth. Even so, examination of these wide-bandwidth estimates yields several important findings. First, the 10-year cumulative effect (1.81%, column 5) is nearly identical to the six-year estimate (1.77%, column 4), indicating a high degree of stability in the overall effect. Second, lead estimates are flat over the entire ten year period preceding a discovery. This strongly suggests discovery timing is not anticipated by firms, buttressing the identification argument. Third, the cumulative innovation effect remains significant over this 10-year window ($p < 0.05$), implying the effects estimated here are permanent, rather than transitory, increases in total firm innovation.[37] Finally, Figure B.1 (left panel) indicates that the 10-year cumulative effect stabilizes around 1.5 – 2 percentage points, and reaches this stable point about five to six years after a discovery, after which average flow increases in innovation cease. 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.

7.2 The lobbying response

Firms' average lobbying response to the discovery of a previously unknown externality (Figure 3, Table 5) is estimated in a similar fashion to the innovation response above. Because the lobbying data are reported every half year, each point in Figure 3 and row in Table 5 represents a six-month period (versus the annual periods shown in Figure 2 and Table 4).

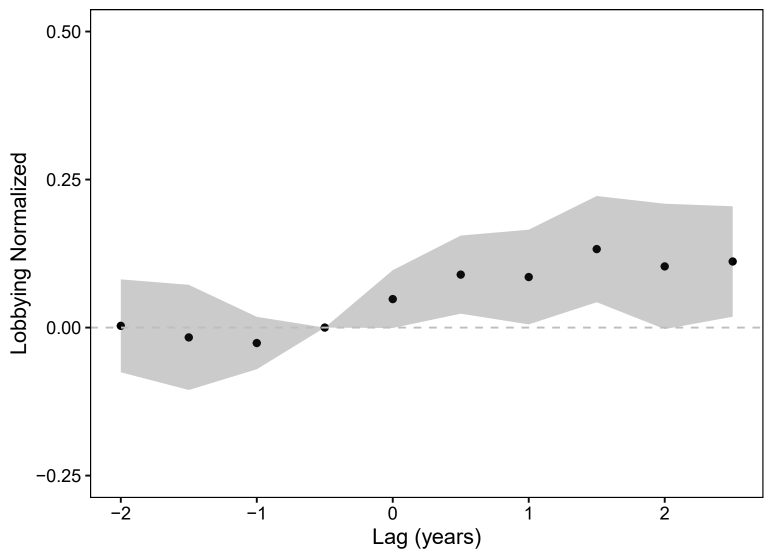


Figure 3: **The cumulative effect of the discovery of a previously unknown externality on firm lobbying investments.** 95% confidence intervals are robust to arbitrary correlations in innovation investments within firm over time, and capture the full intertemporal covariance structure of the estimated dynamic effect. Time 0 is the half-year of publication of a discovery, and the estimator includes five subsequent periods (half-years) of lagged effects. Estimates correspond to column 4 of Table 5.

[37] That is, these estimates indicate that firms do not simply shift innovation 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 B.1.

Estimates multiplied by 100 for interpretability.

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

In contrast to the slower, six-year innovation response, firms increase lobbying over about a two year period in response to the scientific discovery of a previously unknown externality. In the preferred specification, firms increase lobbying investments by a significant ($p < 0.05$) and economically meaningful 11.16% of pre-period lobbying levels (column 4 of Table 5). This 2.5 year cumulative effect implies an average flow increase of 2.2% per period over baseline levels.

Much of the discussion of the first-differenced distributed lag estimator from the previous section applies here as well. I highlight a few key points here. First, because all specifications are estimated in first-differences, a firm fixed effect (in levels) is implicit in each set of estimates. However, 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 (see Figure B.2 for a comparison of estimates with and without firm fixed effects). A firm fixed effect is not necessary in the innovation response estimates because pre-trends, even over 10 leads, are already remarkably flat (Figure B.1, left panel). Second, 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. Third, as before, all estimates require balance of treated firms over their treatment period. 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 insignificant over most lags. As in the previous section, the column 1 identifying assumption requires that discovery rates by independent scientists not vary over time in ways that are correlated with business cycles or changes in governmental regulatory regimes, and that scientists not begin to actively target for study chemicals in certain sectors of the economy more than others over time. 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 over time, then joint sector-half-year fixed effects to allow for sector-specific temporal shocks in discovery rates, and finally adding the firm fixed effect as described in the previous paragraph. 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. Figure D.1 displays the cumulative effects plots corresponding to each of the specifications in columns 1 – 5, which may aid visual comparisons.

As in the previous section, a final estimate pushes the data harder with a wide bandwidth of ten leads and lags (Table 5, column 5; Figure B.1, right panel), still including preferred firm and sector-year fixed effects. Estimating this wide bandwidth sacrifices power as before. Further, though, the CRP panel of lobbying data only starts in 1998, and is shorter than the Compustat panel of innovation data. Examination of these wide-bandwidth estimates yields a similar set of important findings as the innovation wide-bandwidth estimates. First, while 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. Second, the cumulative lobbying effect remains marginally significant over this extended window ($p < 0.1$) with non-decreasing point estimates, implying the estimated effects are permanent, rather than transitory, increases in firm lobbying. Third, Figure B.1 (right panel) indicates that 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, though not statistically different from it. This point of stabilization of the 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. As in the previous section, Column 5 of Table 5 omits several lead and lag estimates for brevity (Figure B.1, right panel, displays all cumulative effect estimates).

7.3 Additional robustness tests

The main estimates of Figures 2 and 3 are robust to multiple variants of the main specification. Tables 4 and 5 display robustness to 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 of the results to alternative bandwidths over which leads and lags are estimated are given in Appendix Figures C.2 (innovation 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 (innovation response) and D.3 (lobbying response). Finally, robustness tests to normalizing control firm outcomes to an early panel year instead of the panel average[38] are given in Appendix Figures C.4 (innovation response) and D.4 (lobbying response).

7.4 Evidence of firm substitution between innovation and lobbying

Having estimated the magnitude of firm first-order innovation and lobbying responses to a shock over the possibility of future regulation, I now turn to the question of whether innovation and lobbying are complements or substitutes in firm actions.

Two independent tests both provide suggestive evidence that innovation and lobbying are on average substitutes in the actions of treated firms in the data. One test suggests that the option to lobby induces less innovation from firms than they otherwise might choose (Figure 4; Table 6, left panel), and another 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 innovation responses (Figure 6). However, the test for innovation spillovers in firm lobbying is inconclusive (Figure E.1).

7.4.1 Innovation response heterogeneity over preexisting lobbying capacity

On average for treated firms, results indicate that firms operating across industries with lower pre-period lobbying tend to innovate more in response to the threat of future regulation (Figure 4; Table 6, left panel), which is consistent with innovation and lobbying as substitutes in firm actions. 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.

[38]In the main specification, control firm innovation is normalized to the firm's panel average market capitalization (and for lobbying control firms, panel average lobbying), which is done because control firms do not have any well defined "baseline" period (Section 4.3). 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.

Yet, for every post-period in the cumulative effects estimate, firms operating in industries with below-median pre-period lobbying exhibit a larger innovation response, and firms in industries with above-median pre-period lobbying exhibit a diminished response.

