NBER WORKING PAPER SERIES

LEARNING BY VIEWING? SOCIAL LEARNING, REGULATORY DISCLOSURE, AND FIRM PRODUCTIVITY IN SHALE GAS

T. Robert Fetter
Andrew L. Steck
Christopher Timmins
Douglas Wrenn

Working Paper 25401
http://www.nber.org/papers/w25401

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2018

We thank DrillingInfo, the Pennsylvania Department of Environmental Protection (DEP), and the Pennsylvania Department of Conservation and Natural Resources (DCNR) for providing access to data, and appreciate insights provided by Lynn Levino (DCNR) and Jeff Brown (DEP) as well as Rob Jacobs, Bob Kleinberg, and Vikram Rao. We are grateful for helpful comments and suggestions from participants in numerous forums including at the National Bureau of Economic Research, Association of Environmental and Resource Economics, and the Norwegian School of Economics, and members of the Duke Public Economics Lab, Duke Environmental Policy seminar, and Duke Energy Initiative Fellows. Fetter acknowledges funding from the Yale Center for Environmental Law and Policy, Duke Environmental Economics Doctoral Scholars, and Resources for the Future; Steck acknowledges funding from the Duke University Graduate School and the Ottis Green Foundation. Timmins acknowledges support from the National Science Foundation (SES-1559481). Any errors or omissions are our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Learning by Viewing? Social Learning, Regulatory Disclosure, and Firm Productivity in Shale Gas

T. Robert Fetter, Andrew L. Steck, Christopher Timmins, and Douglas Wrenn

NBER Working Paper No. 25401
December 2018
JEL No. L24,L51,L71,O3,Q35,Q53,Q55,Q58

ABSTRACT

In many industries firms can learn about new technologies from other adopters; mandatory disclosure regulations represent an understudied channel for this type of social learning. We study an environmentally-focused law in the shale gas industry to examine firms' claims that disclosure requirements expose valuable trade secrets. Our research design takes advantage of a unique regulatory history that allows us to see complete information on chemical inputs prior to disclosure, along with the timing of information availability for thousands of wells after disclosure takes effect. We find that firms' chemical choices following disclosure converge in a manner consistent with inter-firm imitation and that this leads to more productive wells for firms that carefully choose whom to copy — but also a decline in innovation among the most productive firms, whose innovations are those most often copied by other firms. Our results suggest there is a long-run welfare trade-off between the potential benefits of information diffusion and transparency, and the potential costs of reduced innovation.

T. Robert Fetter
Nicholas Institute
Box 90335
2117 Campus Drive
Durham, NC 27708
rob.fetter@duke.edu

Andrew L. Steck
University of Toronto
Rotman School of Management
105 St. George Street Toronto,
ON M5S1J5
Canada
andrew.steck@utoronto.ca

Christopher Timmins
Department of Economics
Duke University
209 Social Sciences Building
P.O. Box 90097
Durham, NC 27708-0097
and NBER
christopher.timmins@duke.edu

Douglas Wrenn
Penn State
Department of Agricultural Economics, Sociology, and Education
112A Armsby Building
University Park, PA 16802
dhw121@psu.edu

1 Introduction

Mandatory disclosure of private information has become a commonly-used regulatory tool in varied settings and with varied aims. On the one hand, disclosure can be socially beneficial by reducing search costs or limiting negative externalities. On the other, firms assert that regulatory disclosures can amount to the publication of valuable trade secrets, dampening incentives to innovate. Policymakers enacting disclosure regulations must balance these two concerns, and the optimal policy depends on the answers to two questions. First, does the disclosed information in fact have business value, such that other firms respond to it? Second, if disclosure enables social learning spillovers that would not otherwise occur, does it limit firms' incentives to invest in costly innovation?

We study both questions in the context of hydraulic fracturing chemicals, and find affirmative evidence for each. Thus, while disclosure can provide benefits — either through the diffusion of best practices, or public health or environmental benefits — policy makers need to weigh those benefits against the long-run social costs from eroded incentives to innovate.

Hydraulic fracturing is an emerging technology that has transformed the nature of energy production in the U.S. and the world, and contributed to substantial economic growth in many areas (Hausman and Kellogg, 2015). It has also raised a number of concerns regarding local economic impacts and averting behavior (e.g. Muehlenbachs et al., 2015; Wrenn et al., 2016; Kirkpatrick and Fetter, 2018) and local environmental impacts, including the use of toxic chemical additives in the fracturing process (Elgin et al., 2012; Stringfellow et al., 2014; Fetter, 2018). Those concerns have led to policies requiring public disclosure of information about the chemicals used, starting in 2010 and now instituted in at least 18 US states.

Hydraulic fracturing is only one example of an industry where disclosure policies have been considered in order to address social welfare concerns. For example, the US Securities and Exchange Commission requires regular financial disclosures in order to protect the interests of public investors; these requirements balance the costs of bookkeeping, regular reporting, and potential revelation of valuable business information against the public's right to know and a lower risk of financial scandals. In an area of growing concern, governments are increasingly interested in regulating and mandating disclosures around technology com-

panies' collection and use of private data; e.g. the EU's recent General Data Protection Regulation, which requires companies to disclose how and when they are collecting user data, as well as providing more information on how that data is used and stored.[1]

In addition to these contexts, governments have used disclosure laws to require transparency over nutritional contents of food, restaurant hygiene, pollution releases, toxics use, renewable electricity generation, drinking water quality, and so on (e.g., Jin and Leslie, 2003; Fung et al., 2007; Bennear and Olmstead, 2008; Delmas et al., 2010; Kim and Lyon, 2011). Information-based policies seem to be especially popular in settings where risks are ill-understood yet not anticipated to be large. They often represent a "pragmatic compromise" (Fung et al., 2007) that is a politically viable response to emerging risks.

In all of these settings, regulators face a potential tradeoff between the discovery and diffusion of new technologies, where discovery is the innovative expansion of the technological frontier, usually as the result of costly trial-and-error, and diffusion is the spread of better technologies through and across industries. As pointed out in Mokyr (1992), both forces represent technological progress from a social point of view, and they are complementary in the long run: neither a high level of discovery nor a high level of diffusion is sufficient on its own for sustained broad societal progress.[2] In this paper we present empirical evidence of this tradeoff in the context of a disclosure law: the fracturing chemicals disclosure law we study appears to increase diffusion at the expense of discovery, highlighting the need for policymakers to weigh these competing forces when crafting disclosure laws.

Firms in the oil and gas industry, as in many sectors, rely on secrecy as a mechanism to capture value from their investments in innovation and maintain their competitive advantage (Wang and Krupnick, 2013; Cohen et al., 2000). Indeed, the industry has often met calls for mandatory disclosure regulations with the argument that fluid formulas represent trade

[1] Interestingly in the technology context, companies such as Facebook can release state-of-the-art tools and algorithms to the public, safe in the knowledge that the value of these algorithms can be unlocked only with their proprietary datasets. For example, see the tools available at: https://facebook.ai/developers/tools. Prufer and Schottmüller (2017) provide a model that illustrates the value of data in this sort of market, and the implications for competition policy.

[2] The patent system is one approach to balancing these concerns, where inventors are promised exclusive rights for a period of time to recoup discovery costs in return for publishing their inventions. Patents provide incomplete protection, however, especially in settings where innovations can be adjusted in small ways that do not substantially change their value but do qualify as novel and non-encroaching (Lemley and Shapiro, 2005).

secrets. In this paper, we shed light on whether innovations in fracturing fluids do in fact constitute a competitive advantage. In particular, we compare detailed, well-level information on inputs to test whether the chemicals used become more similar with public disclosure. We observe evidence that chemical mixes indeed converge following disclosure. We then test whether the use of chemicals more similar to those in high-performing firms' wells improves productivity, and find affirmative evidence. These findings suggest policy-makers need to weigh the tradeoff between right-to-know disclosure rules and industry secrecy. A further question is: to what extent does forced disclosure undermine the incentives for firms in this industry to invest in research and development for future innovations? We close the paper with evidence that among leading firms — which are most likely to conduct experiments, those whose experiments are the most likely to be copied by other firms, and thus (by revealed preference) most likely to engage in high-value experimentation — innovative activity declines following the advent of the disclosure rule.

Our analysis is primarily empirical, and it is set in Pennsylvania. Of the many states where firms are required to disclose hydraulic fracturing chemicals, this state provides a unique data environment in which to analyze our questions of interest. Pennsylvania was one of the earliest states to experience a dramatic increase in exploration and production of unconventional shale gas, and continues to experience extensive unconventional development. Furthermore, Pennsylvania's history features an unusual regulatory episode in which operators were required to disclose information about chemical additives in fluid to the regulator, but not in a format that was easily accessible for the general public (or to one another), for a 14-month period in the height of the boom. We have recovered this information through a combination of "Right-to-Know" law requests and other methods. Observing this information allows us to distinguish the effect of the public disclosure law from other simultaneous phenomena, including a general improvement in technology.

Our paper makes three primary contributions. Our first contribution is to the literature on the relationship between secrecy and innovation: We provide insight into the importance of the social tradeoffs policy makers face when considering disclosure laws. There may be a compelling social benefit from disclosing specific information, such as when production processes involve toxic chemicals in residential areas, but this social benefit must be weighed

against the ability of firms to realize economic returns from their innovations.[3] We provide evidence that innovations in hydraulic fracturing chemicals are valuable, and furthermore that innovation decreases following the disclosure rule — suggesting that this tradeoff may be important at least in this context.

Second, this paper expands the literature on the empirical effects of transparency or disclosure regulations in a new direction. Our analysis is the first, to our knowledge, to combine a study of the effects of information disclosure regulations with social learning. We find strong evidence that disclosure regulations about an emerging technology can enable social learning or spillovers that would not have occurred otherwise.[4] Prior papers in the industrial organization literature have examined the effect of price or quality disclosure on market structure and competition (e.g., Jin and Leslie, 2003; Bollinger et al., 2011; Luco, 2018), while other work has documented the effects of disclosure laws on environmental and health outcomes (e.g., Fung et al., 2007; Bennear and Olmstead, 2008; Delmas et al., 2010; Fetter, 2018) as well as investor behavior (e.g. Hamilton, 1995). Separately, other authors have studied the phenomenon of social learning (Conley and Udry, 2010), without considering the role of disclosure regulations specifically. Our research bridges these streams.

