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

T. Robert Fetter
Duke University

Andrew L. Steck
University of Toronto

Christopher Timmins*
Duke University and NBER

Douglas H. Wrenn
Pennsylvania State University

February 2021

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 firm claims that disclosure requirements expose valuable trade secrets. Our research design takes advantage of a unique regulatory history that allows us to observe 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, Duke University: 2117 Campus Drive, Durham NC 27708; rob.fetter@duke.edu. Andrew Steck: Department of Management, University of Toronto Mississauga and Rotman School of Management, University of Toronto; andrew.steck@utoronto.ca. Christopher Timmins; corresponding author: Department of Economics, Duke University; christopher.timmins@duke.edu. Douglas H. Wrenn: Department of Agricultural Economics, Sociology, and Education, Penn State University; dhw121@psu.edu.

1 Introduction

"Sunlight is said to be the best of disinfectants," wrote US Supreme Court Justice Louis Brandeis in 1914, referring to the idea that requiring firms, governments or other entities to disclose privately held information could reduce externalities or socially undesirable behavior. Regulators have taken this idea to heart in domains from financial markets to food service: mandatory disclosure of private information has become a commonly used regulatory tool in varied settings and with varied aims. An abundant literature in economics (e.g. Jin and Leslie, 2003; Greenstone et al., 2006; Bennear and Olmstead, 2008) explores how mandatory disclosure affects the behaviors of regulated entities that are directly targeted by such rules — most often, negative externalities.

Other effects of information disclosure, such as on technology diffusion and incentives to innovate, have received less attention. These effects could be especially important in the context of an emerging technology, where firms are still refining production methods and conducting experiments to fine-tune inputs. In such a context, requiring firms to reveal private information about production processes or inputs could promote inter-firm learning by reducing other firms' search costs. It could also inhibit investments in innovation by reducing the excess profits that firms can make from successful innovation. Indeed, firms often lobby against mandatory disclosure regulations on the grounds that they undermine the right to trade secrecy and thus constrain innovation. The context of emerging technologies is especially relevant because this is one locus where disclosure rules can be most helpful. When production technologies are still in rapid development and potential risks to the environment or human health are not well documented, regulating through disclosure rather than more prescriptive policy tools can preserve flexibility for innovators to ex-

1

periment, while also providing regulators with more information about potential risks (Fung et al., 2007).

In this paper we study the unintended consequences, with respect to technological diffusion and innovation, of information disclosure laws in hydraulic fracturing for shale gas. Hydraulic fracturing is an emerging technology that has transformed the nature of energy production in the US and the world, even as it has raised a number of concerns regarding impacts on the environment, human health, and local boom-bust cycles (e.g. Hausman and Kellogg, 2015; Muehlenbachs et al., 2015; Wrenn et al., 2016). Public and regulatory concern about these impacts, including potential negative externalities from the use of toxic chemicals in proximity to residential and commercial land uses, led to new legal requirements that firms publicly disclose information about chemicals used in the fracturing process.[1] These laws required firms to reveal previously confidential information about specific inputs in a production process, down to the level of the quantities of individual chemicals used to stimulate hydrocarbon production from individual wells. Prior regulations also required firms to report high-temporal-resolution information about well production. Thus, the implementation of the disclosure laws allowed competitors to observe both inputs and outputs of the production function for individual wells. Our analysis suggests that mandatory information disclosure facilitated technology diffusion through inter-firm learning, but also caused firms to reduce investments in innovation.

Our analysis is set in Pennsylvania, which provides a unique data environment in which to analyze the effects of information disclosure. Pennsylvania, home to the Marcellus Shale, was one of the earliest states to experience a dramatic increase in exploration and production of uncon-

---

[1]In the US, these laws are promulgated at the state level; by 2015, 18 states — including every state with significant fracturing activity — had passed such laws.

ventional shale gas, and continues to experience extensive unconventional development.[2] Pennsylvania's history also features an unusual regulatory episode that facilitates the identification of the effects of disclosure laws. Typically, identifying the effects of information disclosure laws is difficult because researchers cannot observe the disclosed information (in this case, chemical inputs) prior to their implementation. In Pennsylvania, for over a year during the height of the boom, operators were required to disclose detailed information to the regulator, but a combination of accessibility restrictions and regulatory processing delays meant that the information was extremely difficult for either the general public or competitors to access. We recovered pre-public disclosures through a combination of "Right-to-Know" law requests and other methods; by observing this information we can distinguish the effect of the public disclosure law from other simultaneous phenomena, including secular technology change. Later, Pennsylvania altered its reporting platform in such a way as to permit more rapid and essentially costless public observation of chemical and other inputs. We collected this information too, as well as data on output for individual wells during all periods.

To test the effects of disclosure on technological diffusion, we compare detailed, well-level information on inputs, including chemicals, and show that the chemicals used become more similar following the requirement for public disclosure. We demonstrate that this convergence in inputs is not simply an effect of convergence in technology over time, but is rather driven by the visibility of well inputs to other firms. We then test whether the use of chemicals more similar to those in high-performing firms' wells improves productivity, using a measure of well-level input similarity weighted by firm-level productivity. We construct this quality-weighted similarity metric sepa-

---

[2]The Marcellus Shale is the largest natural gas field in the US and one of the largest in the world; its richest deposits underlie northeast Pennsylvania and southwest Pennsylvania (Ikonnikova et al., 2018).

rately for wells that were observable to outsider firms and those that were not (i.e., where any input similarity would be either accidental or driven by independent forces that motivated technological convergence). We find that firms drilled more productive wells when (i) their chemical solutions were more similar to those solutions used by more productive firms *and* (ii) those input sets were visible in the public repository, suggesting that disclosure caused inter-firm learning.

These results raise a question – what is it about visibility that translates the benefits of a productive chemical combination to a copying operator? In a series of additional tests, we explore a possible mechanism. In particular, we focus on the additional information that comes with visibility – in particular, the location in which the copied well was drilled. In Pennsylvania, geography tells the operator a lot about the quality of the shale; in the northeast region of the state, the shale is harder and produces dry gas, while in the southwest region, the shale is softer and also yields liquids. Knowing the location of the well in the visible information regime allows the copying firm to infer the qualities of the shale on which the well was drilled. This additional location-specific information may influence the way in which the chemical combination is implemented yielding greater success. Indeed, we find that when a copying well is drilled either in the same or different region of Pennsylvania, the productivity of the copied wells has no impact on its productivity. When the copied wells are visible, their productivity only matters when the copying operator knows that they were in the same Pennsylvania region.

We next evaluate whether mandated disclosure undermines the incentives for firms to invest in research and development for future innovations. We locate each well in high-dimensional space based on the chemicals used in the fracture job, and identify experimental wells based on their distance from clusters of prior wells in this chemical input space. We demonstrate that firms conduct fewer experiments after mandatory disclosure, and that this effect is particularly pronounced

4

among the most productive operators. We also show that firms are most likely to copy experiments conducted by their most productive peers, and that this copying of experiments is much more common after mandatory disclosure. Taken together, the evidence suggests that the most important experimentation — i.e., that of the most productive firms, which are also the most likely to be copied by peers — falls after mandatory disclosure.

Our paper makes three primary contributions. Our first contribution is to the literature on the relationship between secrecy and innovation. Most of this literature focuses on patent policies, which allow inventors a temporary period of monopoly rights in exchange for public disclosure of their invention. In our empirical setting firms do not apply for patents for chemical mixtures, evidently preferring to preserve monopoly rights by maintaining secrecy over their inventions.[3] However, required public disclosure of a hitherto secret chemical formulation is comparable to reducing the length of a patent.[4] Despite theoretical expositions as early as Nordhaus (1969), the empirical literature on causal effects of reduced patent length on innovative activity is somewhat limited, due to the difficulty of finding settings that permit causal identification. However, prominent papers that have been able to identify causal effects suggest that patent length does not affect firms' investments in research (e.g., Sakakibara and Branstetter, 2001; Lerner, 2009) — a counterintuitive result that Budish et al. (2016) and Williams (2017) suggest is driven by peculiar features of those papers' empirical context, with limited applicability to the general case. These features do not arise in our empirical context, and our study indeed supports the more intuitive conclusion that a shorter duration of monopoly rights decreases investments in innovation.