Firms in high-lobbying industries ("High lobby" in the left panel of Table 6, red in Figure 4) do eventually have a marginally significant cumulative increase of 0.71 percentage points in normalized innovation ($p < 0.1$) six years after a scientific discovery. However, this high-lobby firm response is a statistical and economic zero for the year of a discovery and four years afterwards. In contrast, firms in low-lobbying sectors ("Low lobby" in the left panel of Table 6; blue in Figure 4) have an innovation 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 which reaches an economically important 2.29 percentage point increase in normalized innovation in the six years following a discovery – triple the high-lobby firm response.[39]

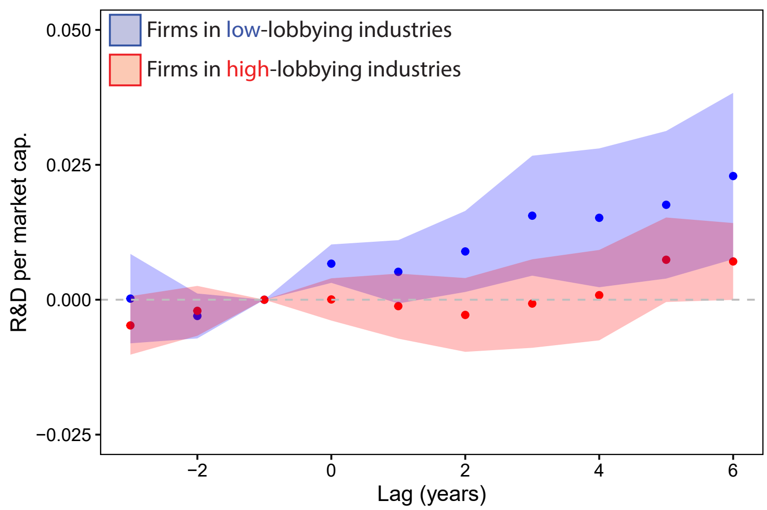


Figure 4: **Heterogeneity over pre-period lobbying in the cumulative effect of the discovery of a previously unknown externality on firm innovation investments.** Firms in industries that had been lobbying more in the pre-period tend to innovate less in the post-period, an important source of heterogeneity in the main innovation response (Figure 2). This is consistent with innovation and lobbying as substitutes in firm actions. Note that pre-period lobbying 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.

[39]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.

7.4.2 Lobbying response heterogeneity over preexisting innovation capacity

Estimates of heterogeneity in the lobbying response exhibit much more noise than those for the innovation response, limiting the ability to draw strong conclusions. As with the innovation heterogeneity estimates, the interaction effect estimated here is highly flexible: each lead and lag indicator has a separate interaction with pre-period innovation (Eq. 6), and nothing constrains the cumulative effects to be consistently above or below each other in every period. Point estimates suggest firms in industries with low pre-period innovation 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-innovation firms is double that of high-innovation firms (20.48% significant increase versus 9.46% insignificant increase). However, estimated magnitudes for the two categories of firms are generally more similar for other lags (excepting lag 3). After lag 1, low-innovation estimates are generally significant while high-innovation 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.

7.4.3 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 3), and indeed industry trade groups in the U.S. facilitate lobbying coordination (Section 4.3; (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.[40]

The left panel of Figure 6 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

[40] 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).

	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 0.78 percent of market capitalization. In contrast, "low lobby" firms increased innovation by a highly significant 1.91 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., three half-years after a discovery, "high R&D" firms increased lobbying by an insignificant 7.33 percentage points. In contrast, "low R&D" firms increased lobbying by a highly significant 18.96 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.
$^{***}p < 0.01, ^{**}p < 0.05, ^{*}p < 0.1$

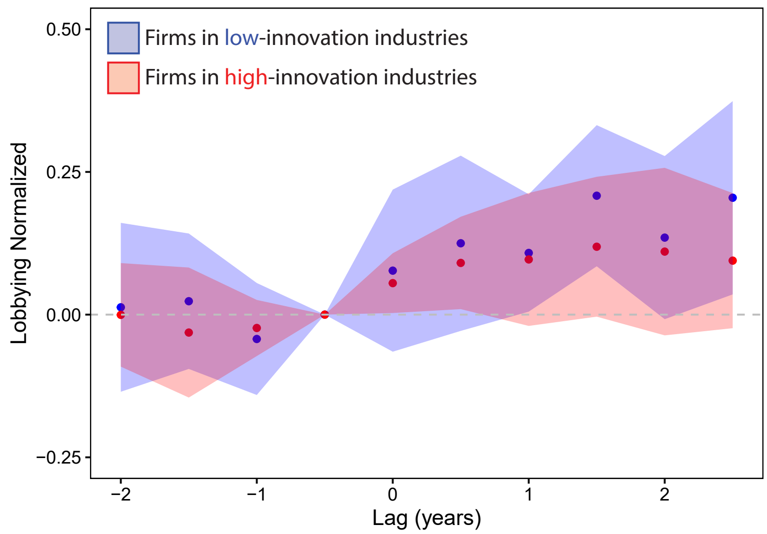
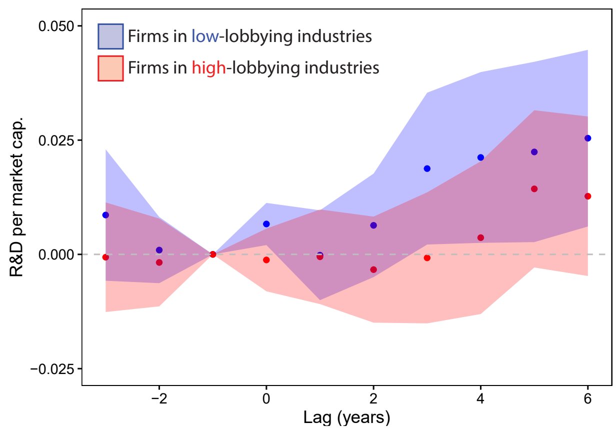


Figure 5: **Heterogeneity over pre-period innovation in the cumulative effect of the discovery of a previously unknown externality on firm lobbying investments.** It may be that firms that had been innovating more in the pre-period tend to lobby less in the post-period, though estimates are too noisy to draw clear conclusions. Pre-period innovation 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.

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 6, 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 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.



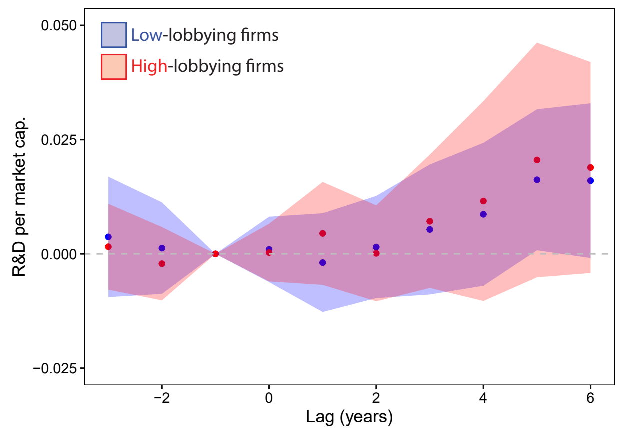


Figure 6: **Lobbying spillovers to firm innovation.** Left: Analogous to Figure 4, with the pre-period lobbying defined as the average lobbying within a 4-digit SIC code, but for the firm-matched subset of the data. Right: Same as left, except pre-period lobbying is defined using a firm's own lobbying 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.

Appendix Figure E.1 presents an analogous test for spillovers from industry innovation to firm lobbying. Unfortunately, the limited firm-matched data lead to some concern regarding the stability of the estimated lobbying response, with large point estimates and standard errors. Taken at face value, the results in Figure E.1 indicate a possible role for innovation spillovers to firm lobbying, but these results are primarily included for completeness.

7.5 Market concentration as a factor in firm strategies

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 innovation 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."[41] The following regression examines heterogeneity over market concentration in firm innovation 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 innovation response is noisy, but results (Figure 7) suggest that firms operating in low-concentration markets with relatively less market power tend to innovate more in response to the regulatory shock than do firms in high concentration markets. In Figure 7, low-concentration firms cumulatively increase innovation by 2.3% ($p < 0.05$) of pre-period market capitalization, but high-concentration firms only increase innovation by 1.1% ($p < 0.05$) of pre-period market capitalization. The innovation 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).[42]

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

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

[42] Given that pre-period lobbying capacity was also found to be a driver of heterogeneity in the innovation response, one might be concerned about omitted variables in the interaction estimates. Testing for heterogeneity in innovation 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 innovation response heterogeneity over pre-period lobbying capacity. Results available on request.