Third, we study the role that chemical additives have played in the development of hydraulic fracturing technology. Several recent papers, beginning with Covert (2015), and continuing with Steck (2018) and Agerton (2018), have documented the rise of hydraulic fracturing and the role that learning plays on different firm decisions. We are the first to seriously consider the role of chemicals in firm learning, and thereby contribute to the literature on hydraulic fracturing specifically and social learning about emerging technologies more generally.

The paper proceeds as follows. In Section 2, we provide context and background for our study, including a review of the emergence of the technologies that have made development of shale gas economically feasible, the environmental concerns that led to the requirement for public disclosure, and the literature on the role of alternative strategies — including secrecy

[3]Theoretically, patents solve the problem of imperfect appropriability, but in practice patents provide only "probabilistic" protection (Lemley and Shapiro, 2005), and many firms depend on secrecy and lead time (among other tactics) to realize returns to costly innovation (Cohen et al., 2000).

[4]This could also be interpreted as a form of technology diffusion facilitated by regulation, though in our context this is not a stated aim of the policy.

— that firms use to preserve competitive advantage. Section 3 describes our data. In Section 4 we present our analysis of how the disclosure rules affected chemical input choices, and how those choices related to well productivity. Section 5 examines innovative activity following the disclosure rule, and Section 6 offers concluding remarks.

2 Background

2.1 Engineering and Geology of Hydraulic Fracturing

The rise of shale gas to prominence in the US energy landscape has been well-documented.[5] The ability to profitably recover hydrocarbons from shale has been largely based on advances in four key areas of technology: horizontal drilling, three-dimensional seismic imaging, micro-seismic fracture mapping, and massive hydraulic fracturing (Wang and Krupnick, 2013). Elements of these technologies have been in development for several decades, spurred by both private and public investments in research and development. Technological advances since the 1970's have ranged from changes in the major compounds comprising fracturing fluid to greater control over directional drilling of wellbores. Foam was replaced by gels in the formulation of fracturing fluid, and order-of-magnitude changes were made in the quantity of proppant used. In the 1990's, there were important advances in the role of directional drilling; in combination with massive hydraulic fracturing, the ability to drill horizontally through a shale formation made shale gas development economical. At the same time, "slick water" fracturing fluid replaced gels. Most recently, fracturing fluid has been refined in multiple dimensions for maximizing output and minimizing costs.

A wide array of chemicals are used in hydraulic fracturing fluid to enhance the productivity of the primary inputs — water and sand. In particular, firms use chemical additives to help open fractures in the rock, transport the proppant along the length of the fracture,

[5]Shale gas grew from 5% of total US dry gas supply in 2004 to 56% in 2015 (http://www.eia.gov/conference/2015/pdf/presentations/staub.pdf). Spurred on by recent developments in hydraulic fracturing and horizontal drilling technologies, natural gas has largely replaced coal in the production of electricity (http://www.cnbc.com/2015/07/14/natural-gas-tops-coal-as-top-source-of-electric-power-generation-in-us.html). The largest contribution from any one shale play to the growth described above has come from the Marcellus Shale, located in Pennsylvania and West Virginia. Due in part to the availability of pre-disclosure data, Pennsylvania is our area of study (see Section 3).

lower viscosity in order to allow faster pumping and higher pressures, minimize fluid loss into the face of the formation, reduce scaling on the formation, reduce chemical corrosion or bacterial growth that might threaten the integrity of metal casings, facilitate breakup of other chemicals post-fracture, and serve other purposes (Stringfellow et al., 2014; Montgomery, 2013; Gulbis and Hodge, 2000). In short, fracturing fluid is a complex mixture in which an additive that improves performance in one dimension may reduce performance in another. Although the cost of the chemicals themselves is usually small in comparison to the overall cost of the well stimulation operation, the proper choice of chemicals may have dramatic effects on the overall cost and productivity of a well.[6]

As of 2014, the total estimated recovery for shale oil wells was on the order of 5%, compared to 50% for conventional oil wells.[7] As engineers seek to improve recovery from shale wells, innovation continues on several elements of the technology, including the use of longer fractures, greater use (per foot) of water and proppant, shorter stages and "micro-perforations", and improved identification of naturally existing fractures through higher-resolution micro-seismic mapping. Designing fracturing fluid for optimal performance is complementary to several of these elements, and represents a significant area of focus for oil and gas engineers, for advancement of shale production technology (Robart et al., 2013; Montgomery, 2013; Gulbis and Hodge, 2000).

In some cases, the quest for superior fracturing fluids has led engineers to consider the use of highly toxic chemicals (Stringfellow et al., 2014). Indeed, many of the chemicals used in fracturing fluid are known to be toxic to human health, or may cause damage to the ecosystem. These risks have raised concerns among environmental groups (Elgin et al., 2012; Haas et al., 2012). Early concerns about shale gas development were driven by the possibility that toxic fracturing fluid might migrate to or be accidentally released into ground water or surface water. The industry responded to these concerns by pointing to the small percentage of fracturing fluid that is actually comprised of substances other than water and sand.[8] Public concerns about risks to water sources were accentuated by the often close

[6]Mark Boling, Southwestern Energy, personal communication.

[7]R. Kleinberg, Schlumberger, April 2014: "Shale Gas & Tight Oil Technology: Evolution & Revolution", presentation to US Association for Energy Economics.

[8]The typical proportion of chemicals in slickwater fracturing fluid, other than water and sand, is 2 to 3%. Nonetheless, for a typical operation that uses on the order of five million gallons of fluid, even 1% of the

proximity of well-pads to residential and other non-industrial land uses, and by a few high-profile incidents of water pollution.[9] Media coverage of fracturing chemicals has highlighted the toxicity of some chemicals along with the industry's desire to maintain secrecy over the specific chemicals involved (Elgin et al., 2012; Haas et al., 2012).

2.2 Policy Trade-offs of Information Disclosure

The environmental and health concerns discussed above prompted calls for disclosure rules. While this sort of information provision seems to be a simple tool, it can have powerful effects: on the positive side, market or political pressure may limit negative externalities at a lower cost than other regulations; on the negative side, to the extent that firms rely on secrecy to maintain competitive advantage, it might undermine incentives to make costly investment in innovation.

These rules, as summarized in Konschnik (2014), may be valuable to stakeholders for other reasons, many of which are more direct. For instance, in the event of an accident, disclosure to emergency medical personnel and medical staff may improve treatment and protect the staff members themselves. In addition, disclosure provides information to nearby landowners and local government authorities so that they can test their water supplies, increasing their ability to bargain with operators. The information provided by disclosure can also help establish liability for contamination, both by allowing particular contaminants to be traced, and by allowing the comparison of contaminant usage across sites. If the information made available extends over a period of time, then disclosure can facilitate the monitoring of environmental releases, exposures, or health impacts over time. In light of these points, disclosure policies can be complementary to other aspects of regulation — though it is worth pointing out that many of these benefits can be achieved just as well through disclosure to a competent regulator as to the broader public. Finally, disclosure satisfies the public's "right-to-know" about possible release of, or exposure to, hazardous materials, in the spirit of the Emergency Planning and Community Right-to-Know Act (EPCRA) of 1986 (although the shale gas industry is currently exempt from some provisions of this Act).

fracturing fluid would represent 50,000 gallons.

[9] See, for example, http://www.vanityfair.com/news/2010/06/fracking-in-pennsylvania-201006.

As an aside, we note that public disclosure is only one approach that regulators might take in order to limit these externalities. For example, command-and-control regulations would require firms to undertake particular technologies or practices, though they are unlikely to be optimal when regulators have incomplete information about production processes. Market-based regulations (e.g., severance taxes and impact fees) modify firms' incentives via price effects, but may affect firms only on the extensive margin, and may be difficult to implement if chemical releases are difficult to monitor.

In contrast to these approaches, information-based regulations simply require regulated entities to disclose elements of their production process that may have external impacts and which would be difficult for outside stakeholders to ascertain without the disclosure requirement. Prior research (Lyon and Maxwell, 2004; Bennear and Olmstead, 2008) has pointed out mechanisms by which these sorts of laws can prompt self-regulation among firms, and they have been used with success in consumer-facing industries (Fung et al., 2007).

Along with these potential benefits however, disclosure can come with costs. In particular, when the regulator promises to disseminate the new knowledge generated by costly innovation, it reduces the incentive to invest in that innovation in the first place.[10] It is easy to forget this cost when considering disclosure's benefits: it is more difficult (or impossible) to measure and is paid only over a longer time horizon. However it should not be dismissed simply for a lack of immediacy: innovation and technological progress are at the root of economic prosperity (Mokyr, 1992; Lucas, 1988).

We conclude this section by noting that much of the technological development that facilitated the "shale revolution" was based on R&D that was directly sponsored by the government, or by government-subsidized R&D in the form of tax credits for development of unconventional shale and tight gas. In response to persistent supply shortages in the 1970s, the US government began to directly and indirectly fund R&D investment in the natural gas industry. The Natural Gas Policy Act of 1978 removed wellhead price controls and provided

[10]The patent system is one framework for addressing this issue. In practice, the protection afforded by patents varies substantially across industries, and would-be innovators in many industries rely on secrecy, lead time, and investments in complementary assets in order to maximize the returns to innovative activity (Cohen et al., 2000; Cohen, 2010; Teece, 1986). A survey administered in 1994 to R&D laboratories in the US manufacturing sector indicated that on average, lab managers in the petroleum and chemicals industries considered secrecy more effective than any other mechanism, including patents, for protecting both product and process innovations (Cohen et al., 2000).

tax incentives for developing new natural gas resources. The Crude Oil Windfall Profit Tax Act of 1980 provided tax credits for developing unconventional fuels, which increased their financial return and reduced their risk. The Department of Energy (DOE) initiated a number of Unconventional Natural Gas Research Programs, including the Eastern Gas Shales Program, the Western Gas Sands Program, the Methane Recovery from Coalbeds Program, the Seismic Technology Program, and the Drilling, Completion and Stimulation Program. While these programs did not have a role in the development of horizontal drilling or 3D seismic imaging technologies, they did play an important role in high-volume fracturing and micro-seismic fracture mapping (Shellenberger et al., 2012; Wang and Krupnick, 2013).

2.3 Disclosure Rules

There are currently eighteen states with significant hydraulic fracturing activity and chemical disclosure laws. There is general uniformity across these states in terms of the required information that operators must disclose, including ingredient name, chemical abstract service (CAS) number, concentration in the fracturing fluid (typically the maximum concentration used in any fracturing stage), the name of the supplier, and the trade name if applicable.