---

[3]A likely explanation is that the protection of monopoly rights that patents provide is always incomplete, and especially so when innovations can be adjusted in small ways that qualify as novel and non-encroaching (Lemley and Shapiro, 2005). See also Cohen et al. (2000), who document the importance that firms place on secrecy for appropriating returns to innovation.

[4]We appreciate an anonymous reviewer for elucidating this point.

Second, we expand the literatures on both interfirm learning, and the empirical effects of information disclosure regulations, by bringing them together. The literature on interfirm learning is relatively small but recent work has analyzed themes such as how managerial knowledge spillovers and peer-to-peer learning through business networks can affect performance (Cai and Szeidl, 2018), including through particularly successful peers who can serve as role models (Dalton et al., 2018). Yet much remains to learn about how firms learn from one another; in particular, we are not aware of literature on disclosure regulations as a mechanism. Meanwhile, an extensive literature on the effects of information disclosure addresses the effect of price or quality disclosure on market structure and competition (e.g., Jin and Leslie, 2003; Bollinger et al., 2011; Luco, 2018), on social or environmental outcomes (e.g., Greenstone et al., 2006; Fung et al., 2007; Bennear and Olmstead, 2008), and on investor behavior (e.g. Hamilton, 1995). Our research bridges these streams: Our analysis is the first, to our knowledge, to study the effects of disclosure regulations on inter-firm knowledge transfer. We find strong evidence that such regulations can enable interfirm knowledge spillovers that would not have occurred otherwise.[5]

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 (2019) and Agerton (2019), have documented the rise of hydraulic fracturing and how learning affects the decisions of firms. 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.[6]

We stop short of attempting to analyze the net effects of the disclosure laws on social welfare,

---

[5]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.

[6]We are aware of one other working paper, still in development, that addresses how chemicals affect well productivity (Fitzgerald and Mason, 2020).

in part because of the difficulty inherent in measuring the value of individual experiments (trials) in a setting where knowledge is advanced by trial and error.[7] However, our results suggest that the laws reduced investment in the discovery of new technologies — that is, the expansion of the technological frontier as the result of costly trial-and-error — even as they increased the diffusion of better technologies through the industry. Mokyr (1992) points out that 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* societal progress.

The paper proceeds as follows. In Section 2, we provide context and background for our study, including background on hydraulic fracturing for shale gas, the role of information disclosure regulations in this industry and others, and related literature on how public policy affects the incentives for firms to invest in innovation. 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. Section 6 addresses a possible alternative explanation by testing the role of contractors in facilitating information flows. Section 7 offers concluding remarks.

---

[7]We also do not address the effects of the disclosure laws on their intended target, the use of toxic chemicals; that question is addressed in Fetter (2019).

2 Background

2.1 Hydraulic Fracturing for Shale Gas

Shale gas grew from 5% of total US dry gas supply in 2004 to 56% in 2015, and forecasts of international energy production indicate shale gas will be a principal driver of world natural gas production for the foreseeable future.[8] The ability to profitably recover hydrocarbons from shale is largely based on advances in four key technology domains: 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 decades, spurred by both private and public investments in research and development. Technological advances since the 1970s 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 fluid and proppant used. The 1990s saw advances in directional drilling, alongside massive hydraulic fracturing that was enabled by the use of "slick water" fracturing fluid in place of gels. Fracturing fluid continues to be refined in multiple dimensions to maximize output and minimize costs.

Operators use a wide array of chemicals in hydraulic fracturing fluid to enhance the productivity of the primary inputs, water and sand. Chemical additives help to open fractures in the rock, transport proppant along the length of the fracture, lower viscosity so as to allow faster pumping at higher pressure, minimize fluid loss into the face of the geologic formation, reduce chemical corrosion or bacterial growth that might threaten the integrity of metal casings, facilitate breakup

---

[8]http://www.eia.gov/conference/2015/pdf/presentations/staub.pdf.

of other chemicals post-fracture, and serve other purposes (Gulbis and Hodge, 2000; Montgomery, 2013). In short, fracturing fluid is a complex mixture, with design specifications that differ depending on the target geologic formation and well depth, and 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 choice of chemicals may have dramatic effects on the overall cost and productivity of a well.[9]

As of 2014, the total estimated recovery for shale oil wells was on the order of 5%, compared to 50% for conventional oil wells.[10] 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 seeking to advance productivity of shale fractures (Gulbis and Hodge, 2000; Montgomery, 2013; Robart et al., 2013).

In some cases, the quest for superior fracturing fluids has led engineers to use chemicals known to have adverse effects on human health or ecosystems (Stringfellow et al., 2014; Fetter, 2019). From the start, public concern about shale gas development was driven partly by the possibility that toxics in fluid — although they represent a small percentage of overall volume[11] — might leak into ground or surface water. Public concern about risk was accentuated by the often close proximity of well-pads to residential and other non-industrial land uses, and by a few high-profile

---

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

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

[11]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 fracturing fluid would represent 50,000 gallons.

incidents of water pollution.[12] 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 Mandatory Disclosure Regulations

The environmental and health concerns over toxics used in fracturing prompted calls for laws that would either ban hydraulic fracturing, prohibit the use of toxics, or force companies to disclose the identity and quantity of chemicals used in individual well fractures. While a handful of countries and US states banned hydraulic fracturing outright, most jurisdictions chose mandatory disclosure as a policy tool that balanced the concerns of regulators, firms, and the public. In this sense, information disclosure was a "pragmatic compromise" (Fung et al., 2007) that allowed continued production and technological development while also providing for regulators' and the public's right to know about proximate risks.

Pennsylvania was the third US state (after Wyoming and Arkansas) to mandate disclosure of fracturing chemicals. All state-level chemical disclosure laws are nearly identical in terms of what must be reported: ingredient name, chemical abstract service (CAS) number, a measure of concentration in fluid (typically the maximum concentration used in any fracturing stage), trade name if any, and supplier name. As of 2016, five of the eighteen states with disclosure laws, including many that passed the earliest such laws, require operators to report to a state regulatory agency or commission; six states require operators to upload forms to FracFocus, an online database created by a multi-state commission in partnership with a non-profit organization (Council and Commission, 2015); and seven states allow firms to choose between FracFocus and the state regulator. Firms

---

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

must also provide well locations and characteristics including vertical depth and water volume.[13]

Disclosure of chemical information increases the ability of local landowners to bargain with operators, since they can test for specific pollutants if they are concerned about infiltration into water sources. Disclosure can also help establish liability for contamination, both by allowing physical contaminant tracing and by allowing comparisons of chemical usage across sites. If the information made available extends over a period of time, then disclosure can facilitate monitoring of environmental releases, exposures, or health impacts over time. Information can also help emergency personnel respond to a spill or accidental exposure. While many of these benefits could be achieved equally well by disclosure only to a competent regulator (and not the public), disclosure satisfies the public's "right-to-know" about possible release of or exposure to hazardous materials, which in the US is codified in the Emergency Planning and Community Right-to-Know Act (EPCRA).[14]

In addition to these benefits, mandatory disclosure has been shown to lead to "self-regulation" (Lyon and Maxwell, 2004) as firms (or other regulated entities) change their behavior so as to reduce the risk or harm that led to regulatory concern. For instance, Jin and Leslie (2003) show that restaurants improve hygiene practices when required to prominently display cards with letter grades in the front window; Greenstone et al. (2006) demonstrate that mandatory financial disclosure leads firm managers to focus on maximizing shareholder value; and Bennear and Olmstead (2008) find that drinking water quality improves when municipal providers must mail violation reports directly to consumers. In the setting of hydraulic fracturing, Fetter (2019) shows that dis-

---

[13]State laws do allow exemptions for disclosure of additives that firms consider confidential business information; operators wishing to invoke this provision must file separate exemptions for individual chemicals but must also (in most cases) report the quantity used. Some states require operators to identify the chemical family to which each proprietary substance belongs.