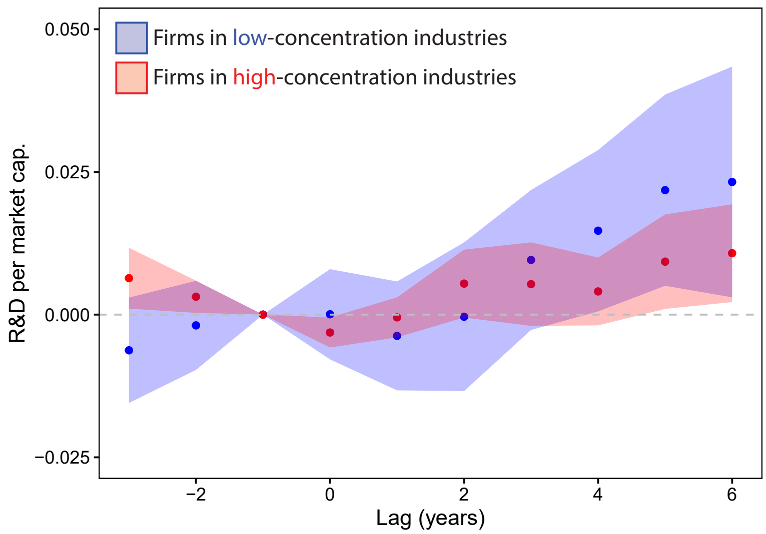


Figure 7: **Heterogeneity over market concentration (HHI) in the cumulative effect of the discovery of a previously unknown externality on firm innovation.** "High concentration" is an indicator for HHI values of 0.18 or greater (see text for details). Results here suggest that firms operating in low-concentration 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.

lobbying investments following an information shock (Figure E.2). This includes both overlap in 95% confidence intervals as well as the magnitudes (and ordering) of the point estimates.

8 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. On the other hand, if the regulatory instrument considered does not rely on private information (e.g. a Pigouvian tax[43]), then lobbying by firms may be nothing more than an effort to protect or appropriate rents. However, even in

[43]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.

the case of Pigouvian taxation, 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 information regarding supply in the regulated and unregulated economies 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 firms' lobbying expenditures.

Second, the heterogeneity in firm innovation responses to the threat of regulation that I estimate is over cross sectional variation in sector average lobbying (Equation 5). While the information shocks firms face may be plausibly exogenous, cross sectional variation in sectoral average lobbying is not. Thus, it may be the case in Figure 4 that firms with stronger innovation responses simply happen to be in industry sectors where lobbying is less prevalent, and vice versa. 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. In doing so, I set aside any direct welfare implications of lobbying on optimal regulatory design. I estimate an average reduction in firms' six-year cumulative innovation responses from 2.3% to 0.7% of firm market capitalization for firms in industries with below median versus above median pre-period lobbying (Table 6, left panel). Taking these estimates at face value, going from below median to above median lobbying is associated with a 1.5 percentage point reduction in normalized innovation. The total pre-period market capitalization of treated firms in sample is \$2.4 trillion, which implies a reduction in innovation 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 innovation estimated here is about \$144 billion (\$36 billion $\times$ 4).

9 Conclusions

This paper introduces the idea that firms with the option to lobby over future regulation affect innovation incentives. Faced with a shock to the probability of future regulation, forward-looking firms have an incentive to innovate (should regulation come to pass), but they also have the option to lobby the regulator. To empirically estimate firm behavior, this paper further introduces novel data on shocks to regulatory expectations, enabling estimation of the dynamics of firm responses to original information shocks.

Empirically, anticipation effects of forward-looking firms make unbiased estimation of their response to regulation difficult, as it is hard to isolate original news about the possibility of future regulation. I overcome this problem by identifying shocks to scientific knowledge from discoveries by independent researchers of previously unknown chemical harms. In particular, I scrape the epidemiological literature for publications of harms from commonly used industrial chemicals, and review over 7,000 publications to construct a novel dataset of scientific discoveries of previously-unknown harms from these chemicals. These discoveries – almost entirely by independent (non-firm) scientists – serve as original shocks to firm information sets over the future regulatory state, which I leverage for estimation.

Estimating the dynamics of firm responses to a discovery, I find that firms slowly increase overall innovation investments by a total of 1.8% of their market capitalization over a six year period, but they quickly increase lobbying investments by a total of 11% of pre-period lobbying levels over a roughly two year period. The average treatment effect (ATE) estimated here can be interpreted as the expected effect of an epidemiological discovery of a low-dose chemical harm, before it is known which particular industry that discovery will affect. This ATE is arguably the policy-relevant parameter regarding similar epidemiological discoveries in the future.

Because total expenditures serve as outcomes, these estimates are net of any within-firm substitution across innovation or lobbying activities. Importantly, while regulatory-induced increases to firm *clean* innovation are well-documented (Popp, 2019), the possibility of a net decrease in *total* innovation – driven by substitution out of other innovation activities – has been difficult to exclude (Dechezleprêtre and Sato, 2017). To my knowledge, the firm innovation response estimated here represents the first firm-level documentation of an aggregate innovation expenditure response to regulation.

Empirical tests for cross-sectional heterogeneity in firm responses suggests that innovation and lobbying are substitutes in firm actions, on average for treated firms in-sample. That is, firms with low lobbying capacities demonstrate a larger innovation response, while those with high lobbying capacities exhibit reduced innovation. A symmetric test for heterogeneity in the lobbying response over innovation capacity is consistent with this finding, but exhibits more noise. Additional tests on a matched subset of the data suggest that lobbying spillovers may drive the estimated heterogeneity in firm innovation responses. This finding is consistent with the known role of trade groups in coordinating lobbying activities among members. While not the main focus of this analysis, this finding suggests a new role through which rival actions may influence firm innovation.

Drawing on a conservative value of innovation from the literature, a back-of-the-envelope calculation suggests that the option to lobby may induce in-sample firms on average to reduce innovation investments by $36 billion, with a forgone social value of about $144 billion.

This analysis faces limitations which present avenues for future work. While this study tests the sign of first order innovation and lobbying responses, the data do not permit me to test whether firms correctly optimize on the margin. Internal frictions (e.g. managerial competition) may prevent firms from fully internalizing their own innovation and lobbying spillovers. Additionally, I test for aggregate innovation and lobbying investments in response to the threat of future regulation, but insight regarding how firms innovate and lobby could be important and may be facilitated by systematic textual analysis of firm patents and long-form lobbying reports. Firm lobbying reports are not uniformly formatted and thus difficult to work with, but would be useful to pursue in future work. Finally, the rapid lobbying response relative to the slower innovation response estimated here could suggest that firms employ lobbying as a means to resolve regulatory uncertainty before investing in innovation. Exploring lobbying as a modulator of firms' option value to the timing of innovation investments (Dixit and Pindyck, 1994) represents another interesting extension.

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, 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.

Akcigit, Ufuk, John Grigsby, Tom Nicholas, and Stefanie Stantcheva. 2022. "Taxation and innovation in the twentieth century." *The Quarterly Journal of Economics* 137 (1):329–385.

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.

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, Nicholas, John Van Reenen, and Heidi Williams. 2019. "A toolkit of policies to promote innovation." *Journal of Economic Perspectives* 33 (3):163–184.

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.

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 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.

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* .

Dixit, Avinash K and Robert S Pindyck. 1994. *Investment under uncertainty.* Princeton University Press.

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

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.

Grossman, Gene M and Elhanan Helpman. 1994. "Protection for Sale." *The American Economic Review* :833–850.

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

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.

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.

Lenway, Stefanie, Randall Morck, and Bernard Yeung. 1996. "Rent seeking, protectionism and innovation in the American steel industry." *The Economic Journal* 106 (435):410–421.

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

Murphy, Kevin M, Andrei Shleifer, and Robert W Vishny. 1991. "The allocation of talent: Implications for growth." *The Quarterly Journal of Economics* 106 (2):503–530.

———. 1993. "Why is rent-seeking so costly to growth?" *The American Economic Review* 83 (2):409–414.

Newell, Richard G, Adam B Jaffe, and Robert N Stavins. 1999. "The induced innovation hypothesis and energy-saving technological change." *The Quarterly Journal of Economics* 114 (3):941–975.

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.

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.

Appendices

A Conceptual framework details

This appendix presents details of the conceptual framework summarized in Section 3. First order conditions are as given in Section 3:

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

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

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

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:

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

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

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

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

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

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 3. To obtain the signs discussed in Section 3, note that $\Delta\pi$ and η' are always the same sign, so that their product is always positive.