There is less uniformity in terms of where the information must be registered. Five of the eighteen states, including several of the states that passed the earliest disclosure rules, require operators to report to a state regulatory agency or commission. Six require that operators report to FracFocus, an online database created by a multi-state commission in partnership with a non-profit organization (Council and Commission, 2015). When uploading information to FracFocus, operators are also asked to provide information about well location and characteristics including vertical depth, volume of water used, latitude and longitude, and well name. The seven remaining states allow operators to choose their reporting location (i.e., to FracFocus or the state).[11]

States have adopted similar approaches to accommodating the need for trade secrets, partly due to the Uniform Trade Secrets Act.[12] In particular, all states allow exemptions

[11] Although Oklahoma notes that the state regulator will upload to FracFocus any information it receives.

[12] The Uniform Trade Secrets Act, which seeks to harmonize standards for trade secret protection, was promulgated by the Uniform Law Commission in 1979 and passed by 46 states.

for the disclosure of additives considered to be confidential business information that firms believe gives them a competitive advantage. Operators must declare an exemption for individual chemical ingredients for which they claim trade secret status. This is accommodated by FracFocus, which allows for uploaded information to include the concentration of the chemical used but not its name or chemical identification number. Some states also require operators to report the chemical family to which the proprietary substance belongs.

3 Data

As this paper seeks to understand the effect of disclosed information, a central empirical challenge is to identify precisely what information is available to each operator at the time of fracturing each new well. This information includes inputs (including chemicals used in stimulation) as well as outputs (gas production). At the time of a given well fracture operation, each operator has access to such data from its own prior wells, as well as (in some cases, and subject to a delay that varies with the extent of disclosure requirements) similar data from other operators. In order to address this empirical challenge, we assemble a rich and novel dataset that includes production inputs and outputs, the date of each fracture, the date (if any) that input information became publicly available for different-operator wells, the operator name, and, if available, the name of the fracturing contractor. We compile this dataset for all unconventional wells in Pennsylvania, using a variety of sources.

We start with all of the unconventional wells in Pennsylvania with completion dates and initial production between January 2007 and May 2015 — 6,545 wells in total. We focus on shale gas development from unconventional reservoirs (as defined by the Pennsylvania Department of Environmental Protection, or DEP) because that is the primary locus of firms' innovation during this period, as well as the focus of the regulatory disclosure laws that are central to our analysis.[13] Our production data comes from DrillingInfo, a national

[13]According to industry engineers and geologists we consulted, the areas of current, active technological innovation that is relevant to fracturing operations are largely distinct between unconventional and conventional production. In other words, learning about production in conventional reservoirs does not transfer readily to provide insights into production in unconventional reservoirs. These experts did advise us that learning about fracturing vertical wells is transferable to fracturing in horizontal wells, and vice versa; thus we include both vertical and horizontal wells in our analysis.

provider of information on the oil and gas industry.

We collected data on inputs (including chemicals) and completion dates from two sources: Well Completion Reports and Stimulation Fluid Additive reports from DEP, and the FracFocus database. Well Completion Reports, which operators must submit within 30 calendar days following completion, contain the name of the operator (and, in many cases, the fracking contractor), well location, and information about the perforation and stimulation process.[14] Effective in February 2011, operators also had to submit information on chemicals used in the stimulation process, including the chemical name, chemical identification number, and concentration of each additive in the fracturing fluid.

Under the February 2011 law, operators had to submit information about chemicals to DEP either along with the Well Completion Report or on a separate DEP form, the Stimulation Fluid Additive report. Some operators elected to submit chemical additive information to the FracFocus registry as well as on state forms, often submitting a printout of the FracFocus disclosure form to the state rather than using the official state form. Operators from other states, too, were uploading chemical additive information to FracFocus at the same time. Most such disclosure was voluntary at the time: Wyoming and Arkansas had also passed chemical disclosure laws as of February 2011, but both provided standalone websites for this purpose. By April 2012, when Pennsylvania switched to FracFocus as its required reporting site, several other states had passed chemical disclosure regulations; some required the use of FracFocus, and others used a state registry instead.

Notably, information that operators submitted to FracFocus was more readily observable by competitors and the public; from the start, the site allowed users to download individual PDF files with the chemical composition of fracturing fluid as well as well location, operator name, water volume, and other items.[15] In contrast, Well Completion Reports and Stimulation Fluid Additive reports submitted to the DEP were available for review by a paid subscription service or by in-person review at regional DEP offices. Subscribers could view Well Completion Reports through the state-run Exploration and Development Well Informa-

[14] Although the chemical disclosure requirement is fairly new, operators have been required to submit Well Completion Reports since 1989; see http://www.pacode.com/secure/data/025/chapter78/s78.122.html.

[15] Initially FracFocus was designed to inhibit large-scale downloads, but at least two users (one nonprofit and one for-profit) successfully scraped the database by late 2012 and one of these entities provided a public download soon after (Skytruth, 2013).

tion Network (EDWIN), in which some chemical disclosure forms were available as scanned PDF documents.[16] However, in many cases there was a long lag after well completion before stimulation chemicals and other information became available in the system, especially during the height of the fracturing boom. For instance, one of the authors accessed the system several times in 2012-13 and found lag times for stimulation fluid additive reports of 18 months — or more in some cases — were not uncommon. We also found that several dozen reports for wells completed in 2011 or early 2012 were uploaded into the system between December 2015 and January 2017. Regulatory staff responsible for uploading stimulation chemical reports told us that the most important contributing factors for delayed upload were a cumbersome workflow and inadequate staff resources relative to workload, which is also supported by contemporaneous internal assessments (e.g., see Pennsylvania Department of Conservation and Natural Resources, 2010).

Organizations that subscribed to EDWIN — as well as members of the public — could also review physical copies of the Well Completion Reports and Stimulation Fluid Additive reports, without having to wait for upload into the system. However, to do this, they would have had to identify the well permit number, contact the appropriate regional DEP office, file an official records request, schedule an appointment (typically three to four weeks in advance), and would then be allowed to review a limited number of hard copy documents on site, on the order of 25 per day. Furthermore, in our own experience using this in-person inspection system, we found that some requested hard copy documents (on the order of 20% for some regional offices and times) were not available on the appointment date, typically due to internal processes by which other departmental users had "checked out" the hard copy reports. For this reason, we do not believe operators would have expected others to observe their fluid contents, prior to the April 2012 rule revision that *mandated* public disclosure of chemical information on FracFocus.

We assume full information flow within each firm's internal organization: that is, that each operator had full information about the inputs and outputs for all prior same-operator wells immediately upon completion and initial production. To determine the date by which operator i could learn about the inputs used for subsequent wells fractured by operator $j \neq i$,

[16]Prior to 2015, EDWIN was known as the Integrated Records and Information System, or IRIS.

we incorporate information on the date (if ever) that a well was uploaded to FracFocus. We obtain these dates from Konschnik and Dayalu (2016), who in turn obtained them from the administrators of the FracFocus registry. In the case of wells for which operator $j \neq i$ submits data only to the DEP, and does not upload a report to FracFocus, we assume that operator i does not observe the chemical input data during the period of our study (i.e., by May 2015).

To capture productivity, we use the standard industry metric of initial gas output (first six months of production) per foot of wellbore.[17] Total output is highly correlated with initial output (Newell et al., 2016), and dividing by the length of the perforated interval normalizes output by well size to facilitate comparison across wells. In the process of creating this metric, we find that operators failed to provide the length of the perforated interval for 168 wells, so we cannot use these in the analysis.[18] We drop an additional 282 wells that have nonsensical completion dates (the recorded completion date is more than 30 days after the date of initial production) or insufficient information on upload date. Finally, we drop 1,233 wells for which we lack information on key inputs, i.e., chemicals or water volume.[19]

This leaves a sample of 4,862 wells. For each, we observe identifying information (operator, location, and completion date), output (initial gas production per perforated foot), and inputs (volume of water and chemical additives to the stimulation fluid). Data on chemical inputs is certainly the most difficult-to-access element that we observe; as noted above, we collect this data from two sources: FracFocus, and DEP reports. The former are available in digital form by combining a downloadable database from FracFocus itself (for wells fractured from approximately May 2013) and a dataset released by an environmental NGO that scraped the "version 1" FracFocus website Skytruth (2013).

The latter source — chemical information from DEP reports — are substantially less accessible, as they are not amenable to optical character recognition. The DEP and firms

[17] To ensure the analysis is not driven by outliers, we winsorize per-foot initial gas production at the 99th percentile.

[18] This omission of perforated interval appears to be out of line with basic reporting requirements.

[19] Nine of these wells appear to be out of compliance with Pennsylvania law, based on their lack of chemical disclosure and completion date after February 2011. Another possibility is that operators provided chemical information for these wells, but the reports had not been digitized and uploaded to EDWIN or its predecessor system by January 2017 when we searched the database for Well Completion Reports and Stimulation Fluid Additive reports.

employed a wide variety of formats over the time period we analyze — over ten different formats, often with different headers on the pages of interest — and also use numbers that are sometimes handwritten or crossed out and overwritten, as well as overlaid date stamps and raster images. Thus, we employed a team of research assistants to digitize the relevant information manually. This effort took about 1,800 person-hours over 4 months, and involved the entry of about 200 data items per report. We verified the quality of our data entry process with systematic checks for consistency and reasonableness, comparison of duplicate entries by different workers, and random spot checks in which we compared the original reports to hand-entered data.

Despite our careful collection and entry of input data, not all variables are available for every well. In addition to fluid volume we also considered incorporating information on proppant volume, and on the number of stages in a frac; however, proppant volume is available for only 3,747 wells (77%), information on stages per frac is available for only 3,421 wells (70%), and only 2,802 wells (58%) have information on both proppant and stages. For the wells where we do have complete information, we analyzed the variation in initial production per foot that is explained with and without these additional variables and found that including proppant increases the portion of variation explained (based on adjusted R^2) from 0.4579 to 0.4873 (nearly 3 percentage points); including proppant and stages increases that to 0.4881.[20] We concluded that to include these variables would add only a small increment to explanatory power, but result in a substantial cut in sample size.