[14]Note that the shale gas industry is currently exempt from some provisions of this Act.

closure laws similar to those we study caused operators to reduce the use of toxic chemicals.

The question of whether mandatory information disclosure facilitates technological diffusion is less well studied — in fact we are aware of no prior study on this topic. This may be because mandatory disclosure rules often target performance aspects of firms (e.g., restaurant hygiene, drinking water quality violations, corporate financial reports) that are unlikely to convey important details about production processes. However, in settings where the information sought relates directly to the production process, disclosure could facilitate diffusion. This appears to be the case in shale gas, and may also be relevant for other contexts in which companies must report chemical use or releases, such as the European Pollutant Release and Transfer Registry or the US Toxics Release Inventory. To the extent that regulators consider information disclosure as a compromise policy that carries little political or economic risk, our study sounds a note of caution about the potential for unintended consequences related to both technology diffusion and innovation.

2.3 Innovation, Secrecy, and Disclosure

Since innovation and technological progress are at the root of economic prosperity, there is a compelling public interest in ensuring that would-be innovators can secure returns to costly investments in research and development (Mokyr, 1992). As a result, there is an extensive literature on how firms realize these returns and how public policy can facilitate innovation (Williams, 2017). The observation that each successful invention indicates pathways others can follow and stimulates the development of related ideas led to the "grand bargain" of patent systems, which provide inventors with a temporary period of monopoly rights in exchange for publicly disclosing details of their innovation. Nordhaus (1969) identifies patent length as a key policy variable that balances the

deadweight loss arising from an extended period of monopoly power and the incentives for firms to invest in innovation.

Although much of the economic literature on innovation focuses on patent policy and, indeed, uses patents as the key measure of innovation outcomes, in practice many firms use secrecy as a primary tool to realize returns from their investments, and evidently consider secrecy more effective. For instance, Cohen et al. (2000) surveyed 3,240 R&D laboratories in the US manufacturing sector and asked laboratory managers to judge the effectiveness of six appropriability mechanisms (secrecy, patents, other legal mechanisms, lead time, complementary sales and service, and complementary manufacturing facilities and knowledge). With few exceptions, managers in every one of the 34 industry codes surveyed (including chemicals and petroleum) reported that secrecy was more effective than any other mechanism, including patents, for protecting both product and process innovations.[15]

In the sense that mandating disclosure of a trade secret is comparable to shortening the length of a patent, it results in the benefits and costs that Nordhaus (1969) wrote about: social benefits arising from the reduced duration of monopoly power, and social costs from the weakened incentives to invest in innovation. One key difference is that patent applicants must reveal details of their innovation in the application, thus allowing others to see some of the key ideas that paved the way for a successful invention, whereas those who employ trade secrecy are under no such obligation. As such, the forced disclosure of trade secrets can also promote faster take-up of technological improvements — not just in terms of adoption of the specific innovation but also in terms of revealing a successful formulation to other would-be inventors, and thus stimulating further in-

---

[15]Cohen et al. (2000) asked managers to categorize the percentage of their innovations for which they considered each mechanism an effective tool for appropriating returns. In the petroleum industry, for instance, managers identified secrecy as an effective tool for 57% of process innovations, and patents as an effective tool for 37%.

novation in that direction. Of course, whether the net effect on innovation is positive or negative is an empirical question.

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. At the time of a given fracture operation, each operator has access to input and output data from its own prior wells as well as data from other firms on well locations, output, and — in some cases, and subject to a delay that varies with the extent of disclosure requirements — inputs. 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, the operator name, and, if available, the name of the contractor firm that helped to perform the frac job. We compile this dataset for all unconventional wells in Pennsylvania.

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.[16] Gas

---

[16]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.

production data comes from Enverus DrillingInfo, a national 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 fracturing contractor), well location, and various details about the perforation and stimulation process. Effective in February 2011, operators also had to submit information on chemicals used in stimulation, including the chemical name, identification number, and concentration in fracturing fluid.[17]

Under the February 2011 law, operators had to submit information about chemicals to DEP either attached to 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, sometimes 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. All disclosures to FracFocus were voluntary at the time: Wyoming and Arkansas had also passed chemical disclosure laws as of February 2011, but both required firms to report to a state registry. 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.

Information that operators submitted to FracFocus was more readily observable by competitors and the public: from the start, the registry allowed users to download individual PDF files

---

[17]Although the chemical disclosure requirement was new in 2011, operators have been required to submit Well Completion Reports since 1989; see http://www.pacode.com/secure/data/025/chapter78/s78.122.html.

with fracturing fluid chemical composition, well location, operator name, water volume, and other items.[18] 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 Information Network (EDWIN), in which some chemical disclosure forms were available as scanned PDF documents.[19] 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 a contemporaneous internal assessment (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.

---

[18]Initially FracFocus was designed to inhibit bulk downloads or scraping, though 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).

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

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 the designers of each frac job had complete information about 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 $j \neq i$, 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 received 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.[20] 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.[21] We drop an additional 282 wells that have nonsensical completion dates (the

---

[20]To ensure the analysis is not driven by outliers, we winsorize per-foot initial gas production at the 99th percentile.
[21]This omission of perforated interval appears to be out of line with basic reporting requirements.

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.[22]

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). We also observe gas-in-place, an estimate of underlying resource quality, from a University of Texas Bureau of Economic Geology study that uses shale thickness, total organic content, and other geological variables (Ikonnikova et al., 2018).[23]

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 organization that scraped the "version 1" FracFocus website Skytruth (2013). The DEP reports are substantially less accessible, as they are not amenable to optical character recognition. The DEP used 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

---

[22]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.

[23]As Agerton (2019) points out, because gas-in-place is estimated from geological data rather than well production, it is not affected by firms' drilling decisions and thus provides an exogenous measure of resource quality. Furthermore, since it is based on the type of information that all firms should have access to, we assume that it is in firms' information sets before they fracture new wells.

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.

Unfortunately, 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; including proppant and stages increases that to 0.4881.[24] We concluded that to include these variables would add only incrementally to explanatory power, but result in a substantial cut in sample size and potentially introduce issues of sample selection.

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. A more detailed look at upload

---

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

times is available in Figure 3.

[Figure 3 about here]

Table 4 provides a summary of information about the wells for which we observe chemical 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.)[25] 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 5 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 4 about here]

[Table 5 about here]

---

[25]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.

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 chemicals to be declared proprietary, create conditions by which firms could learn from each other? Second, did the learning enabled by disclosure 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.

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 that suggests firms are copying from other operators' 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 predicts higher productivity. Together, these findings demonstrate that firms are learning through mandatory chemical disclosure.

4.1 Well-to-Well Similarity Measures

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 $\mathscr{C}$ denote the set of all possible chemicals, and let $x_{ic} \in [0,1]$ be the concentration of chemical $c \in \mathscr{C}$ in well i. The sum of all such concentrations is therefore $\sum_{\mathscr{C}} x_{ic} = 1$. 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:

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

- $B_{ij} \equiv \dfrac{\sum_{\mathscr{C}} x_{jc} y_{ic}}{\sum_{\mathscr{C}} x_{jc}} = A_{ji}$ is the converse.

We use the Jaccard index, under which the pairwise similarity between wells i and j can be defined as:[26]

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

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 4.3. (A value $s_{ij} = 1$ implies that the hydraulic fracturing fluids used in wells i and j are indistin-

---

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

guishable; $i = j \Rightarrow s_{ij} = 1$. 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 would be a concern in our setting.

4.2   Disclosure and Similarities

We calculate the similarity index for each (i, j) well-pair in our data. Our chemical data on 4,862 wells gives us 11.8 million such pairs. Figure 4 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.