B Testing pre-period parallel trends

The main innovation and lobbying specifications include three periods and four periods of leads, respectively. Here I show results for 10 periods of leads and lags for both responses (Figure B.1). Note that lobbying is observed every six months so 10 periods is five years, and the lobbying panel (which starts in 1998) is much shorter than is the innovation panel (which starts in 1950). Estimated effects show more noise than the corresponding main estimates (Figures 2 and 3) for two reasons. First, more coefficients are estimated. Second, treated firms require more pre- and post-period data to enter the balanced panel, thus fewer treated firms enter estimation over this longer time horizon. See Tables 4 and 5 for counts of treated firms. These caveats aside, the point-estimates of overall cumulative effects are remarkably consistent with those estimated in the main models, as is the temporal structure of the responses.

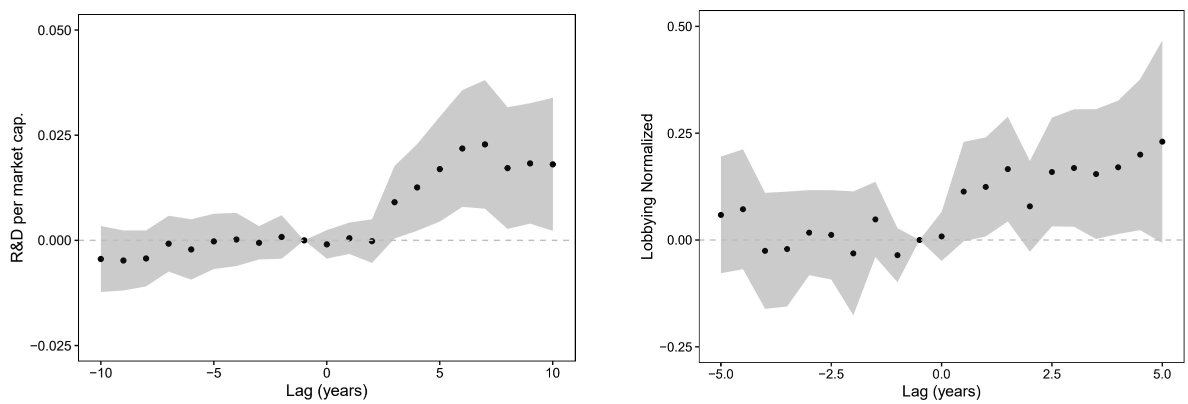


Figure B.1: The cumulative effect of the discovery of a previously unknown potential externality on the innovation (right) and lobbying (left) investments of affected firms. These estimates include an extended set of leads and lags in firm responses, increasing the noise in the estimated effects considerably. However, additional leads serve as an even stronger partial test for differential trends, showing that treated and control firms exhibit parallel trends for many periods before treatment occurs. The additional lags show that the point-estimates of the treatment effects stabilize within the shorter lag periods estimated in the preferred specifications (Figures 2 and 3), and they stabilize at similar values. Estimates here include 10 leads and lags for both responses. 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 B.2 presents the reasoning for including firm fixed effects in the main specification of the lobbying response. The blue plot in Figure B.2 displays lobbying estimates over 10 leads and lags (following Figure B.1, right panel), but *excludes* firm fixed effects from the first differences estimator. Pre-trending behavior is quite clear in these estimates and statistically significant in two pre-periods ($p < 0.05$). The grey plot in Figure B.2 includes firm fixed effects, which eliminate pre-trending behavior, and is included here for visual reference (it is identical to the right panel of in Figure B.1).

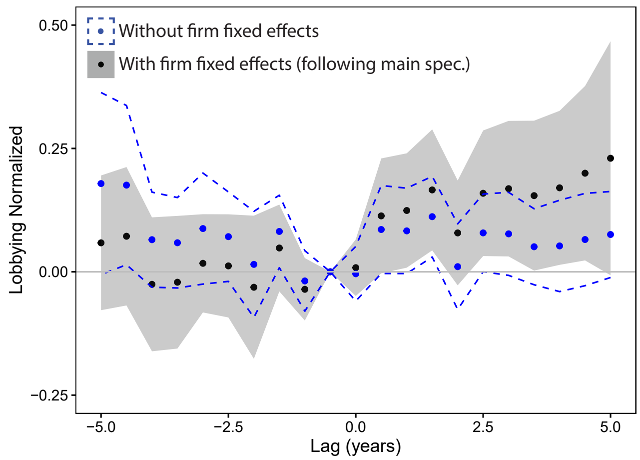
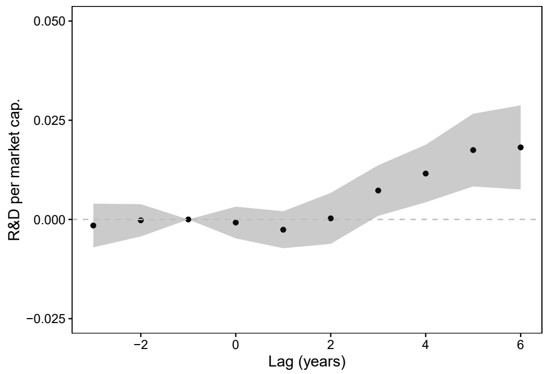


Figure B.2: **Inclusion of firm fixed effects eliminates pre-trends in the lobbying response.** Excluding firm fixed effects in the first differences lobbying estimator results in pre-trending behavior in the lobbying response over an extended set of 10 leads and lags (blue dashed plot). Estimates in grey follow the main specification and include firm fixed effects in the first-differences estimator, eliminating pre-trends. Note that estimates in grey replicate those in the right panel of Figure B.1 above, and are included here for reference. Confidence intervals are as in Figure B.1.

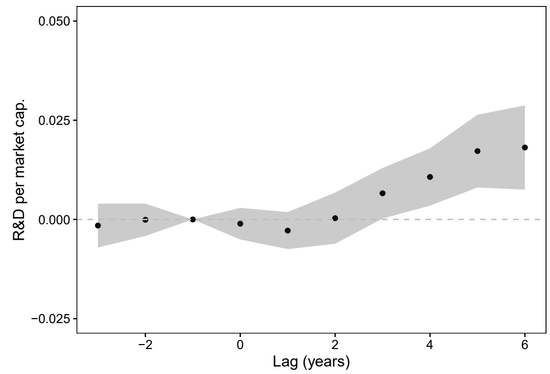
Note that firm fixed effects are not necessary in the innovation estimates, as pre-period behavior is already remarkably flat, even when estimating over 10 leads and lags (Figure B.1, left panel).

C Robustness of the innovation response

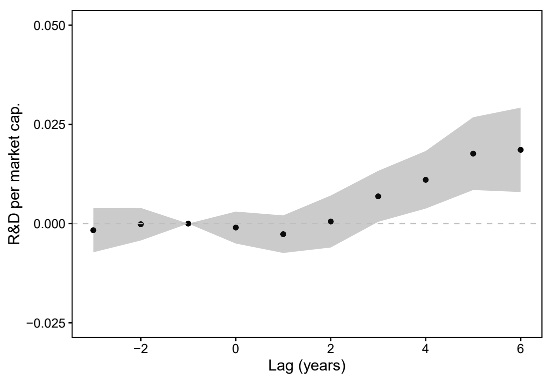
This section demonstrates robustness of the innovation response to alternative fixed effects specifications (Figure C.1) and lag specifications (Figure C.2), and to dropping individual discoveries (Figure C.3, only the most influential dropped discoveries shown).