For most of the wells in our sample (4,595), chemical information is from FracFocus; we have information from DEP for 267 wells. It is worth noting that for many of the wells whose formulations were eventually released on FracFocus, the upload date was substantially later than the completion date. For wells fracked prior to the mandatory public disclosure law in April 2012, the median and mean time lapse from completion to upload was 113 days and 231 days, respectively (compared to 49 days and 99 days for wells fracked after mandatory disclosure). We use the well-specific upload date from Konschnik and Dayalu (2016) to identify the information set available to each operator at the time of fracturing each well.

Table 1 provides a summary of information about the wells for which we observe chemical

[20]For consistency, this comparison includes only those wells for which proppant and stages are available.

input information. Although we have other information for wells beginning in 2007, chemicals data are generally available only for wells starting in 2010. (Though mandatory disclosure started in Pennsylvania in 2011, some operators voluntarily submitted chemical information for some wells prior to this date — perhaps in an effort to forestall the very regulations that would require mandatory disclosure of all wells.)[21] The table shows that the median and mean number of chemicals per well has increased slightly over time, fluid volume has increased somewhat more substantially, and gas production per foot has generally increased.

Table 2 provides additional information for wells in the sample for which we observe chemical use. This table demonstrates that operators continued to innovate by introducing new chemicals each year, with the largest innovation (in terms of number of new chemicals) occurring in 2011-2012 — right around the time that the state was issuing laws requiring mandatory disclosure. Companies also retired some chemicals from use (that is, we do not observe their subsequent use in our sample).

Table 1: Summary Information for Wells in Sample with Chemicals Data

	Fluid Volume		# Chemicals / Well		First 6 Month Gas per Foot	
Year	Median	Mean	Median	Mean	Mean	Median
2010	4.8	4.8	11	11.5	171	141.6
2011	4.3	4.4	9	10.1	157.2	123.5
2012	4.1	4.4	11	12.3	172.3	122.9
2013	5.7	6	13	15.3	192	148.5
2014	7.8	8.2	13	15.7	190.1	150.4
2015	8.5	8.6	14	15.6	171.7	150.7

Notes: Data sources are described in Section 3. Chemicals are listed, non-proprietary chemicals with legitimate CAS numbers. Fluid volumes are expressed in millions of gallons. Gas volumes are expressed in thousands of cubic feet (mcf).

4 Analysis

In this section, we perform a series of empirical analyses designed to answer two questions. First, did mandatory disclosure of chemicals, even with provisions allowing for some chem-

[21] Lyon and Maxwell (2004) document several explanations for why firms would voluntarily (but selectively) disclose valuable private information about production processes, including to forestall more stringent regulation.

Table 2: Additional Information for Wells in Sample with Chemicals Data

			Chemicals		
Year	# Wells	# Operators	# New	# Retiring	# Total
2010	103	14	91	5	91
2011	1145	35	65	20	145
2012	1230	32	96	22	214
2013	1108	33	62	65	255
2014	1098	32	34	108	228
2015	177	19	6	–	134

Notes: Data sources are described in Section 3. Chemicals are listed, non-proprietary chemicals with legitimate CAS numbers. New chemicals refers to the count of unique chemicals appearing in our dataset for the first time in a given year. Retiring chemicals refers to the count of unique chemicals appearing in our dataset for the last time in a given year. Total chemicals refers to the count of unique chemicals used in a given year.

icals to be declared proprietary, create conditions by which firms could learn from each other? Second, did the learning enabled by disclosures have value, enabling operators to improve productivity through their chemical choices? After considering these questions we also address, in Section 5, whether mandatory disclosure reduced operators' innovative experimentation with new chemical formulations.

For our initial analysis, which serves primarily as motivation for deeper exploration, we define two periods: pre- and post-disclosure. Thanks to Pennsylvania's regulatory history, the pre-disclosure period covers the 14 months from February 2011 to April 2012. During this period we as researchers are able to observe the chemicals used for all wells, but it is implausible that operators had access to this information (see Section 3). The post-disclosure period covers all of the wells from April 2012 through the end of our sample.

If the answer to both questions is affirmative, we would expect to see convergence in inputs across wells, with firms using available information (revealed by the disclosure laws) to imitate successful wells. After confirming that this is the case, we examine each of the two questions more closely to rule out alternative explanations. To answer the first question, we examine how the design of fracturing fluids changed after disclosure. After finding a convergence in the chemicals used consistent with copying from disclosed wells, we turn to the second question. To answer it, we estimate a non-linear model relating well productivity

to an index of input similarities with other wells. We find evidence that using chemicals more similar to those used in high-productivity firms' disclosed wells is associated with higher productivity. Together, these findings are consistent with firms learning through chemical disclosure.

4.1 Change in Well Productivity

As a starting point in our analysis, we consider the hypothesis that disclosure led to convergence in well productivity. To investigate this, we run a simple regression at the well-level:

$$g_i = \beta_0 + \beta_1 f_i + \gamma_i + \epsilon_i, \tag{1}$$

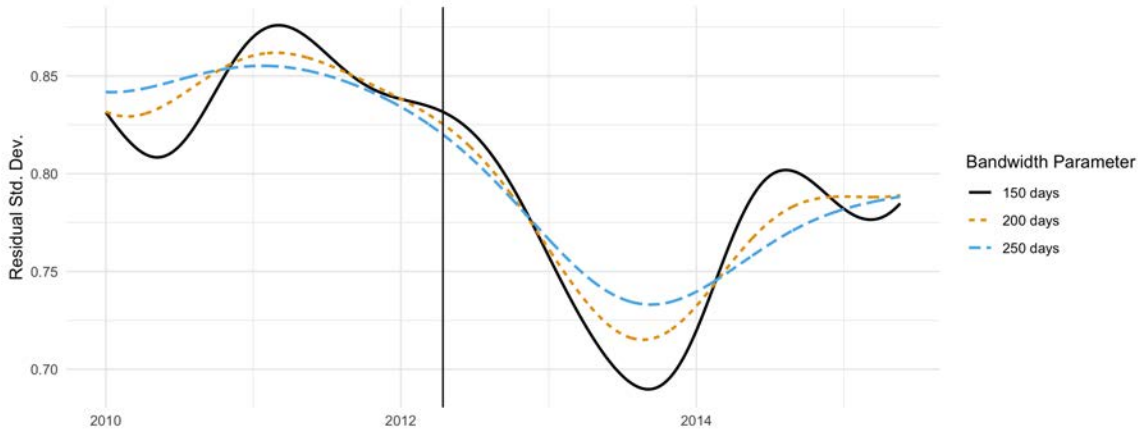
where the i subscript indicates a well, g denotes log per-foot first-6-month gas production, and f is log fracturing fluid volume; γ_i is a set of fixed effects for each well's county and year.[22] We then take the estimated residuals $\hat{\epsilon}_i$ and calculate a Gaussian kernel-weighted standard deviation over time. The results are shown in Figure 1. The curves plot the estimated standard deviation, for various bandwidth choices; the vertical black line marks April 2012, the date that disclosure came into effect. All three bandwidths show a marked fall in the standard deviation of the estimated residuals after disclosure. This fact motivates our following analysis.

4.2 Well-to-Well Similarity Measures

It is possible that the drop in residual standard deviation described in Figure 1 is simply the result of the shale gas industry maturing, with individual operators all converging towards a production frontier. One way to test the learning hypothesis would be to look for evidence that poor performers improve their productivity specifically when they copy the chemical mixtures of more successful firms.

[22]The sample for this regression and the accompanying figure is slightly different from that of the rest of the paper, as it does not require the chemical data. We use wells for which we have data on gas production and fluid volumes, and both variables are recorded to have a positive value. This filter leaves us with 5,279 wells through 2015. The trend observed in Figure 1 is quite similar under the chemicals-observable sample used in the rest of the paper, or other reasonable specifications of Equation 1.

Figure 1: Kernel-Weighted Standard Deviation of Productivity Residuals over Time



Notes: The curves indicate the kernel-weighted estimated residual standard deviation over time, with different bandwidths. The vertical line indicates the date that public disclosure came into effect. See text and footnote 22 for details on the sample used.

We first consider the effect of Pennsylvania's disclosure rules on operators' chemical choices. If chemical disclosure compelled the release of valuable (formerly private) information, we might expect to find that operators' chemical recipes exhibit more similarity after disclosure than beforehand. Our detailed chemical input data (discussed in Section 3) allows us to explore this hypothesis by constructing measures of similarity for each well-to-well pair we observe.

We begin by introducing some notation, defining our similarity index of choice, and providing a summary of the index in our data. Let $\mathcal{C}$ denote the set of all possible chemicals, and let $x_{ic} \in [0, 1]$ be the concentration of chemical $c \in \mathcal{C}$ in well i. The sum of all such concentrations is therefore $\sum_{\mathcal{C}} x_{ic} = 1$. We also define the binary variable $y_{ic} = \mathbb{1}\{x_{ic} > 0\}$. Then we can define the following quantities for a given (i, j) pair:

- $A_{ij} \equiv \dfrac{\sum_{\mathcal{C}} x_{ic} y_{jc}}{\sum_{\mathcal{C}} x_{ic}}$ is the concentration-weighted share of well j chemicals in well i's formula; and
- $B_{ij} \equiv \dfrac{\sum_{\mathcal{C}} x_{jc} y_{ic}}{\sum_{\mathcal{C}} x_{jc}} = A_{ji}$ is the converse.

We use the Jaccard index, under which the pairwise similarity between wells i and j can

be defined as:[23]

$$s_{ij} \equiv \frac{A_{ij}B_{ij}}{A_{ij} + B_{ij} - A_{ij}B_{ij}}.$$

The Jaccard index has several appealing properties. It is bounded by [0,1], which allows for easier integration with our other metrics, notably the quality-similarity index defined in Section 4.4. (A value $s_{ij} = 1$ implies that the hydraulic fracturing fluids used in wells i and j are indistinguishable; $i = j \Rightarrow s_{ij} = 1$. For completely dissimilar fluids that have no chemicals in common, $s_{ij} = 0$.) Because the index is based on sets rather than vectors, it is not sensitive to effects that can manifest with vector-based metrics in high-dimensional data. For instance, a Euclidean distance metric based on vector endpoints might identify a pair of wells that each have a large diversity of chemicals, but none in common, as "close", because both endpoints are relatively "close", in a multidimensional sense, to the origin. The Jaccard index is not sensitive to this effect, which is a concern in our setting.