[Figure 4 about here]

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:

$$s_{ij} = \beta_0 + \beta_1 DiffOper_{ij} + \beta_2 DiffOper_{ij} * visible_{ij} + \phi_{ij} + \varepsilon_{ij} \tag{1}$$

23

where s_{ij} is the Jaccard similarity index defined above, $DiffOper_{ij}$ 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 reiterating that the visibility of the earlier well's chemicals cannot be taken for granted, even if there is a long lag between the wells: Section 3 discusses the high variance of lags between completions and chemical disclosures. We therefore use the upload date to determine visibility rather than simply the date of the completion. We also include fixed effects for the first well's operator, the second well's operator, and depending on the specification, 5th order polynomials for the second well's completion date, the days between the two wells, and the geographic distance between the two wells are represented by the ϕ_{ij} term. The results of this regression, estimated separately with and without the fixed effects, are shown in Table 6.[27]

For the purposes of inference, we note that the usual OLS standard error assumptions are not satisfied for Equation (1) — given that the regression takes place at the well-pair level, we must account for the possible correlation of the error term across well-pairs that share a particular well. We there calculate our standard errors in Table 6 as the result of a 1,000 iteration bootstrap routine. In each iteration we draw a full bootstrap sample with replacement, and the standard errors reported are the standard deviations of the calculated bootstrap coefficients.[28]

[Table 6 about here]

A few things are worth noting from Table 6. First, all columns show a significantly negative

---

[27]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.

[28]We are grateful for a discussant pointing out that another approach would be to use the asymptotics developed in Honoré and Powell (1994). They consider a more general case, but the implementation of their method is impractical in our setting where we have so many wells.

estimate of the coefficient on $DiffOper_{ij}$; this indicates that well-pairs with different operators tend to use less-similar chemical mixes, providing statistical confirmation of the trends visible in Figure 4. Second, the coefficient on the interaction of $visible_{ij}$ with $DiffOper_{ij}$ 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. As the various columns show, these results are robust to the inclusion of a variety of temporal and spatial controls that might independently affect the similarity of the inputs for a pair of wells. Incorporating 5th-degree polynomial functions of the completion date of the second well, the number of days between the completion dates of the two wells, and the geographical distance that separates the well-pair does not alter the sign nor the approximate magnitude of the core results.

4.3   Effect of Information on Productivity

Next, we estimate a model in an attempt to understand the relationship between disclosure and information, an operator's 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.

4.3.1 Model

We begin by defining the regression equation of interest:

$$\ln gpf_i = \beta X_i + \theta_{f(i)} + \alpha_{OBS} \ln(1 + QS_i^{OBS}) + \alpha_{UNOBS} \ln(1 + QS_i^{UNOBS}) + g(t) + \varepsilon_i \qquad (2)$$

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 appears to have throttled production, the region where the well is located, and the natural gas price at the time of the frac.[29] The term $\theta_{f(i)}$ is a productivity fixed effect for the firm $f(i)$ that drilled well i.[30] Finally, we include a time control term in $g(t)$ — in the main specification this takes the form of a linear time trend, and we present results under alternative controls in Appendix B.

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.[31]

---

[29]Throttling production refers to intentionally restricting hydrocarbon flow, which is a process operators sometimes engage in for geological, economic, or logistical 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.

[30]An anonymous referee suggested the idea of using a measure of well-level productivity rather than firm-level productivity, as we have done here. Unfortunately, we are not familiar with any econometric method that allows residuals (i.e. well productivy) to be estimated and used simultaneously.

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

These quantities are defined:

$$QS_i^{OBS} \equiv \sum_{f \in \mathscr{F} \backslash f(i)} \left(\theta_f \frac{1}{|\mathscr{J}_f^{OBS}|} \sum_{j \in \mathscr{J}_f^{OBS}} s_{ij} \right), \tag{3}$$

$$QS_i^{UNOBS} \equiv \sum_{f \in \mathscr{F} \backslash f(i)} \left(\theta_f \frac{1}{|\mathscr{J}_f^{UNOBS}|} \sum_{j \in \mathscr{J}_f^{UNOBS}} s_{ij} \right). \tag{4}$$

In these equations, $\mathscr{F}$ indicates the set of firms in the data, $\mathscr{J}_f$ 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 4.1. 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 $\mathscr{J}_f^{OBS}$, and those whose were not fall into $\mathscr{J}_f^{UNOBS}$. 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 (2) thus accounts both for how similar well i is to its comparison wells $\mathscr{J}$, and also how successful those wells are (as measured by their firm fixed effects, θ). QS_i^{UNOBS} is an analogous quality-similarity index for the recently drilled but *unobserved* wells.

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 and shown in Figure 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.

If firms realize private information during the fracturing operation that (i) alters their choices of inputs, including chemicals, and (ii) is unobservable to us as econometricians, this could threaten

identification of the key quantities of interest. This concern appears to be nonexistent in our setting, based on our conversations with several geologists and petroleum engineers with extensive experience in designing and implementing frac operations. This is because (i) design decisions about fracturing configurations, including chemical mixtures, are made weeks to months in advance of the actual stimulation operation, and (ii) due to logistical limitations and the difficulty of coordinating the already-complex fracturing process, any last-minute changes can be only marginal. See also Covert (2015), who finds the same in his study of fracturing operations in North Dakota.

4.3.2 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 (2) 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 (2) 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 (2)using non-linear least squares (NLLS), and present results in Table 7.

[Table 7 about here]

The three columns of Table 7 consider different possible look-back restrictions on $\mathscr{J}$, e.g. restricting wells j to have been fractured within 180 days of well i's fracture. From the first column it can be seen that increasing fluid volume by 100% predicts a 21.6% increase in productivity, and throttling the well predicts a 43% decrease ($e^{-0.569} - 1$) 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 $e^{0.390} - 1 = 47.7\%$ more productive per foot, and a \$1 increase in the Henry Hub price of natural gas predicts an 8.1% increase in productivity. Conveniently, the coefficient on

28

QS_i^{OBS}, estimated to be statistically significant at the 1% level, can be understood as an elasticity: a 100% increase in $(1 + QS_i^{OBS})$ implies a 7.6% increase in productivity per foot.[32] In turn, the coefficient corresponding to QS_i^{UNOBS} has a much smaller point estimate, and is not statistically distinguishable from zero.

The second and third columns present analogous results for different specifications of the $\mathscr{I}$ used in constructing the QS_i: column 2 restricts the consideration to wells completed within 360 days, and column 3 uses 720 days. 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% level, while $\hat{\alpha}_{UNOBS}$ remains statistically indistinguishable from zero. The declining point estimates of $\hat{\alpha}_{OBS}$ with increasing length of look-back period reflect a form of attenuation bias: similarity to more recent wells (conditional on QS) explains more of the variation in productivity, but extending the look-back period puts equal weight on the similarity to earlier wells that have less influence on current productivity. More simply, what matters most for productivity gains is copying the recent wells of high-quality operators, and using a longer look-back period dilutes the information content of how QS^{OBS} affects productivity.

Table 8 provides another reference for interpreting these estimates. Each entry in Table 8 represents a fitted value $\hat{gpf}$ calculated using the estimates of Equation (2) presented in Table 7. Entries in the top half of the table use the operator and location from the first two columns to set θ_f and *northeast* and set all other variables to sample means, with the exception of QS^{OBS}. Instead, QS^{OBS} is set at the 25th or 75th empirical percentile as suggested by the column headings. Results are presented for the 180-day, 360-day, and 720-day specifications. The bottom half of the

---

[32]Some caution is warranted when interpreting these results, as it is not necessarily straightforward 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.

table performs a similar exercise, but keeps QS^{OBS} at the sample mean and varies fracturing fluid volumes at the 25th or 75th percentiles.

As an example, a well completed by Hunt in the northeast portion of Pennsylvania, that is average in every other way, would move from a predicted production of 403 to 520 thousand cubic feet of gas (MCF) per linear foot of fracture in the first 6 months if its QS^{OBS} were increased from the 25th to the 75th empirical percentile, using the estimates from the 180-day specification. If the same well had an average QS^{OBS} but its fluid volume moved from the 25th to 75th empirical percentile, we would expect to see increased production in the first 6 months from 442 to 512 MCF per foot.

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.

[Table 8 about here]

The results above incorporate information about geographic variation by using a fixed effect for well location, specifically, whether the well is located in the highly productive northeastern region of the Marcellus. To control more finely for geographic or geologic variation, we also consider a similar regression that conditions on the amount of gas in place. These results, presented in Appendix Table B.1, do not substantially differ from those in Table 7, although the magnitude of the coefficient on QS^{OBS} is somewhat lower.