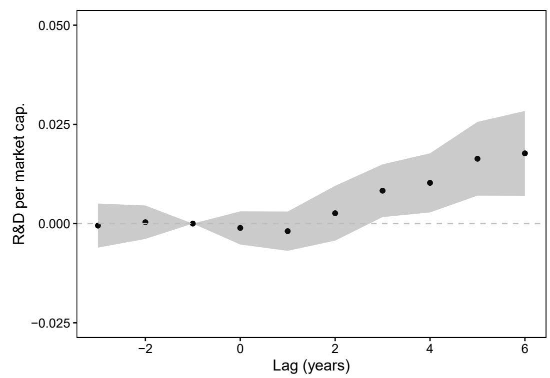
(a) No fixed effects



(b) Year

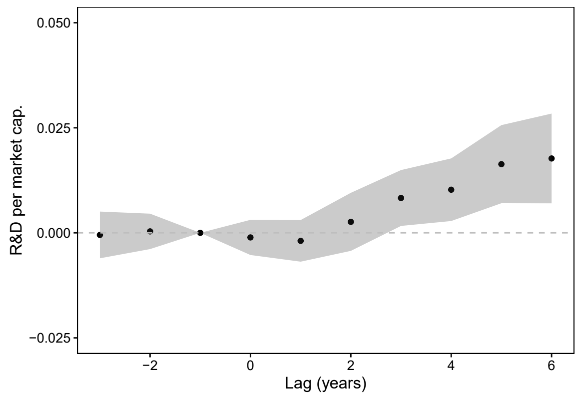


(c) Sector + year

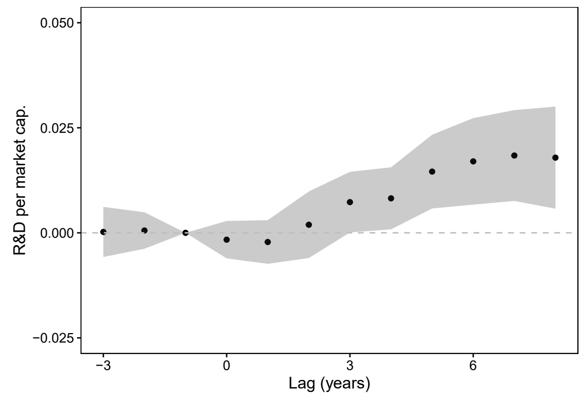


(d) Sector×year

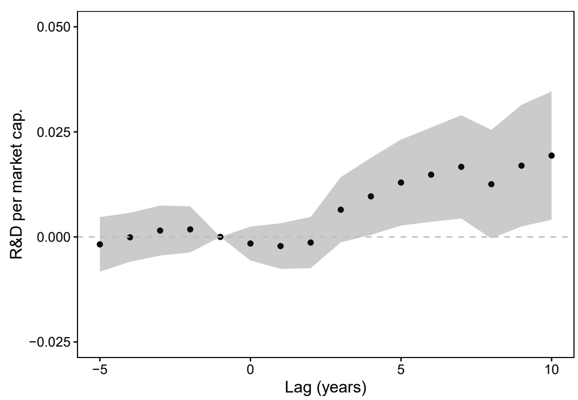
Figure C.1: Robustness of the innovation response to alternative fixed effects specifications. Panel (d) is the preferred specification presented in the main text, included here for comparison. 95% confidence intervals are robust to arbitrary correlations in innovation investments within firm over time,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



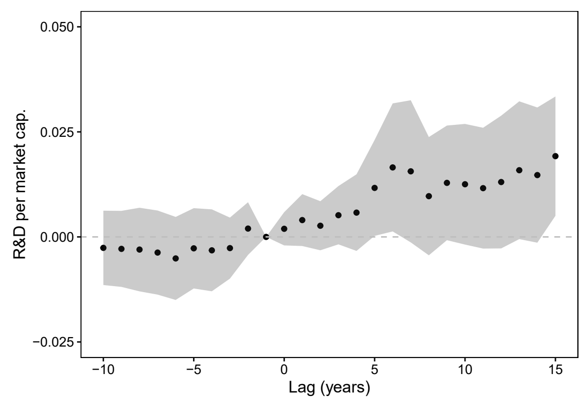
(a) 3 leads, 6 lags



(b) 3 leads, 8 lags



(c) 5 leads, 10 lags



(d) 10 leads, 15 lags

Figure C.2: Robustness of the innovation 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.

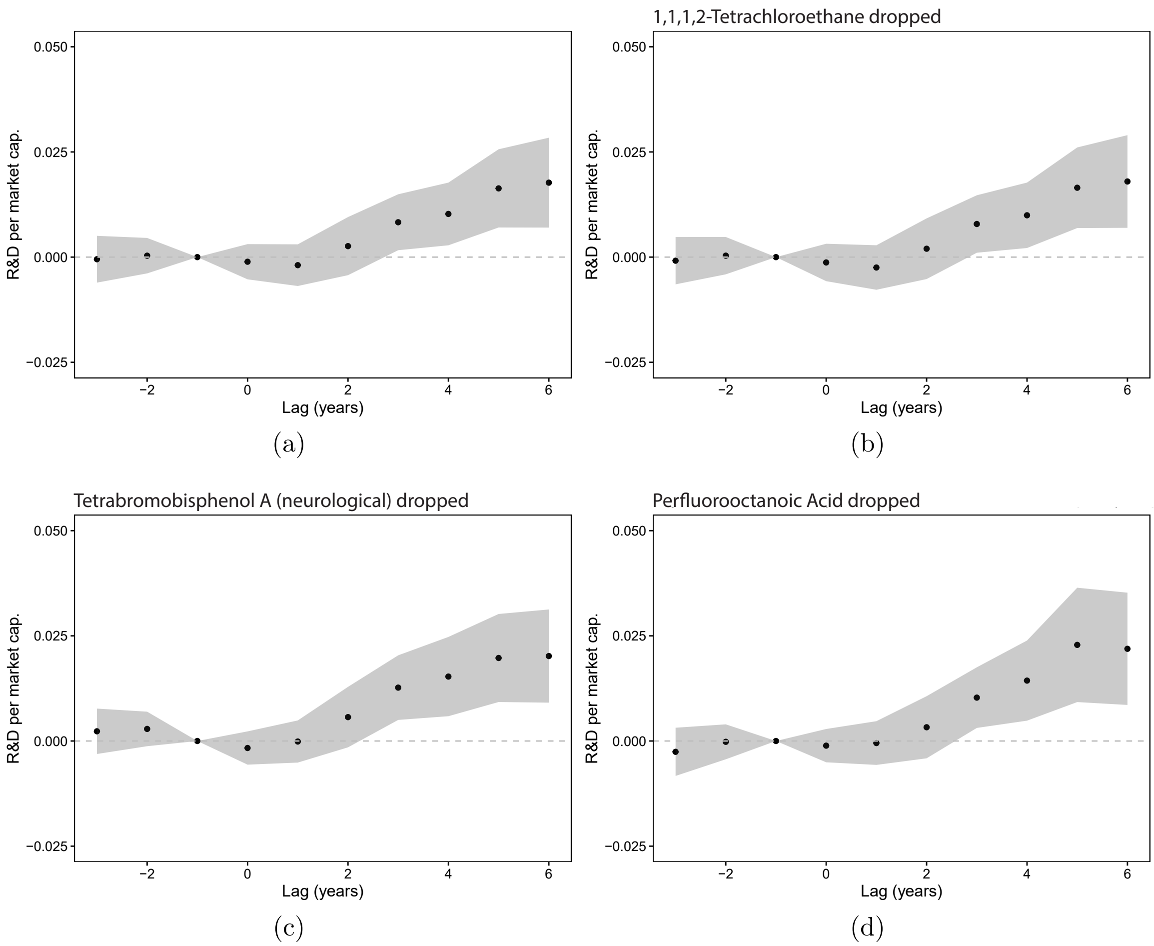


Figure C.3: Robustness of the innovation response to dropping individual discoveries. Panel (a) is the preferred specification presented in the main text, included here for comparison. Panel (b) is representative of dropping a random discovery. Panels (c) and (d) are the two discoveries that *most* affect 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.