4.3 Disclosure and Similarities

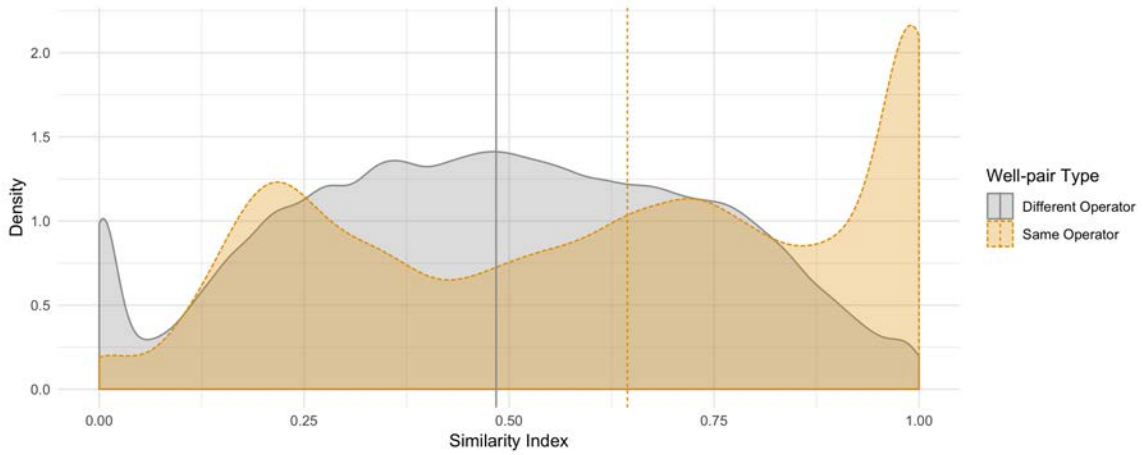
We calculate the similarity index for each i, j well-pair in our data. Our chemical data on 4,944 wells gives us 12.2 million such pairs. Figure 2 shows the distribution of these measures for two sub-samples of pairs: those where both wells are drilled by the same operator, and those where the wells are drilled by different operators. The figure also plots the medians of the two distributions. A few observations are noteworthy: the different-operator distribution has a mass point at $s_{ij} = 0$ and a median slightly below $s_{ij} = 0.5$. In contrast, the same-operator distribution has a large mass point at $s_{ij} = 1$ and a median above $s_{ij} = 0.6$. This confirms what we might have expected: intra-operator well-pairs tend to use more similar chemical mixes than inter-operator well-pairs.

We next consider how well-pair similarities have changed with the advent of disclosure. We do this by running a regression of the following form:

$$s_{ij} = \beta_0 + \beta_1 oper_{ij} + \beta_2 oper_{ij} * visible_{ij} + \phi_{ij} + \epsilon_{ij} \tag{2}$$

[23] See chapter 3 of Leskovec et al. (2014) for more details on the Jaccard index.

Figure 2: Jaccard Similarities by Same / Different Operator Status



Notes: The curves plot estimated densities for well-pair similarities s_{ij}, conditional on whether the wells in the pair had the same or different operators. The vertical lines indicate the distribution medians.

where s_{ij} is the Jaccard similarity index defined above, $oper_{ij}$ is a binary variable equal to 1 if wells i and j are drilled by different operators, and $visible_{ij}$ is a binary variable equal to 1 if *either* the wells were completed by the same operator, *or* the chemicals used in the earlier well of the well-pair were disclosed prior to the later well's fracture. It is worth noting that the visibility of the earlier well's chemicals cannot be taken for granted, even if there is a long lag between the wells; we therefore use the upload date to determine visibility rather than simply the date of the completion. Section 3 discusses the high variance of lags between completions and chemical disclosures. Fixed effects for the first well's operator, the second well's operator, and the year of the second well's completion are represented by the ϕ_{ij} term. The results of this regression, estimated separately with and without the fixed effects, are shown in Table 3.[24]

For the purposes of inference, we note that the usual OLS standard error assumptions are not satisfied for Equation 2 — given that the regression takes place at the well-pair level, we must account for the possible correlation across well-pairs that share a particular well. Therefore the standard errors we report in Table 3 are calculated from a standard bootstrap routine with 1,000 iterations.[25]

[24]We drop from this analysis roughly 8,000 well-pairs where each of the two wells was completed on the same day, having no way to assign a "first" or "second" operator.

[25]We are grateful for a discussant pointing out that another approach would be to use the asymptotics developed in Honoré and Powell (1994), who consider a more general case.

Table 3: Determinants of Well-Pair Similarities

	Similarity Index	
	(1)	(2)
Different Operator	-0.126^{***}	-0.195^{***}
	(0.000)	(0.000)
Different Operator x Visible	0.005^{***}	0.009^{***}
	(0.000)	(0.000)
Constant	0.606^{***}	
	(0.000)	
Operator and Year FEs		✓
Observations	11,808,695	11,808,695
R^2	0.014	0.339

Notes: Results are from the regression described in (2); the similarity index is the Jaccard Abundance index, described in detail in Section 4.2. Standard errors are calculated from a standard bootstrap routine, with 1,000 iterations. *, **, and *** indicate statistical significance at the 10, 5, and 1 percent levels.

A few things are worth noting from Table 3. First, both columns show a significantly negative estimate of the coefficient on $oper_{ij}$; this indicates that well-pairs with different operators tend to use less-similar chemical mixes, confirming the graph in Figure 2. Second, the coefficient on the interaction of $visible_{ij}$ with $oper_{ij}$ is significantly positive, indicating that different-operator well-pairs use more similar chemical mixes when the earlier-used mix is known to the later-fractured well. This is consistent with disclosure facilitating the transfer of knowledge about chemical mixes, and motivates our further analysis in the next section.

4.4 Effect of Information on Productivity

Next, we estimate a model in an attempt to understand the relationship between disclosure and information, the choice of chemicals, and well productivity. The model accounts for important observable variables, and additionally relates firm and well productivity to each other through the channel of disclosed information. We find that disclosed chemical information can play a statistically and economically significant role in improving well productivity.

We begin by defining the regression equation of interest:

$$\ln gpf_i = \beta X_i + \theta_{f(i)} + \alpha_{OBS} \ln(1 + QS_i^{OBS}) + \alpha_{UNOBS} \ln(1 + QS_i^{UNOBS}) + \epsilon_i \,, \qquad (3)$$

where gpf_i is the per-foot quantity of gas produced over the first six months for well i, X contains observables such as the amount of fracturing fluid injected, a binary indicator of whether the well operator appeared to have throttled production, the region where the well is located, and the price of natural gas when the well is fractured, and $\theta_{f(i)}$ is a productivity fixed effect for the firm $f(i)$ that drilled well i.[26] In addition to these linear terms, and an assumed iid error term, we also include two additional terms containing objects QS_i^{OBS} and QS_i^{UNOBS}, which are included as natural logs in order to facilitate an elasticity interpretation.[27]

These quantities are defined:

$$QS_i^{OBS} \equiv \sum_{f \in \mathcal{F} \setminus f(i)} \left(\theta_f \frac{1}{|\mathcal{J}_f^{OBS}|} \sum_{j \in \mathcal{J}_f^{OBS}} s_{ij} \right) , \qquad (4)$$

$$QS_i^{UNOBS} \equiv \sum_{f \in \mathcal{F} \setminus f(i)} \left(\theta_f \frac{1}{|\mathcal{J}_f^{UNOBS}|} \sum_{j \in \mathcal{J}_f^{UNOBS}} s_{ij} \right) . \qquad (5)$$

In these equations, $\mathcal{F}$ indicates the set of firms in the data, $\mathcal{J}_f$ is a set of wells j drilled by firm f before well i, and s_{ij} is the same well-pair similarity index as given in Section 4.2. The superscripts denote whether or not the chemical data for well j was uploaded and observable prior to well i's fracture — those wells whose chemicals were observable fall into $\mathcal{J}_f^{OBS}$, and those whose were not fall into $\mathcal{J}_f^{UNOBS}$. QS_i^{OBS} is thus a quality-weighted measure of the similarity between well i and those wells whose chemical information was observable when well i was fractured. Its inclusion in Equation 3 thus accounts both for how similar well i is to its comparison wells $\mathcal{J}$, and also how successful those wells are (as measured by their firm

[26] Throttling production refers to intentionally restricting hydrocarbon flow, which is a process operators sometimes engage in for geological, economic, or logistic reasons. We classify a well as having been throttled if the ratio of first 18 months' production to first 6 months' production is greater than 3, which is true for 11% of the wells in our data.

[27] We shift the function by adding 1 before taking logs, as many QS_i are equal to 0.

fixed effects, θ). QS_i^{UNOBS} is an analogous quality-similarity index for the recently drilled but *unobserved* wells.

We consider different possible look-back restrictions on $\mathcal{J}$, e.g. restricting wells j to have been fractured within 180 days of well i's fracture. Results under several different assumptions are presented in Table 4.

This separation of recently fractured wells into sets whose chemicals are observed and unobserved is a key component of the analysis. Identification is coming not just from the similarity between well i and its recently fractured peers, but *also which of those peers had observable chemical information* at the time. As discussed in Section 3, there is significant variation in the lag between completion and upload: following disclosure, the mean and median lags are 99 and 49 days respectively.

4.4.1 Estimation and Results

The inclusion of the QS_i terms render the firm fixed effects θ_f non-linear parameters — the same θs that appear on their own in Equation 3 are the components of the QS^{OBS} and QS^{UNOBS}. We can only infer the 'quality' portion of the QS objects from the estimated θs, and Equation 3 restricts that those estimated θs account for the observable and unobservable similarities between wells. The two quantities are therefore calculated in a single estimation procedure: we estimate Equation 3 using a non-linear least squares (NLLS) approach, and present results in Table 4.