An alternative explanation that could generate some of the same results we see here would be unobserved shocks, common across firms, that directly affect both input similarities s_{ij} and output gpf_i. For instance, there might be secular technological convergence that coincidentally

occurred around the same time as the disclosure regulation, such as if the industry as a whole is evolving toward an equilibrium in the set of chemical mixtures considered most effective. If that evolution occurs concurrently with the disclosure regulation, this could generate the pattern of input convergence we see in Section 4.2. Furthermore, if the chemical-input convergence also happens simultaneously with convergence in other production technology choices, this might also explain part of the increase in productivity that we attribute to firms' better copying of higher-quality chemical mixtures when they are visible (i.e., $\hat{\alpha}_{OBS}$).

It is not clear, in this alternative version of events, why $\hat{\alpha}_{OBS}$ would differ significantly from zero whereas $\hat{\alpha}_{UNOBS}$ would not. However, to help rule out whether unobserved shocks simultaneously affect input similarities and output, we estimate two other versions of Equation (2): one with a year fixed effect, and one restricted to wells fractured within a narrower time window around the disclosure law effective date, so as to reduce the effects of secular technological change or convergence.[33] The results are shown in Appendix Tables B.2 and B.3 respectively. The key results do not change in either specification — $\hat{\alpha}_{OBS}$ remains positive and significant (and the magnitude actually increases from that shown in Table 7), while $\hat{\alpha}_{UNOBS}$ remains smaller in magnitude and not distinguishable from zero. We conclude that mandatory chemical disclosure affected inter-firm knowledge transfer even factoring in other phenomena that may have facilitated technological convergence at this time.

It is worth noting that a sophisticated, large, and forward-looking firm might choose to improve its long-term 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.

---

[33]We are grateful to an anonymous reviewer for suggesting the latter test.

4.3.3 Productivity, Visibility, and Proximity

THIS IS CHRIS' LANGUAGE.

(I BELIEVE)

With an eye towards possible mechanisms underlying the role of visibility, We next explore one dimension over which visibility might convey information important for the successful adoption of another firm's fracking technology – geography. Specifically, visibility adds additional information about the location of a well, and location conveys an additional information about shale quality. In the case of Pennsylvania, the quality of the shale differs systematically with region – in the northeast part of the state, the Marcellus shale is hard and produces dry gas, while in the southwest the shale is softer and also yields a variety of liquids. [WE MIGHT BE ABLE TO IMPROVE UPON THIS DESCRIPTION OF THE SHALE.] Visibility therefore provides additional information about the type of shale that a fracking combination was used on, and the recipient of that additional information can then optimize the application of that combination in ways that an operator who inadvertently chances on the same combination cannot. I.e., we would expect an operator to make better use of a successful technology if they knew that they were applying it to the same type of shale. In a set of additional specifications, we test this hypothesis with a set of difference-in-differences regressions that look both before and after chemical combinations become visible, *and* we control for whether the combination was applied to the same type of shale.

[INSERT TABLE HERE.]

We see that, in the absence of visibility, copying a successful well conveys no advantage regardless of whether or not the well was drilled in the same region – i.e., similar shale conveys no advantage if the copying operator does not know that it is similar. When visible, copying a suc-

cessful well drilled in the *other* region of Pennsylvania also carries no advantage, suggesting that drilling on a similar shale is important to successful copying. The *only* term for which we find an advantage from copying a successful well is when visibility and similar region are combined.

<p align="center">**THIS LANGUAGE IS PRESENT FOR BOTH OPTIONS A AND B:**</p>

In this subsection, we explore the apparent relationship between well productivity, chemical similarity and visibility, and geographical proximity between well-pairs.

Let us begin by restating our interpretation of the results in Table 7: the combination of a positive and significant estimate for $\hat{\alpha}_{OBS}$ with a statistically insignificant estimate of $\hat{\alpha}_{UNOBS}$ is consistent with mandatory chemical disclosures allowing for the imitation of the inputs of the most productive firms, and that imitation leading to higher productivity at the well-level. We do not claim a causal interpretation, but the high degree of variation and granularity in chemical choices and visibility timing argue against correlated unobserved shocks.

The biggest puzzle that remains is: why does the observed productivity increase for wells that use chemical inputs more similar to those used by high-productivity firms *depend on the fact that those chemicals were disclosed*?[34] That is, any sensible model of well productivity should be invariant to disclosures made to the regulator: the only things that matter are the underlying geology, the well inputs, and their interactions. If engineers use inputs *X* at location *Y*, we expect the well would experience the same production regardless of whether those engineers were aware that another firm had successfully used inputs similar to *X* on another well.

As a solution to this puzzle, we observe that the while the input choices we observe for each well are extensive, they are not exhaustive. If some of the unobserved choices that are made depend on a well's specific location or geology, then it is plausible that the visibility of well inputs

---

[34]We thank an anonymous referee for clarifying this puzzle.

adds more value for well-pairs that are located closer together. In other words, the disclosure of inputs, in combination with prior success, conveys a *tacit* knowledge [ROB: INSERT CITATION AND/OR FOOTNOTE ABOUT TACIT KNOWLEDGE IN STRATEGY CONTEXT] that is unobserved (indeed unobservable) to the econometrician.

<div align="center">

THIS IS OPTION A:

USING WHETHER THE TWO WELLS ARE WITHIN 35KM.

NB THIS OPTION CANNOT USE ADDITIONAL NLLS RESULTS.

</div>

In support of this proposal, we observe a stark shift in the well-pair similarity distribution between those well-pairs whose first well's inputs are invisible vs. visible when the geographic distance between the two wells is small.[35] This shift is visible in the distributions plotted in Figure 1.

<div align="center">

FIGURE 1: SIMILARITIES BY VISIBILITY AND DISTANCE
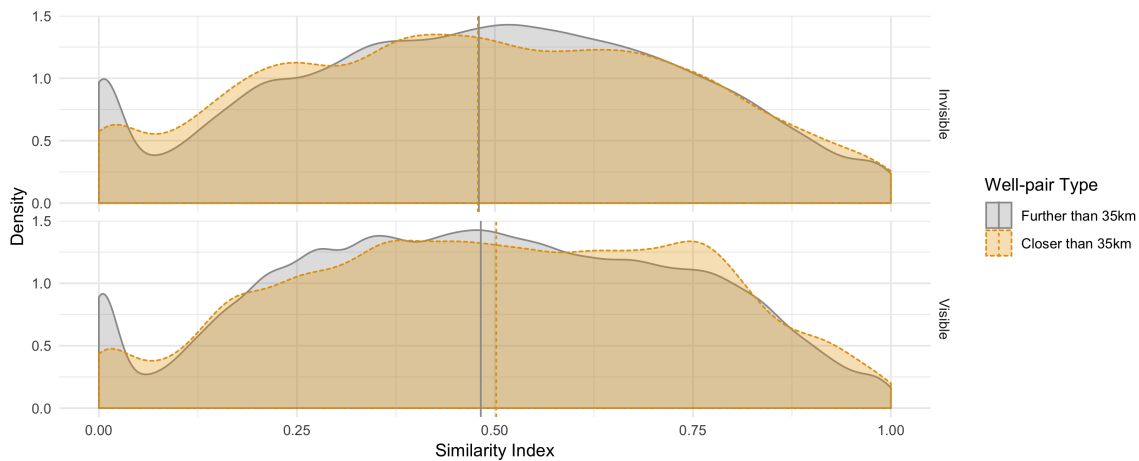
</div>

Notes: The curves plot estimated densities for well-pair similarities s_{ij}, conditional on whether the inputs from the first well were visible when the second well was stimulated and whether the distance between the two wells is less than 35km. The vertical lines indicate the distribution means. All distributions include only those well-pairs where each well is completed by a different operator.