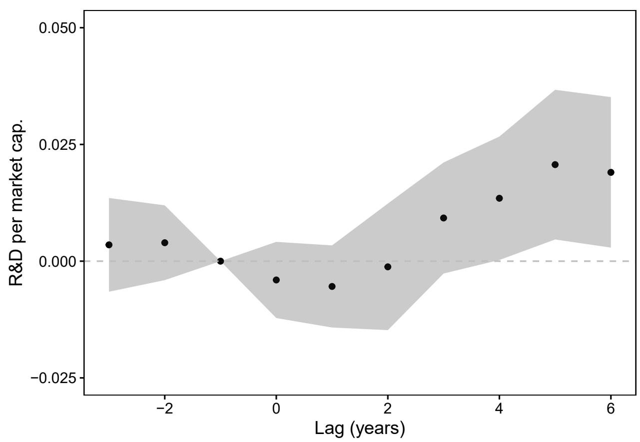
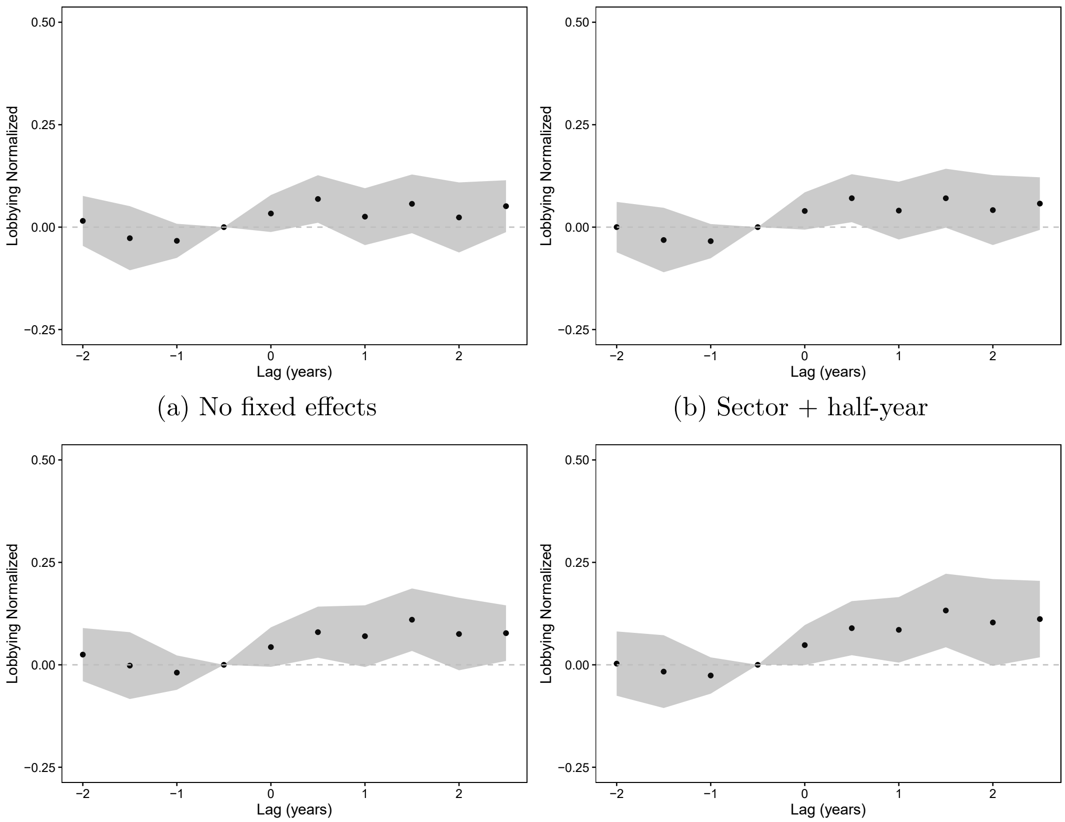


Figure C.4: Robustness of the innovation response to normalizing controls at a common year. Estimates follow the main specification, but normalize control firm innovation expenditures to market capitalization centered on a single common year (1980). In the main specification, control firm innovation 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).



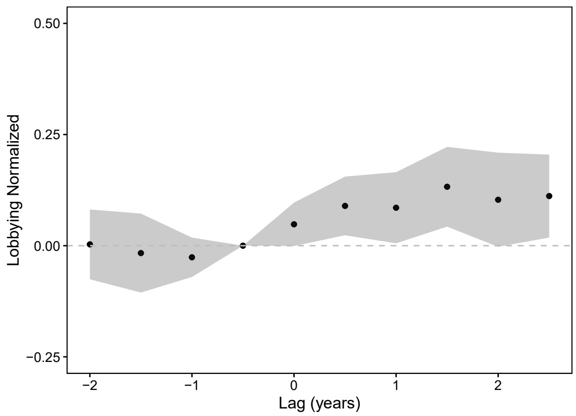
(a) No fixed effects

(b) Sector + half-year

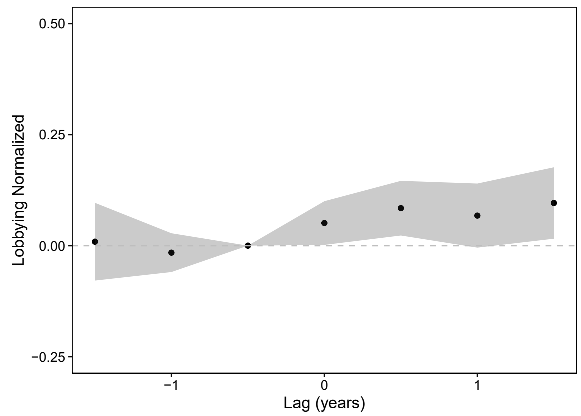
(c) Sector×half-year

(d) Firm + Sector×half-year

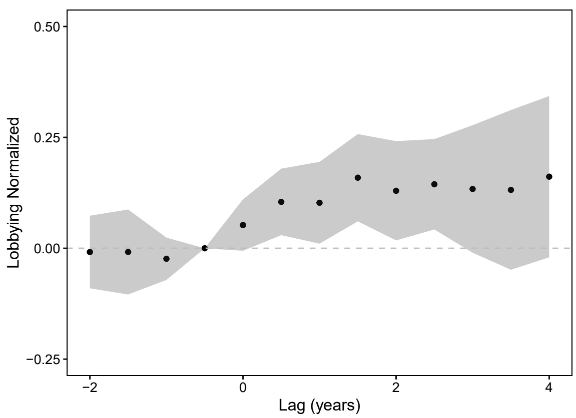
Figure D.1: Robustness of the lobbying response to alternative fixed effects specifications. Panel (d) is the preferred specification presented in the main text, included here for comparison. 95% confidence intervals are robust to arbitrary correlations in lobbying investments within firm over time,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



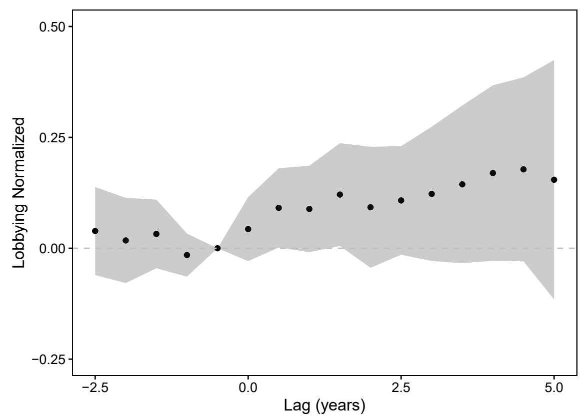
(a) 4 leads, 5 lags



(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.