From the first column of Table 4 it can be seen that increasing fluid volume by 100% is associated with a 18.7% increase in productivity, and throttling the well is associated with an $e^{-0.570} - 1 = -43.4\%$ decrease in productivity per foot, both statistically significant at the 1% level. In turn, those wells drilled in the northeast region of Pennsylvania are estimated to be $e^{0.398} - 1 = 48.9\%$ more productive per foot, and a \$1 increase in the Henry Hub price of natural gas is associated with a 6.6% increase in productivity. Conveniently, the coefficient on QS_i^{OBS}, estimated to be statistically significant at the 5% level, can be understood as an elasticity: a 100% increase in $(1 + QS_i^{OBS})$ is associated with a 4.2% increase in productivity.[28] In turn, the coefficient corresponding to QS_i^{UNOBS} has a much

[28]Some caution should be used when interpreting these results, as it is not necessarily straightforward

Table 4: Quality-Similarity Results

	$\ln GPF_i$			
$\mathcal{J}$ Restriction	180 Days	360 Days	720 Days	Unrestricted
ln Fluid Volume	0.187^{***}	0.187^{***}	0.187^{***}	0.183^{***}
	(0.019)	(0.019)	(0.019)	(0.020)
Throttle	-0.570^{***}	-0.570^{***}	-0.570^{***}	-0.576^{***}
	(0.036)	(0.036)	(0.036)	(0.037)
Northeast	0.398^{***}	0.398^{***}	0.397^{***}	0.434^{***}
	(0.043)	(0.043)	(0.043)	(0.043)
HH Price	0.066^{***}	0.065^{***}	0.064^{***}	0.062^{***}
	(0.014)	(0.014)	(0.014)	(0.014)
α_{OBS}	0.042^{**}	0.036^{**}	0.030^{*}	0.026^{*}
	(0.017)	(0.017)	(0.017)	(0.015)
α_{UNOBS}	0.012	0.016	0.020	-0.018
	(0.017)	(0.017)	(0.017)	(0.025)
Operator FEs	✓	✓	✓	✓
N	4,862	4,862	4,862	4,862

Notes: Coefficients are estimated via non-linear least squares (NLLS). Each column conducts the estimation using a different specification of $\mathcal{J}$ when constructing the QS_i^{OBS} and QS_i^{UNOBS} weights. The different columns represent *, **, and *** indicate statistical significance at the 10, 5, and 1 percent levels. See text for details.

smaller point estimate, and is not statistically distinguishable from zero.

The second through fourth columns present analogous results for different specifications of the $\mathcal{J}$ used in constructing the QS_i: column 2 restricts the consideration to wells completed within 360 days, column 3 uses 720 days, and column 4 places no time restriction on the construction. The estimated coefficients on the controls $\hat{\beta}$ are relatively unchanged from the first column. In all specifications, $\hat{\alpha}_{OBS}$ is estimated to be positive and statistically significant at the 5% or 10% level, while $\hat{\alpha}_{UNOBS}$ is never statistically distinguishable from zero. Furthermore, a pattern in the estimates of $\hat{\alpha}_{OBS}$ is also evident across the columns: as a longer lookback period is used, the point estimates decrease.

Altogether, these results suggest that those wells that used a chemical mix that was more similar to those used in the recent, disclosed wells of the more productive operators realized significantly greater productivity than those that did not. In other words, the information revealed through disclosure can be valuable for those operators who pay attention.

It is worth noting that a sophisticated, large, and forward-looking firm might choose to improve its value not by maximizing QS and imitating the current best practice, but rather by experimenting in an attempt to find new, superior chemical combinations. We examine this issue in more depth in Section 5.

4.4.2 Mechanism: Learning Through Contractors

One possible channel for the transfer of information about chemical mixtures is the contractors who are hired to hydraulically fracture the wells. Contractors perform a variety of roles; operators hire them to assist with jobs that may include drilling, cementing, well logging operations, and other tasks, as well as designing and conducting the fracturing job. The DEP requires operators to provide information about contractors on Well Completion Reports. According to these reports, for some wells in our data, operators hired up to 40 contractors to assist in various roles; the median number of contractors per well is 8. The roles these contractors play are often, but not always, specified on the reports. We were able to identify the fracturing contractor for 3,529 wells, about 73%.

to "increase QS_i": each well can only have a single chemical mixture, and thus any changes that make s_{ij} higher for some j may make it lower for others.

To test whether contractors facilitate the transfer of information about chemical mixtures, and whether that role was changed by the institution of disclosure rules, we perform a series of regressions of well-pair similarity indices on dummies for: (i) if the two wells share the same contractor and (ii) if the first well in the well-pair's chemicals were disclosed prior to the second well's fracture ($visible_{ij}$). For these regressions, we consider only inter-operator well-pairs.

$$s_{ij} = \beta_0 + \beta_1 contractor_{ij} + \beta_2 visible_{ij} + \beta_3 contractor_{ij} * visible_{ij} + \phi_{ij} + \epsilon_{ij} \tag{6}$$

For all of these regressions, we restrict ourselves to the sample of well-pairs with data on the fracturing contractor *for both wells* (see above). As in Section 4.3, we include fixed effects for the first and second wells' operator, and the year the second well was completed, and calculate standard errors from 1,000 bootstrap iterations. The results of these regressions are shown in Table 5.[29]

When it is the only non-constant regressor, the dummy for sharing a contractor is associated with an increase in the similarity index of 0.101. In the second column, the dummy for visibility is included: this inclusion has no discernible effect on the contractor coefficient. The third column adds an interaction term between these two dummies. The coefficient for same contractor increases, the coefficient on visibility is altered slightly, and the interaction term is estimated to be negative and significant. The role of the contractor in facilitating a similar chemical mix is reduced by about a fifth when the first well's chemicals are disclosed, suggesting that operators are not as reliant on contractors as a source of chemical information when that information is being published. These results suggest that the contractor channel is associated with more similar wells, but that this channel is less active when chemicals are publicly visible.

[29] As in the analysis in Section 4.3, we drop well-pairs where the wells were completed on the same day. See Footnote 24.

Table 5: Contractors' Role in Similarity of Well-Pairs

	Similarity Index		
	(1)	(2)	(3)
Same Contractor	0.101***	0.101***	0.122***
	(0.000)	(0.000)	(0.001)
Visible		−0.012***	−0.010***
		(0.000)	(0.000)
Contractor x Visible			−0.028***
			(0.001)
Operator and Year FEs	✓	✓	✓
Observations	5,811,335	5,811,335	5,811,335
R^2	0.360	0.360	0.360

Notes: Results are from the regression described in (6); similarities are from the subsample where the fracturing contractor can be identified with confidence. See text for details. Standard errors are calculated through a standard bootstrapping routine, with 1,000 iterations. *, **, and *** indicate statistical significance at the 10, 5, and 1 percent levels.

5 Experimentation

Our findings in the previous section suggest that the mandatory disclosure policy forced operators to reveal economically valuable information. A natural follow-on question is whether the policy also resulted in decreased investments in innovation, or experimentation, among the firms involved. To the extent that exploration and production firms in the shale gas industry rely on secrecy as a mechanism to capture value from their investments in innovation and thus maintain competitive advantage, the erosion of secrecy via mandatory disclosure may also erode private incentives to experiment with new formulations. If the policy resulted in a decrease in experimentation, this would suggest that policymakers considering disclosure policies face a trade-off between benefits of these policies (by providing information to other stakeholders) and potential costs (of reduced innovative activity).

To test for a change in innovative activity, we start by measuring operators' use of experimental chemical combinations. We designate a chemical combination as experimental based

on comparing the chemical identities and proportions used in each well, to the combinations used in previously developed wells. We use a density-based clustering algorithm, DBSCAN, to identify clusters of wells that have similar chemical input combinations. The algorithm also identifies those wells that do not fit into any such cluster, which will lead us to our definition of an experimental well.

One approach would be to label a well as experimental if its corresponding chemical mixture is novel compared to all prior wells. However, this method would be mechanically biased toward finding less experimentation over time since, as more wells are drilled, any new well is more likely to look similar to some prior well. To avoid this, we compare each well to a comparison set that contains the 1000 most recently developed wells, keeping the size of the comparison set constant over time.[30]

The DBSCAN clustering algorithm (Ester et al., 1996) provides an intuitive, density-based approach to identifying clusters of similar items (chemical combinations used for well stimulation, here) and items that do not fit any cluster. The algorithm groups together points that have many nearby neighbors, and also identifies points that are too far away from others, or in low-density regions; points in the latter category are labeled outliers or "noise points". Specifically, the algorithm designates an ϵ-ball around each point and scans for other data points within that ball. The algorithm then continues in iterative fashion, constructing an ϵ-ball around the new collection of points and again searching for data points within the new (larger) ball. To be designated as a cluster, a collection of points must have a minimum cardinality (set as a tuning parameter). Observations that fall outside any cluster — because there is no other observation within the ϵ-ball, or because the observations within a candidate cluster are too few to meet the minimum points threshold — are considered noise points. We interpret these as experimental wells, because they represent novel combinations of chemical inputs that are relatively distant from prior practice.

Compared to alternative clustering algorithms in common use, DBSCAN has several advantages. First, unlike methods that require the number of clusters to be pre-specified as an input parameter (e.g., k-means clustering or finite mixture models), it identifies the

[30]Similar results hold for different alternative values of this "lookback" set, as well as lookback sets based on the number of days rather than number of wells. These results are available from the authors upon request.

number of clusters based on the observed data. Second, DBSCAN does not require the assignment of every point to a cluster — that is, it identifies "noise points" in addition to the clusters. It is also effective in discovering clusters that do not form convex sets. The algorithm also requires the selection of just two parameters, and established data-driven guidelines are available for choosing both. It also accommodates any distance measure (we use the Jaccard distance, $d_{ij} \equiv 1 - s_{ij}$, to be consistent with prior sections). Finally, the algorithm is computationally efficient, and has been used in a wide variety of applications (Ester et al., 1996).