Both the top and bottom panels of Figure 1 plot two distributions of well-pair similarities: the

---

[35]We use an arbitrary 35km cutoff here, but similar results hold at other cutoffs or, e.g., using same-county status as a proxy for distance.

grey/solid distribution is comprised of those well-pairs where the two wells are located further than 35km from each other; the orange/dashed distribution includes well-pairs where the two wells are within 35km. The top panel plots those well-pairs where the inputs of the first well are *invisible* when the second well is stimulated, while in the bottom panel the first well's inputs are *visible* when the second well is stimulated. For all four distributions the mean is plotted as a vertical line. Finally, it is worth emphasizing that none of the well-pairs plotted in Figure 1 have the same operator for both wells.

Looking at Figure 1, a major difference emerges. When we compare how similarities shift for within-35km well-pairs when inputs from the first well become visible, we see a noticeable rightward shift in the distribution. In contrast when we examine how visibility affects similarities for well-pairs further than 35km from each other, there is no effect at the mean and does not appear to be any effect along the rest of the distribution.

We move from a graphical to a statistical examination of this heterogeneity in Table 1. We perform two separate Wilcoxon Rank Sum tests with one-sided alternative hypotheses. The two cases are for 'invisible' or 'visible' well-pairs, depending on whether the first well's inputs have been disclosed. In each case the alternative hypothesis is that the CDF of the well-pair similarities when the two wells are located within 35km of each other lies below the CDF of the the similarities when the wells are further apart.

The first column of Table 1 shows the results from the test when we consider 'invisible' well-pairs. As can be seen, the estimated location difference is statistically indistinguishable from zero. In contrast, when we consider 'visible' well-pairs, the location difference is calculated to be positive and significant. Taken together, we can see that the visibility of the first well's inputs seems to have little (unconditional) distributional effect when the wells are located further apart, but a statis-

	Invisible	Visible
Difference in location	-0.00241	0.02009
95% Confidence Interval	[-0.00315, Infinity]	[0.01961, Infinity]
W (test statistic)	3.65e11	3.53e12
p-value	1	<2.2e-16

Notes: This table presents the statistical results of two Wilcoxon Rank Sum tests with a one-sided alternative hypothesis. The alternative hypothesis is that the CDF of the < 35km distribution lies below that of the > 35km distribution for well-pairs where the first well's inputs are invisible and visible respectively.

tically significant effect when the wells are closer than 35km. This finding lends statistical support to the above conjecture that the disclosures convey some tacit information above and beyond the list of ingredients; and this sort of tacit information could explain the difference in the coefficient estimates on $\hat{\alpha}_{UNOBS}$ and $\hat{\alpha}_{OBS}$ in Table 7.

THIS IS OPTION B:

USING WHETHER THE TWO WELLS ARE IN THE SAME REGION.

NB THIS OPTION CAN USE ADDITIONAL NLLS RESULTS.

TABLE 2: WELL-PAIR SIMILARITIES BY VISIBILITY AND REGION

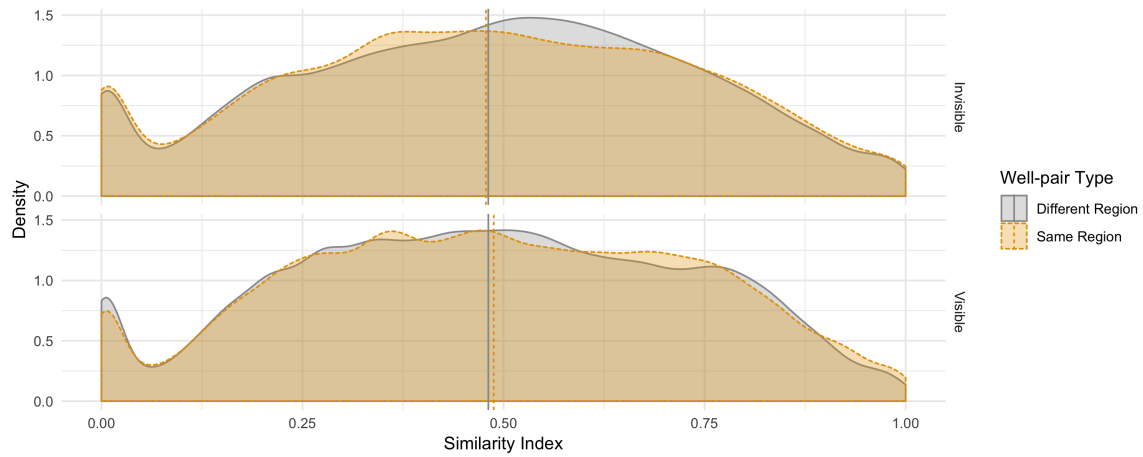|	Invisible	Visible
Difference in location	-0.00294	0.00599
95% Confidence Interval	[-0.00339, Infinity]	[0.00570, Infinity]
W (test statistic)	9.62e11	8.69e12
p-value	1	<2.2e-16

Notes: This table presents the statistical results of two Wilcoxon Rank Sum tests with a one-sided alternative hypothesis. The alternative hypothesis is that the CDF of the same-region distribution lies below that of the different-region distribution for well-pairs where the first well's inputs are invisible and visible respectively.

In support of this proposal, we observe a stark shift in the well-pair similarity distribution between those well-pairs whose first well's inputs are invisible vs. visible when the geographic distance between the two wells is small.[36] This shift is visible in the distributions plotted in Figure

---

[36]We use a binary variable denoting whether the two wells lie within the same region of PA, defined as in Table 7.

2.

Notes: The curves plot estimated densities for well-pair similarities s_{ij}, conditional on whether the inputs from the first well were visible when the second well was stimulated and whether the two wells are located in the same region of Pennsylvania. The vertical lines indicate the distribution means. All distributions include only those well-pairs where each well is completed by a different operator.

Both the top and bottom panels of Figure 2 plot two distributions of well-pair similarities: the grey/solid distribution is comprised of those well-pairs where the two wells are located in different regions from each other; the orange/dashed distribution includes well-pairs where the two wells are within the same region. The top panel plots those well-pairs where the inputs of the first well are *invisible* when the second well is stimulated, while in the bottom panel the first well's inputs are *visible* when the second well is stimulated. For all four distributions the mean is plotted as a vertical line. Finally it is worth emphasizing that none of the well-pairs plotted in Figure 2 have the same operator for both wells.

Looking at Figure 2, a major difference emerges. When we compare how similarities shift for intra-region well-pairs when inputs from the first well become visible, we see a noticeable rightward shift in the distribution. In contrast when we examine how visibility affects similarities for inter-region well-pairs, there is no effect at the mean and does not appear to be any effect along

Similar results hold when we use distance-based measures as well.

the rest of the distribution.

We move from a graphical to a statistical examination of this heterogeneity in Table 3. We perform two separate Wilcoxon Rank Sum tests with one-sided alternative hypotheses. The two cases are for 'invisible' or 'visible' well-pairs, depending on whether the first well's inputs have been disclosed. In each case the alternative hypothesis is that the CDF of the well-pair similarities when the two wells are located in the same region lies below the CDF of the the similarities when the wells are in different regions.

TABLE 3: WELL-PAIR SIMILARITIES BY VISIBILITY AND REGION

	Invisible	Visible
Difference in location	-0.00294	0.00599
95% Confidence Interval	[-0.00339, Infinity]	[0.00570, Infinity]
W (test statistic)	9.62e11	8.69e12
p-value	1	<2.2e-16

Notes: This table presents the statistical results of two Wilcoxon Rank Sum tests with a one-sided alternative hypothesis. The alternative hypothesis is that the CDF of the same-region distribution lies below that of the different-region distribution for well-pairs where the first well's inputs are invisible and visible respectively.

The first column of Table 3 shows the results from the test when we consider 'invisible' well-pairs. As can be seen, the estimated location difference is statistically indistinguishable from zero. In contrast, when we consider 'visible' well-pairs, the location difference is calculated to be positive and significant. Taken together, we can see that the visibility of the first well's inputs seems to have little (unconditional) distributional effect when the wells are located in different regions, but a statistically significant effect when the wells are in the same region. This finding lends statistical support to the above conjecture that the disclosures convey some tacit information above and beyond the list of ingredients; and this sort of tacit information could explain the difference in the coefficient estimates on $\hat{\alpha}_{UNOBS}$ and $\hat{\alpha}_{OBS}$ in Table 7.