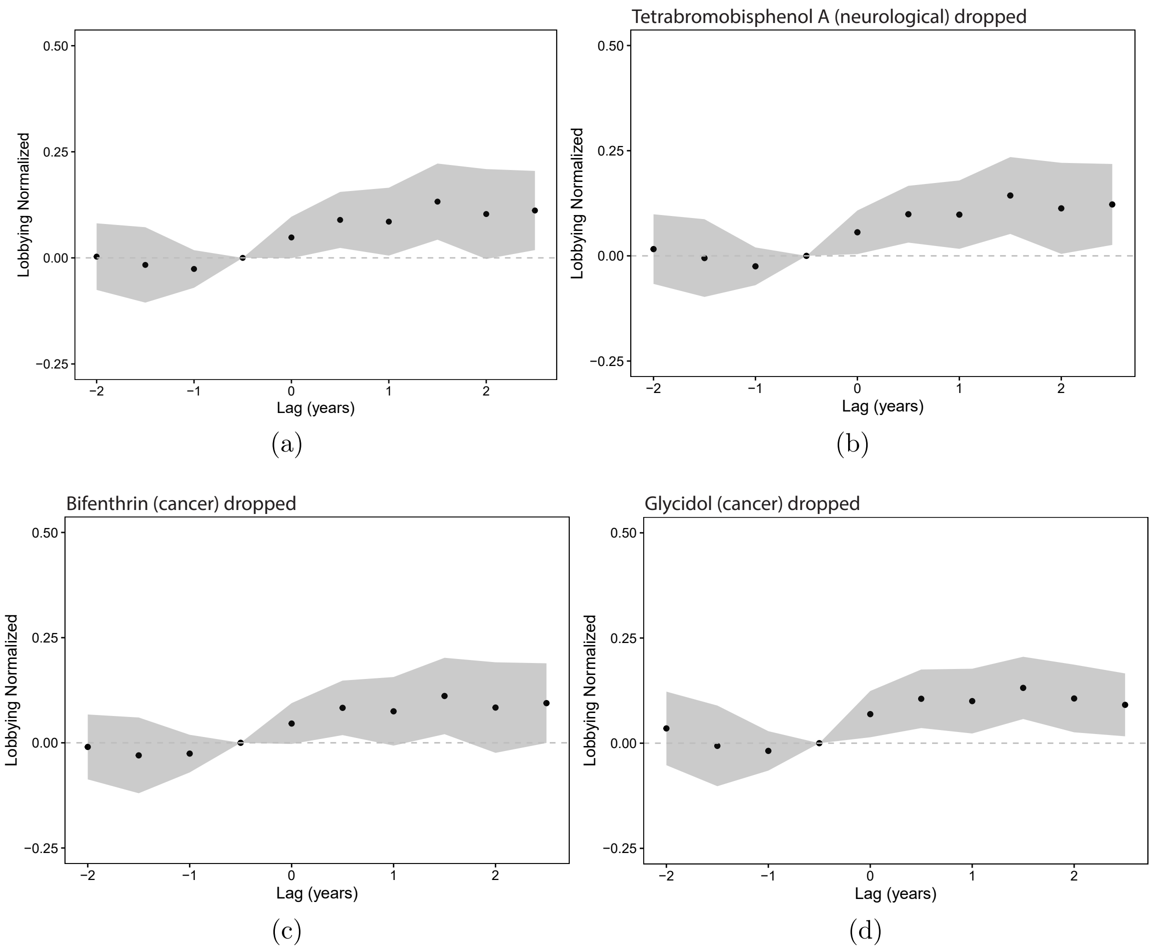


Figure D.3: Robustness of the lobbying response to dropping individual discoveries. Panel (a) is the preferred specification presented in the main text, included here for comparison. Panel (b) is representative of dropping a random discovery. Panels (c) and (d) are the two discoveries that *most* affect 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.

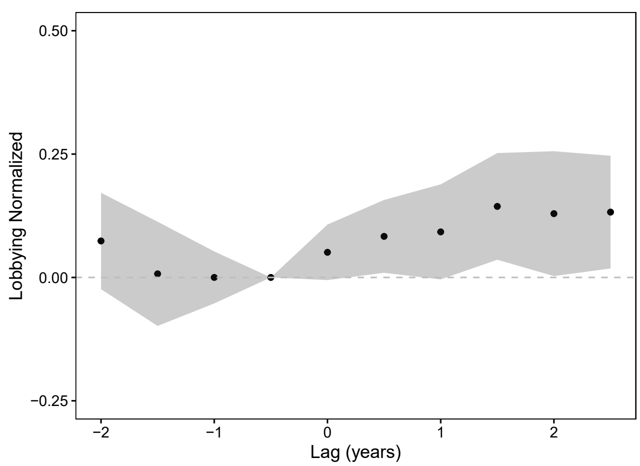


Figure D.4: Robustness of the lobbying response to normalizing controls at a common year. Estimates follow the main specification, but normalize control firm lobbying expenditures to a baseline centered on a single common year (2000). In the main specification, control firm lobbying is normalized to 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 Additional heterogeneity in the lobbying response

This appendix presents additional tests for heterogeneity in firm lobbying responses to a regulatory threat. These results are generally too noisy to lend them a high level of confidence, and so should be interpreted with caution.

E.1 Tests for innovation 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 7.4.3. However, results here were too unstable for inclusion in the main text.

As described in the main text, results 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 7.4.3). The left and right panels of Figure E.1 are directly analogous to the spillover test in Figure 6. That is, the left panel displays an estimate of main specification for firm lobbying heterogeneity over pre-period sectoral innovation (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.1, 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.1 is expanded relative to those in the main text to accommodate uncertainty in these estimates.

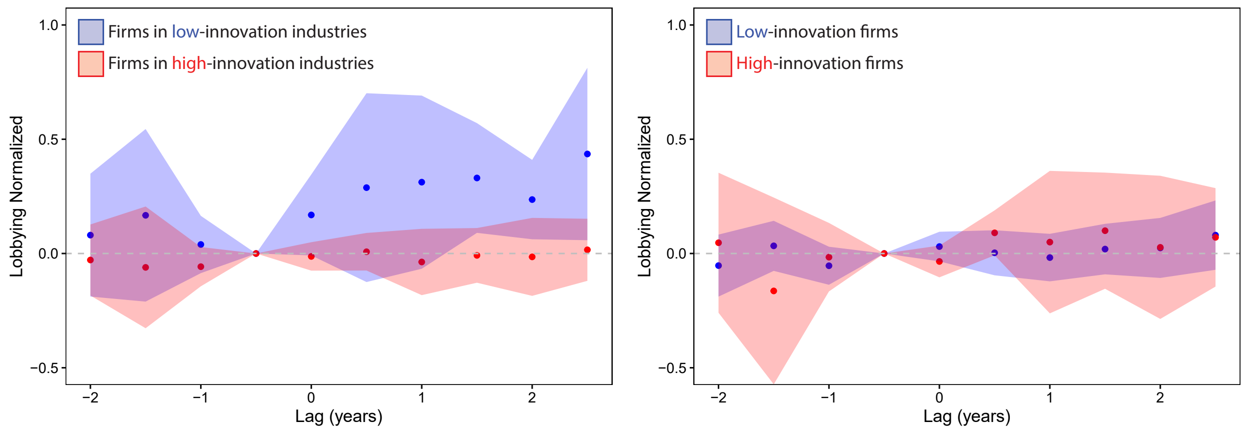


Figure E.1: Innovation spillovers to rival firm lobbying are noisy. Left: The main specification for firm lobbying heterogeneity over pre-period sectoral innovation (as displayed in Figure 5 with modified y-axis), but using a more limited dataset composed of individual firm matches across the Compustat and Center for Responsive Politics datasets. See Section 7.4.3 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. 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.

E.2 Heterogeneity over market concentration in the lobbying response

This section displays heterogeneity over market concentration (industry sector HHI) in firm lobbying responses to a regulatory threat. This estimate is analogous to the innovation response estimate of Figure 7, 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.

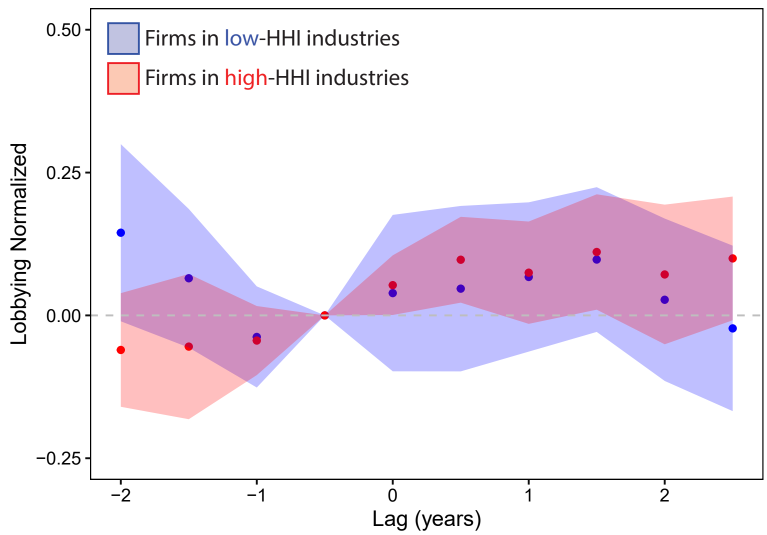


Figure E.2: Heterogeneity over market concentration (HHI) in the cumulative effect of the discovery of a previously unknown potential externality on the lobbying investments of affected firms. "High-HHI" defined as in Figure 7. Little meaningful heterogeneity in the lobbying response over HHI is resolved. 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.