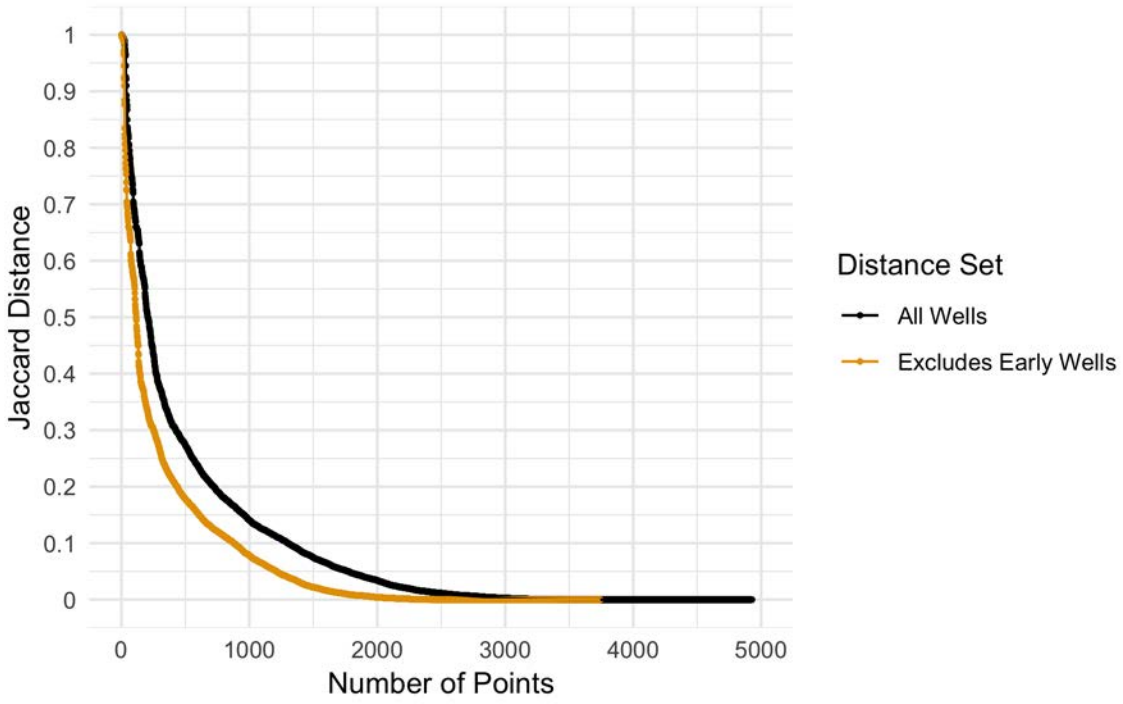
DBSCAN requires the selection of two parameters: the minimum number of points in a cluster, and the value of ϵ. Similar to the question of, say, the critical value to use from a test distribution to determine statistical significance, there is some discretion in the choice of these two parameters, but also a common practice. For choosing the minimum-points parameter, a common heuristic in ordinary or dense matrices is to use the number of dimensions in the data or, for sparse matrices, the intrinsic dimensionality (Johnsson, 2016; Bruske and Sommer, 1998). In our setting the latter applies: there are 379 chemicals used across all of the wells in our data, but on average each well uses 13 chemicals, and the median well uses 11 (the first and third quartiles use 8 and 16, respectively). That is, if the formulation of the median well in our dataset (ordered by the number of chemicals used) were represented as a vector in 379-dimensional space, the value of all but 11 elements would be zero. To estimate the intrinsic dimensionality we use the method of Bruske and Sommer (1998), which uses a form of principal components analysis applied to local subspaces of the observed data to identify the minimum number of parameters needed to describe the well-chemicals data. For our data the resulting estimate of intrinsic dimensionality varies from 11 to 16 depending on the number of local subspaces chosen; accordingly, for the results that follow, we set the minimum-points parameter at 15.[31]

As suggested by Ester et al. (1996), we choose ϵ based on the location of the "knee" of a sorted k-nearest-neighbor (kNN) distance plot: that is, a graph that shows the distance from each point to its kth-nearest neighbor, where k is set at the threshold value for minimum points. Figure 3 shows these distances for our data, based on a 1000-well lookback and a

[31] We have also run the analysis using alternative values within the same range, and found similar results.

threshold cluster size of 15. (The figure includes two curves: one that includes all wells in the dataset, and one that omits the first 1000 wells in the dataset, which do not have a full 1000-well comparison set.) Reading the sorted kNN distance plot from right to left, the knee of the plot shows where the 15th-nearest-neighbor distance increases dramatically; this corresponds to a value of ϵ that would result in inclusion of the sparsest concentration of points as a cluster (Ester et al., 1996). Points to the left of this threshold, with a higher distance to their 15th-nearest-neighbor, are considered to be experiments; points to the right of the threshold fall into an existing cluster. Looking closely at the graph, there appears to be a substantial increase in the percentage of noise points around a distance of 0.3. Thus, in the results that follow, we set ϵ at 0.3. Results are quite similar when we use other reasonable values of ϵ.

Figure 3: Jaccard Distance to 15th-Nearest Neighbor



Notes: Sorted Jaccard distance indicating the proximity of each well's chemical mix to its 15th-nearest neighbor, within a comparison set of the 1,000 most recent prior wells. The outer (black) curve includes all wells, while the inner (yellow) curve excludes the earliest 1,000 wells ranked by completion date, as they do not have the full set of comparison wells.

With the two necessary parameters in hand, we measure the Jaccard distance (i.e., one

minus the Jaccard similarity) from each well's chemical combination to those of the 1,000 most recent prior wells. We then use this distance matrix and the DBSCAN algorithm to classify the chemical combination used for each well, at the time of its fracture date, into either an existing cluster of similar chemical combinations, or an experimental combination if it is too distant from an existing cluster. Table 6 shows the resulting number of experimental wells, distinguished by time (prior to the mandatory disclosure rule, or after it) and by operator quartile (using the operator fixed effects, θ_f, recovered from estimating Equation 3).

Table 6: Experimentation by Operator Quartile

Period	Operator Quartile	Experiments	Mean Monthly Experiments	Experiments as % of wells	% of Other-Operator Wells Following
Pre-disclosure	1	15	2.1	12.3	0.00
Pre-disclosure	2	16	2.3	6.5	0.10
Pre-disclosure	3	6	0.9	2.0	0.50
Pre-disclosure	4	16	2.3	15.0	0.19
Post-disclosure	1	2	0.1	1.0	0.00
Post-disclosure	2	21	0.6	2.7	0.27
Post-disclosure	3	51	1.4	3.7	0.59
Post-disclosure	4	44	1.2	5.7	1.23

Notes: Number of experimental wells equals number of chemical combinations that are not within a density-based cluster relative to the most recent 1000 wells (starting in October 2011, after 1000 wells have been developed). Operator quartiles are based on the firm fixed-effects, θ_f recovered in Section 4.4 from Equation 3. Average number of experimental wells per month is based on 7 (37) months in the pre (post) disclosure period. A later well "follows" an experiment if its distance to the novel chemical combination is within ϵ.

Although the absolute number of experimental wells increases between the period before mandatory disclosure and the period afterwards (in total, and for all operator quartiles except the lowest), the average rate of experimentation drops (overall, and for all quartiles except the third). This is true when measuring the average incidence of experiments per month, as well as the percentage of wells that are experimental as opposed to being part of an existing cluster (that is, a relatively well-established chemical combination or design). The table also shows that top-quartile operators engage in more experimentation than any other quartile (experimenting with chemical combinations for about 15% of their wells in

the pre-disclosure period, and 5.7% afterwards). Finally, the table shows the percentage of temporally subsequent other-operator wells whose chemical use "follows" the novel chemical combinations that represent experiments at the time of their use. (We define the chemical use of later well j as following that of earlier experimental well i if the Jaccard distance is within ϵ, that is, $0 \leq (1 - s_{ij}) \leq \epsilon$.) Comparing other-operator experiment-following for each quartile, before and after disclosure, shows that other-operator following increases after disclosure for all quartiles (except for experiments conducted by the lowest quartile of operators, which are never followed by other operators).[32] The increase in other-operator following is greatest for experiments conducted by the most productive operators, suggesting that the disclosure laws helped to facilitate the diffusion of innovation conducted by the most productive firms. At the same time, the decrease in experimentation among these firms (and others) is consistent with disclosure eroding incentives to innovate by reducing the ability of firms to exclusively realize the fruits of their investments in innovation.

We further document the operators' experimentation activity before and after mandatory public disclosure by performing a set of difference-in-differences regressions on the percent of experimental wells fractured by operator j in month t. These take the form:

$$\begin{aligned} pctexp_{jt} = \beta_0 + \beta_1 top_j + \beta_2 post_t + \beta_3 top_j * post_t + \beta_4 price_t + \\ \beta_5 newchems_t + \beta_6 recentwells_t + \epsilon_{jt}\,, \end{aligned} \tag{7}$$

where $post_t$ is an indicator that takes the value one for months in which public disclosure of chemicals was mandatory, and top_j is an indicator that takes the value one if operator j is in the top quantile of operators based on the operator fixed effect recovered from Equation 3 (we use top quartile in some specifications, and top two quartiles in others). We also condition on the shale gas price, the number of wells fractured in the most recent quarter (by any operator), and the number of new chemicals introduced by operators in month t. In addition, in some specifications, we add a cubic time trend to account for any unobserved time-varying factors that may affect experimentation. We restrict this analysis to start from October 2011 because, prior to this date, the comparison set for the clustering analysis

[32]The difference between each of the percentages in the right-most column of Table 6 is statistically significant at $p < 0.001$, except for the two zero values for bottom-quartile operators between pre and post disclosure.

contains fewer than 1,000 wells.

We estimate using a linear probability model, despite its well-documented limitations, because of the ease of interpretation and because the marginal effects of greatest interest are those associated with binary predictors. The results are shown in Table 7. The coefficients of greatest interest are those on the dummies for top-producing operators, post-mandatory-disclosure, and their interaction. As in Table 6, we see from columns 3 and 4 that top-quartile operators conduct more experimentation than other operators, and also reduce experimentation following mandatory disclosure (albeit with a p-value of 0.16 in column 3, and 0.18 in column 4, neither statistically significantly different from zero at conventional confidence levels). Columns 1 and 2 show that these results do not hold when *top* is defined more widely to include top-half operators. Together, these results echo the story of Table 6: top-quartile operators conduct more experiments (and, as we saw in Table 6, theirs are the experiments that are more frequently imitated), and decrease experimentation following disclosure.

We leave a more complete model of experimentation for further research, but reiterate here our suggestive evidence of two phenomena. First, the imitation of experiments by other operators increases after mandatory disclosure. Second, the leading operators in terms of overall well productivity — who account for the most experimentation, the experiments that are most worthy of being followed, and thus (by the revealed preference of other firms) the experimentation that is most successful — decrease their experimental innovative activity in the wake of the disclosure requirement. This finding serves as a cautionary note to regulators hoping for a free lunch; it appears in our setting that disclosure disseminated valuable knowledge, but also that it was followed by reduced experimentation.