As a final test of this hypothesis, we perform another regression based on Equation 2 with one

difference: we now use four α parameters along with four QS measures. We continue to differentiate between well-pairs where the first well's inputs are observed or unobserved, and we now introduce differentiation based on whether each well-pair's two wells are located in the same region or different regions. That is, the four combinations are unobserved and same region, unobserved and different region, observed and same region, and observed and different region. The results of this regression are presented in Table **??**.

[Table **??** about here]

As can be seen in Table **??**, we find a positive and statistically significant estimate for $\hat{\alpha}_{OBS,SameRegion}$, but for none of the other $\hat{\alpha}$s. We take this as further evidence consistent with our conjecture that valuable and tacit information is being conveyed through the chemical disclosures.

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 shale gas exploration and production firms rely on secrecy to capture value from investments in innovation, the erosion of secrecy via mandatory disclosure may also reduce incentives to experiment with new formulations.

We identify experimental chemical combinations by using a density-based clustering algorithm, locating each well in high-dimensional space based on the chemicals used in the fracture job and identifying wells that do not fit into groups as experimental. We demonstrate that the frequency of experimentation fell after mandatory disclosure, especially among the most productive

operators. Finally, we show that when operators mimic the experiments conducted by their peers — which, not surprisingly, happens significantly more often after mandatory disclosure — they are most likely to copy those experiments conducted by the most productive operators. Taken together, the evidence suggests that the most important experimentation — i.e., that of the most productive firms, which are also the most likely to be copied by their peers — fell after mandatory disclosure. This suggests that policymakers considering disclosure policies face a trade-off between socially beneficial impacts (e.g., technological diffusion and perhaps reduced toxicity) and costs that accrue from reduced innovation.

5.1 Measuring Experimentation

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. We choose the DBSCAN algorithm as it has several valuable properties for our setting: it does not require a pre-specified number of clusters (thereby minimizing 'supervision' from the researchers), it accommodates any distance measure, it works well with non-convex sets, and it does not need to assign every point to a cluster. This last property will lead us to our definition of an experimental well. More details on the DBSCAN algorithm and our implementation are available in Appendix A.

For consistency with prior sections, we measure proximity in chemical input space by using the Jaccard distance $d_{ij} \equiv 1 - s_{ij}$. A baseline approach might 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 1,000 most recently developed wells, keeping the size of the comparison set constant over time.[37]

5.2 Effects of Disclosure on Experimentation

Our key outcome of interest from the clustering analysis is wells that use an experimental chemical combination and thus are classified as noise points. We first consider the incidence of experimentation over time. Figure 5 shows a bin scatterplot in which each point represents a single calendar month, and measures the proportion of wells for which operators used an experimental combination of chemicals in the fracturing fluid. The vertical line indicates the month at which full public disclosure of chemicals became mandatory. The graph provides suggestive evidence of a decrease in experimentation after mandatory disclosure. In the months prior to mandatory disclosure (April 2012), while the percentage from month to month was somewhat volatile, there is no strong trend upward or downward prior to the disclosure rule. After mandatory disclosure, the average rate of experimentation remains volatile but generally falls.[38] The decline in experimentation over time is not monotonic, but is nonetheless evident in the graph, which also shows that in eleven of the post-disclosure months operators conducted no experiments at all.

[Figure 5 about here]

---

[37]Similar results hold for different alternative values of this "look-back" set, as well as look-back sets based on the number of days rather than number of wells. For the sake of brevity, these results are available from the authors upon request.

[38]As noted previously, since we apply the clustering algorithm to a comparison set of constant size, this is not a mechanical result of the passage of time.

A t-test, weighted by the number of wells fractured each month, verifies the difference in experimentation is significant: the mean incidence of experimentation prior to mandatory disclosure is 6.6 percent compared to 3.8 percent afterwards (significantly different at p=0.0057). This effect is also robust to controlling for the output price (see Appendix Table B.4.)

Another question of interest is how disclosure changes the degree to which operators copy the experiments conducted by their peers. The analysis in Section 4.3 demonstrates the effects of disclosure on copying in general; how much of this copying reflects inter-firm learning about experimental wells, in particular? To investigate this question we begin by defining the concept of firms "following" the experiments of other operators. We define the chemical use of later well j as *following* that of earlier experimental well i if the Jaccard distance is less than ε, that is, $0 \leq (1 - s_{ij}) \leq \varepsilon$, where for consistency we use the same value of ε that is used to define the clusters and experiments (see Appendix A for details). Figure 6 shows the incidence of inter-firm following of experimental wells, compared to operator quality conditional on location, price, and input choices — that is, the operator fixed effects θ_f recovered from estimating Equation (2). The figure demonstrates that it is the experiments of higher-quality firms that are most often copied by their peers.

[Figure 6 about here]

Table 9 offers more evidence on how the incidence of experiment following — and also experimentation — varies by operator quality, both before and after mandatory disclosure. The table echoes Figure 5 in showing that experimentation declined post-disclosure, with operators in all quartiles decreasing the mean monthly number of experiments and their proportion relative to total wells fractured (except for those in the third quartile, whose level of experimentation was approximately the same in both periods). Comparing the extent of inter-firm copying of experiments

42

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.[39] Furthermore, 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) suggests that disclosure erodes incentives to innovate by reducing the ability of firms to exclusively realize the fruits of their investments in innovation.

[Table 9 about here]

We leave a more complete model of experimentation for further research, but reiterate here the evidence of two notable 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: it appears in our setting that disclosure disseminated valuable knowledge, but also that it reduced incentives for firms to invest in innovative experiments.

---

[39]The difference between the values in the right-most column of Table 9, comparing pre- and post-disclosure for each quartile, is always statistically significant at $p < 0.001$ except for the two zero values for bottom-quartile firms.

6 The Role of Contractors

Having provided evidence for the transfer of valuable information about chemical mixtures, in this section we consider a possible channel of such information transfer that is unrelated to disclosure regulations: contractors. Contractors perform a variety of roles, not limited to the fracture itself: operators hire them to assist with tasks related to drilling, cementing, well logging, and other tasks, as well as designing and conducting the fracturing job. In Pennsylvania, the DEP requires operators to provide information about contractors on Well Completion Reports. These reports indicate that 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 DEP reports; as a result we were able to identify the fracturing contractor for 3,529 wells, about 73% of our sample.

6.1 Contractors and Input Similarities

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 chemicals used for earlier well in the well-pair are disclosed prior to the later well's fracture ($visible_{ij}$). For these regressions, we consider only inter-operator well-pairs.

$$s_{ij} = \beta_0 + \beta_1 contractor_{ij} + \beta_2 visible_{ij} + \beta_3 contractor_{ij} * visible_{ij} + \phi_{ij} + \varepsilon_{ij} \qquad (5)$$

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 4.2, we include fixed effects for the first and second wells' operator and for 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 10.[40]

[Table 10 about here]

When it is the only non-constant regressor, the dummy for sharing a contractor predicts 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.

6.2   Contractors, Input Choices, and Productivity

In addition to examining contractors' role in facilitating similar inputs across wells fractured by different operators, we also test the role they play in influencing the relationship between productivity and chemical choice — that is, whether the relevant entity for learning from other firms'

---

[40]As in the analysis in Section 4.2, we drop well-pairs where the wells were completed on the same day. See Footnote 27.

disclosures is the operator or the fracturing contractor. To do so, we run a variant of the test embodied in Equation (2) and Table 7, in which we restrict the estimation to wells that are fractured by large operators. Our reasoning is that large operators, who have more specialized technical staff and differentiated roles, may take a more direct role in the design of the fracturing job; whereas for smaller operators, whose staff tend to play more generalist (albeit still highly technical) roles, the fracturing contractor may be more likely to play the leading role in design and implementation of the fracture job. Thus, running a "big-operator-only" version of the test in section 4.3 allows us to test whether the relationship between chemical choices, productivity, and input disclosure holds for the subset of wells in which the operator plays the key role in directing chemical choices and fracture design.