F Time series variation in the data

Time series variation in innovation and lobbying In the time series of firm investments, both innovation (Figure F.2a) and lobbying (Figure F.2d) are increasing in real terms for both treated and untreated firms. Additionally, levels of innovation 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 in outcome 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 6).

As discussed in the main text, to account for scale effects, I normalize innovation investments by the firm's market capitalization (Figure F.2b) and lobbying investments by the firm's average lobbying expenditure[44] (Figure F.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 F.2b and F.2e).

Because innovation expenditures often exhibit a unit root (which I also cannot reject, see Section 6), 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 F.2c and F.2f). Of course, it may be the case (for example) that some firms are trending upward in their innovation or lobbying investments while others are trending downward, and these groups of firms are differentially affected by scientific discoveries over time. These differential trends can still be present even if cross-sectional and temporal differences in aggregate are not. I address these concerns in my empirical specification.

One additional detail in the data is worth pointing out. The first-differenced, normalized innovation data (Figure F.2c) exhibit a large positive spike for both treated and untreated firms

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

in 2008 and a subsequent negative spike in 2009. This appears to be due to the Great Recession, which began in late 2008. Firm innovation 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 year innovation expenditures fell and firm market values partially recovered, leading to low levels of normalized innovation and a large (negative) first difference. Interannual shocks such as these are controlled for by sector-year fixed effects in my empirical specification.

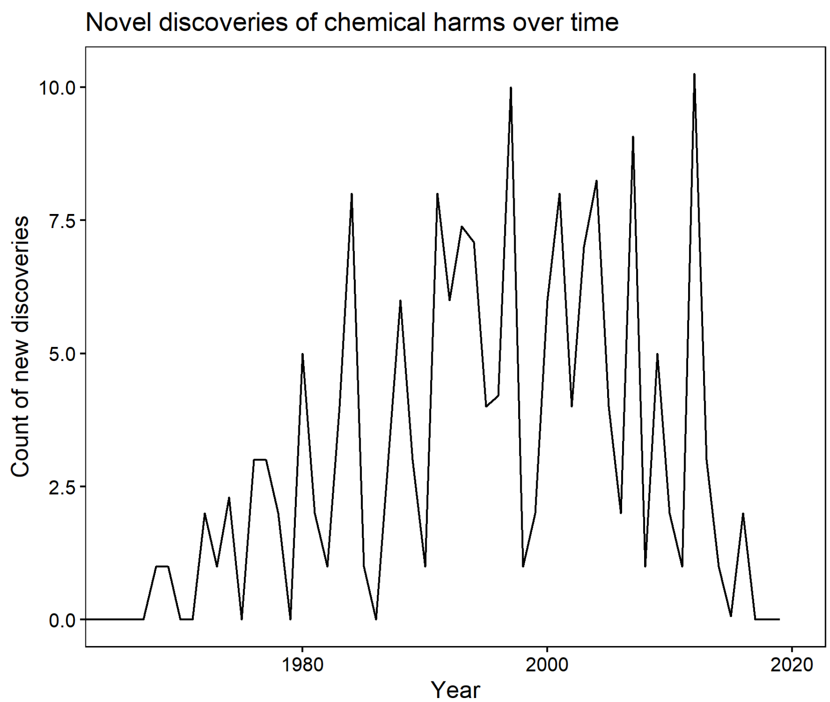


Figure F.1: **Time series of the count of discoveries of previously-unknown chemical harms by year.** Discoveries of harms are by independent epidemiological researchers, not R&D scientists employed at the sample firms. While the rate of new discoveries of chemical harms increases over time, discoveries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

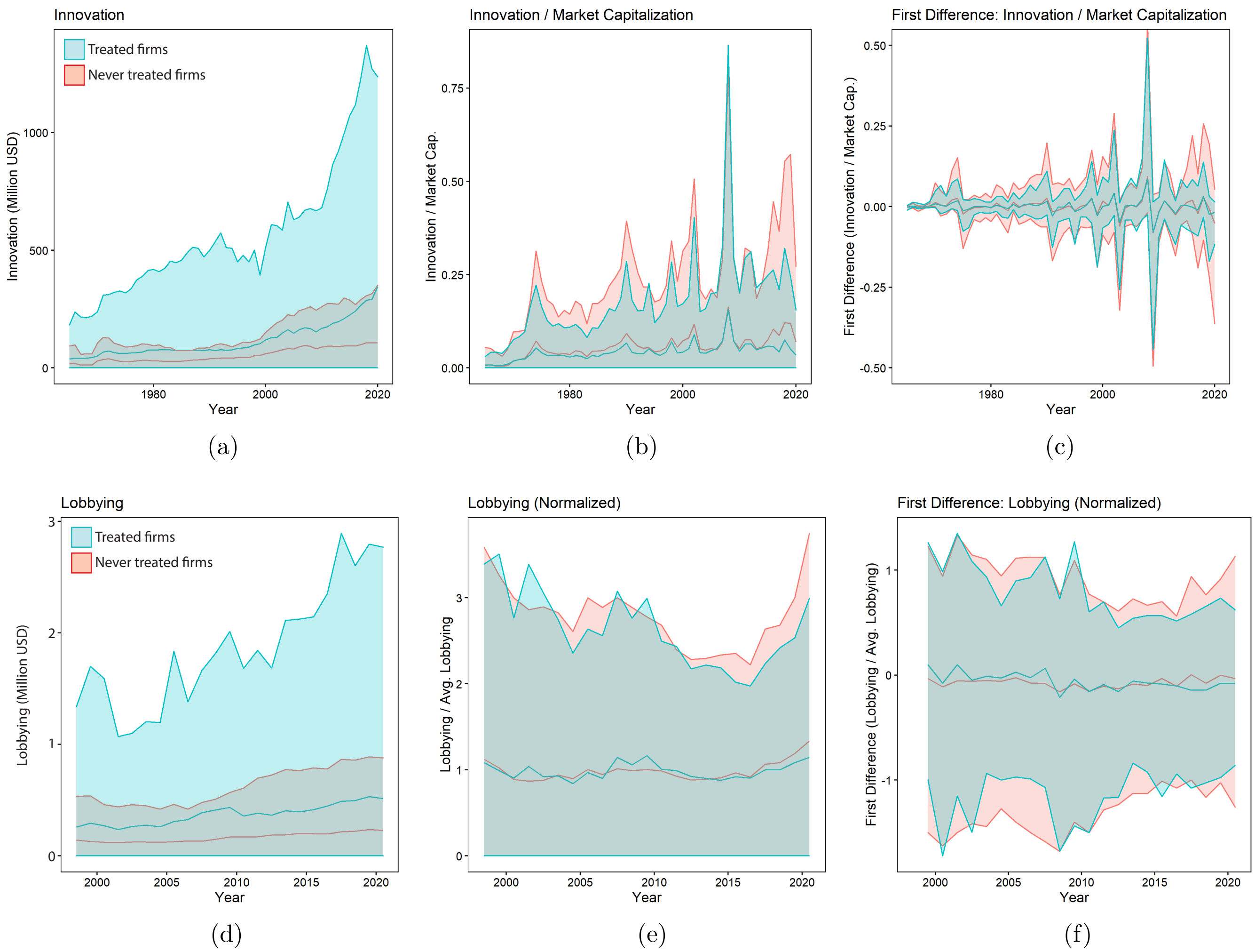


Figure F.2: **Time series variation in innovation and lobbying.** Time series variation in innovation (a-c) and lobbying (d-f) across treated firms (blue) and untreated firms (red). Innovation and lobbying expenditures are shown in real terms. Levels of firm innovation and lobbying investments exhibit 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.

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.