6 Discussion

Disclosure laws, often motivated by concerns about negative externalities that arise in part from asymmetric information, are increasingly popular in part due to perceived low social costs. Prior research shows that disclosure laws can induce voluntary self-regulation, suggesting low private costs as well. We consider a novel question — whether disclosure laws create pathways for knowledge transmission that were previously inaccessible or overly costly

Table 7: Experimentation Before and After Disclosure

	Percent Experiments			
	(1)	(2)	(3)	(4)
Intercept	−0.007	1.708	−0.021	1.634
	(0.044)	(1.065)	(0.041)	(1.074)
New Chemicals	0.000	0.000	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)
Output Price	−0.015	−0.036**	−0.015*	−0.038***
	(0.009)	(0.014)	(0.009)	(0.014)
Recent Wells	0.000***	0.000	0.000***	−0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Post Disclosure	0.002	0.026	0.021	0.043**
	(0.018)	(0.022)	(0.017)	(0.021)
Top 50% Operator	0.039	0.038		
	(0.035)	(0.035)		
Post * Top 50% Oper.	−0.005	−0.004		
	(0.033)	(0.033)		
Top 25% Operator			0.122**	0.121**
			(0.050)	(0.050)
Post * Top 25% Oper.			−0.074	−0.071
			(0.053)	(0.053)
Cubic Time Trend		X		X
Observations	807	807	807	807
R^2	0.017	0.021	0.036	0.040

Notes: Results are from OLS regressions described in Equation (7). Top 25% and top 50% operators are determined based on the rank-order of the operator fixed effect predicted by Equation (3) (with results in Table 4), corresponding to restricting the lookback set $\mathcal{J}$ to a period of 180 days. Standard errors in parentheses are based on 500 block bootstrap replications; bootstrap samples are drawn such that the number of wells for each operator is the same as in the original data. Results using other lookback sets are comparable, and are available from the authors on request. The unconditional mean of the dependent variable is 0.053. *, **, and *** indicate statistical significance at the 10, 5, and 1 percent levels.

— and assemble a rich and novel dataset that is uniquely well positioned to answer it. In our context, a chemical disclosure law required shale gas operators in Pennsylvania to reveal detailed inputs of their production function. We assemble detailed data on inputs and outputs both before and after disclosure and ask, specifically, (i) if disclosure increases the transmission of information between firms, and (ii) whether the associated erosion of private returns to innovation thereby reduces innovative activity.

Our first finding is that chemical inputs converged following the mandatory public disclosure law, and that the use of chemical mixtures that are more similar to those in high-performing wells is associated with significantly increased well productivity. The evidence suggests that disclosure laws opened a new channel for social learning, and that operators who exploited this channel were able to increase their wells' productivity. This finding supports the argument that public disclosure laws can erode the competitive advantages of top-performing firms.

Our second finding is that experimental configurations declined following disclosure, especially among the top-performing firms, whose experiments are most likely to be imitated by others. This evidence increases the concern over erosion of competitive advantages, and suggests that policymakers ought to take seriously the potential long-run costs of reduced innovation. Nonetheless, the net effect of such disclosure laws on public welfare over the longer term remains an open question: any welfare loss from reduced innovation needs to be weighed against the gains from transparency and dissemination.

References

Agerton, Mark J., "Learning Where to Drill: Drilling Decisions and Geological Quality in the Haynesville Shale," 2018. Unpublished manuscript.

Bennear, Lori S. and Sheila M. Olmstead, "The impacts of the "right to know": Information disclosure and the violation of drinking water standards," *Journal of Environmental Economics and Management*, September 2008, *56* (2), 117–130.

Bollinger, Bryan, Phillip Leslie, and Alan Sorensen, "Calorie posting in chain restaurants," *American Economic Journal: Economic Policy*, 2011, *3* (1), 91–128.

Bruske, J. and G. Sommer, "Intrinsic dimensionality estimation with optimally topology preserving maps," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, May 1998, *20* (5), 572–575.

Cohen, Wesley M., "Fifty Years of Empirical Studies of Innovative Activity and Performance," in Bronwyn H. Hall and Nathan Rosenberg, eds., *Handbook of The Economics of Innovation, Vol. 1*, Elsevier BV, 2010, pp. 129–213. Chapter 4.

_ , Richard R. Nelson, and John P. Walsh, "Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not)," Working Paper 7552, National Bureau of Economic Research February 2000.

Conley, Timothy G and Christopher R Udry, "Learning about a New Technology: Pineapple in Ghana," *American Economic Review*, March 2010, *100* (1), 35–69.

Council, Ground Water Protection and Interstate Oil & Gas Compact Commission, 2015. FracFocus Chemical Disclosure Registry. Online website.

Covert, Thomas, "Experiential and Social Learning in Firms: The Case of Hydraulic Fracturing in the Bakken Shale," February 2015. Unpublished manuscript.

Delmas, Magali, Maria J. Montes-Sancho, and Jay P. Shimshack, "Information Disclosure Policies: Evidence From the Electricity Industry," *Economic Inquiry*, April 2010, *48* (2), 483–498.

Elgin, B., B. Haas, and P. Kuntz, "Fracking Secrets by Thousands Keep U.S. Clueless on Wells," *Bloomberg*, November 2012.

Ester, Martin, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu, "A Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," in "Proceedings of the Second International Conference on Knowledge Discovery and Data Mining" KDD'96 AAAI Press 1996, pp. 226–231.

Fetter, T. Robert, "Fracking, Toxics, and Disclosure," 2018. Unpublished manuscript.

Fung, Archon, Mary Graham, and David Weil, *Full Disclosure: The Perils and Promise of Transparency*, New York: Cambridge University Press, 2007.

Gulbis, J. and R. M. Hodge, "Fracturing Fluid Chemistry and Proppants," in K.G. Nolte and M.J. Economides, eds., *Reservoir Stimulation*, 2000. Chapter 7. Third Edition.

Haas, B., J. Polson, P. Kuntz, and B. Elgin, "Fracking Hazards Obscured in Failure to Disclose Wells," *Bloomberg*, August 2012.

Hamilton, James T., "Pollution as News: Media and Stock Market Reactions to the Toxics Release Inventory Data," *Journal of Environmental Economics and Management*, January 1995, *28* (1), 98–113.

Hausman, Catherine and Ryan Kellogg, "Welfare and Distributional Implications of Shale Gas," April 2015.

Honoré, Bo E and James L Powell, "Pairwise difference estimators of censored and truncated regression models," *Journal of Econometrics*, 1994, *64* (1-2), 241–278.

Jin, G. Z. and P. Leslie, "The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards," *The Quarterly Journal of Economics*, May 2003, *118* (2), 409–451.

Johnsson, Kerstin, "Structures in High-Dimensional Data: Intrinsic Dimension and Cluster Analysis." PhD dissertation, Faculty of Engineering, LTH August 2016.

Kim, Eun-Hee and Thomas P. Lyon, "Strategic environmental disclosure: Evidence from the DOE voluntary greenhouse gas registry," *Journal of Environmental Economics and Management*, May 2011, *61* (3), 311–326.

Kirkpatrick, A. Justin and T. Robert Fetter, "Averting Expenditures and Desirable Goods: Consumer Demand for Bottled Water in the Presence of Fracking," November 2018. Unpublished manuscript.

Konschnik, Kate, "Goal-Oriented Disclosure Design for Shale Oil and Gas Development," *Natural Resources Journal*, 2014, *55* (1), 319.

Konschnik, Katherine and Archana Dayalu, "Hydraulic fracturing chemicals reporting: Analysis of available data and recommendations for policymakers," *Energy Policy*, Jan 2016, *88*, 504–514.

Lemley, Mark A and Carl Shapiro, "Probabilistic Patents," *Journal of Economic Perspectives*, April 2005, *19* (2), 75–98.

Leskovec, Jure, Anand Rajaraman, and Jeffrey D. Ullman, *Mining of Massive Datasets* 2014.

Lucas, Robert E, "On the mechanics of economic development," *Journal of monetary economics*, 1988, *22* (1), 3–42.

Luco, Fernando, "Who benefits from information disclosure? The case of retail gasoline," 2018.

Lyon, Thomas P. and John W. Maxwell, *Corporate Environmentalism and Public Policy*, Cambridge University Press, 2004.

Mokyr, Joel, *The lever of riches: Technological creativity and economic progress*, Oxford University Press, 1992.

Montgomery, Carl, "Fracturing Fluid Components," in "Effective and Sustainable Hydraulic Fracturing," InTech, May 2013.

Muehlenbachs, Lucija, Elisheba Spiller, and Christopher Timmins, "The Housing Market Impacts of Shale Gas Development," *American Economic Review*, December 2015, *105* (12), 3633–3659.

Newell, Richard G., Brian C. Prest, and Ashley Vissing, "Trophy Hunting vs. Manufacturing Energy: The Price-Responsiveness of Shale Gas," Working Paper 22532, National Bureau of Economic Research August 2016.

Pennsylvania Department of Conservation and Natural Resources, "Gap Analysis Report," Technical Report, Well Data Gap Analysis Project November 2010.

Prufer, Jens and Christoph Schottmüller, "Competing with big data," 2017.

Robart, Christopher, Michael Ruegamer, and Alex Yang, "Analysis of U.S. Hydraulic Fracturing Fluid System Trends," Technical Report SPE 163875 2013.

Shellenberger, Michael, Ted Nordhaus, Alex Trembath, and Jesse Jenkins, "Where the Shale Gas Revolution Came From: Government's Role in the Development of Hydraulic Fracturing in Shale," May 2012.

Skytruth, "Fracking Chemical Database," 2013. Online database.

Steck, Andrew L., "Industry Dynamics with Social Learning: Evidence from Hydraulic Fracturing," 2018. Unpublished manuscript.

Stringfellow, William T., Jeremy K. Domen, Mary Kay Camarillo, Whitney L. Sandelin, and Sharon Borglin, "Physical, chemical, and biological characteristics of compounds used in hydraulic fracturing," *Journal of Hazardous Materials*, June 2014, *275*, 37–54.

Teece, David J., "Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy," *Research Policy*, 1986, *15* (6), 285–305.

Wang, Zhongmin and Alan Krupnick, "A Retrospective Review of Shale Gas Development in the United States: What Led to the Boom?," RFF Discussion Paper 2013.

Wrenn, Douglas H., H. Allen Klaiber, and Edward C. Jaenicke, "Unconventional Shale Gas Development, Risk Perceptions, and Averting Behavior: Evidence from Bottled Water Purchases," *Journal of the Association of Environmental and Resource Economists*, December 2016, *3* (4), 779–817.