Table 11 shows the results of this test among the set of wells fractured by operators with at least 50 wells. In sign and significance, the results are virtually identical to those in Table 7; indeed, the magnitudes of the coefficients on QS_i^{OBS} are somewhat larger among wells fractured by larger operators — for instance, the results in column (1) suggest that a 100% increase in $(1 + QS_i^{OBS})$ predicts a 9.07% increase in productivity per foot, and results are of similar magnitude for other columns representing alternative choices for the look-back period.[41]

From both tests — the former regarding the role of contractors in facilitating similar inputs across different-operator wells, and the latter regarding the role of contractors in productivity-enhancing information diffusion — we conclude that while contractors do facilitate greater similarity of inputs across well-pairs, the disclosure laws we study still play a key role. Specifically, mandatory information disclosure allows important, productivity-enhancing knowledge transmis-

---

[41]Appendix Table B.5 shows the results for operators with at least ten wells, and the results are again comparable to those in Table 7.

sion across operators over and above that which arises due to shared contractors.

7 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 new question — whether disclosure laws create pathways for knowledge transmission that were previously inaccessible or overly costly — 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 implies significantly increased well productivity. The evidence suggests that disclosure laws opened a new channel for inter-firm 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

47

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," 2019. 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.

Budish, Eric, Benjamin N. Roin, and Heidi L. Williams, "Patents and Research Investments: Assessing the Empirical Evidence," *American Economic Review*, May 2016, *106* (5), 183–187.

Cai, Jing and Adam Szeidl, "Interfirm relationships and business performance," *The Quarterly Journal of Economics*, 2018, *133* (3), 1229–1282.

Cohen, Wesley M., 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.

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.

Dalton, Patricio, Bilal Zia, Julius Rüschenpöhler, and Burak Uras, "Learning Business Practices from Peers: Experimental Evidence from Small-scale Retailers in an Emerging Market," WorkingPaper, Tilburg University February 2018.

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," 2019. Unpublished manuscript.

Fitzgerald, Timothy and Charles F. Mason, *The Importance of Productivity Effects of Hydraulic Fracturing to Broader Exploitation of Unconventional Natural Gas Deposits* April 2020. Working Paper.

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

Greenstone, M., P. Oyer, and A. Vissing-Jorgensen, "Mandated Disclosure, Stock Returns, and the 1964 Securities Acts Amendments," *The Quarterly Journal of Economics*, May 2006, *121* (2), 399–460.

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.

Ikonnikova, Svetlana A., Katie Smye, John Browning, Robin Dommisse, Amin Gherabati, Gürcan Gülen, Scott Hamlin, Casee Lemons, Frank Male, Guin McDaid, Ken Medlock III, Bridget Scanlon, Mark Shuster, Scott Tinker, and Emilian Vankov, Final Report on Update and Enhancement of Shale Gas Outlooks, University of Texas, Bureau of Economic Geology 2018.

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.

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.

Lerner, Josh, "The Empirical Impact of Intellectual Property Rights on Innovation: Puzzles and Clues," *American Economic Review*, April 2009, *99* (2), 343–348.

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

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.

Nordhaus, William, *Invention, Growth, and Welfare: A Theoretical Treatment of Technological Change*, MIT Press, 1969.

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

Pennsylvania Geologic Survey, "EDWIN (Exploration and Development Well Information Network)." Well completion reports and fluid stimulation additive reports retrieved from database of the Pennsylvania Geological Survey. Online at https://edwin.onbaseonline.com, accessed from 2014 to 2017.

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

Sakakibara, Mariko and Lee Branstetter, "Do Stronger Patents Induce More Innovation? Evidence from the 1988 Japanese Patent Law Reforms," *The RAND Journal of Economics*, 2001, *32* (1), 77.

Skytruth, "Fracking Chemical Database," 2013. Online database at www.skytruth.org.

Steck, Andrew L., "Industry Dynamics with Social Learning: Evidence from Hydraulic Fracturing," 2019. 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.

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.

Williams, Heidi L., "How Do Patents Affect Research Investments?," *Annual Review of Economics*, August 2017, *9* (1), 441–469.

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.

TABLE 4: SUMMARY INFORMATION FOR WELLS IN SAMPLE WITH CHEMICALS DATA

Year	Fluid Volume Median	Mean	# Chemicals / Well Median	Mean	First 6 Month Gas per Foot 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).

Year	# Wells	# Operators	Chemicals		
			# 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.

TABLE 6: DETERMINANTS OF PRODUCTION INPUT SIMILARITIES

	(1)	(2)	(3)	(4)
Different Operator	−0.195***	−0.197***	−0.198***	−0.191***
	(0.000)	(0.000)	(0.000)	(0.000)
Different Operator	0.009***	0.012***	0.013***	0.013***
x Visible	(0.000)	(0.000)	(0.000)	(0.000)
Polynomial Date		✓	✓	✓
Polynomial Time Between			✓	✓
Polynomial Distance				✓
Operator FEs	✓	✓	✓	✓
Observations	11,808,695	11,808,695	11,808,695	11,808,695
R^2	0.339	0.340	0.341	0.341

Notes: Results are from the regression described in Equation (1); the similarity index is the Jaccard Abundance index, described in detail in Section 4.1. An observation is a pair of wells, as described in the text. Standard errors are calculated from a standard bootstrap routine with 1,000 iterations. Polynomial controls are all 5th degree. *, **, and *** indicate statistical significance at the 10, 5, and 1 percent levels.

$\mathscr{J}$ Restriction	$\ln GPF_i$		
	180 Days	360 Days	720 Days
Log Fluid Volume	0.216***	0.216***	0.215***
	(0.0198)	(0.0198)	(0.0198)
Throttling	−0.569***	−0.572***	−0.571***
	(0.0351)	(0.0351)	(0.0351)
Northeast	0.390***	0.391***	0.391***
	(0.0421)	(0.0421)	(0.0422)
Output Price	0.0810***	0.0788***	0.0745***
	(0.0141)	(0.0149)	(0.0148)
QS^{OBS}	0.0756***	0.0659***	0.0611***
	(0.0158)	(0.0170)	(0.0166)
QS^{UNOBS}	−0.0177	−0.0131	−0.0137
	(0.0164)	(0.0159)	(0.0155)
Date Controls	Linear	Linear	Linear
Observations	4,794	4,794	4,794
RMSE	0.715	0.715	0.715

Notes: Coefficients are estimated via non-linear least squares (NLLS); standard errors are shown in parentheses. Each column represents a different estimation that reflects a different specification of the "lookback period" $\mathscr{J}$ for constructing the QS_i^{OBS} and QS_i^{UNOBS} weights. Superscripts $^{+}$, *, **, and *** indicate statistical significance at the 10, 5, 1, and 0.1 percent levels. See text for details.

Operator	Northeast	180 days		360 days		720 days	
		q25	q75	q25	q75	q25	q75
Hunt		272.9	352.1	286.7	360.1	372.5	463.1
Hunt	✓	403.2	520.3	424.0	532.5	550.9	684.8
Chesapeake		700.9	904.5	730.6	917.5	944.7	1,174.3
Chesapeake	✓	1,035.5	1,336.3	1,080.5	1,356.9	1,397.0	1,736.6
Hunt		299.1	346.8	310.0	359.6	401.1	464.8
Hunt	✓	441.8	512.3	458.5	531.8	593.2	687.4
Chesapeake		768.2	890.8	789.9	916.3	1,017.2	1,178.7
Chesapeake	✓	1,134.9	1,316.0	1,168.2	1,355.1	1,504.2	1,743.0

Notes: Each entry represents predicted first sixth month gas output per foot given the estimates of Equation 2 presented in Table 7. Entries in the top half of the table use the estimated operator effect θ_f for the operat in column 1, set the variable *Northeast* as indicated in column 2, and use a value of QS^{OBS} at the 25th or 75th empirical percentile as indicated by the column labels. All other variables are set to sample means. The third and fourth columns correspond to estimates from the 180-day lookback specification, the fifth and sixth columns use the 360-day specification, and the seventh and eighth columns use the 720-day specification.

The bottom half of the table repeats a similar exercise, but sets QS^{OBS} to the sample mean and sets *Fluid Volume* to the 25th or 75th empirical percentile respectively.

Period	Operator Quartile	Experiments	Mean Monthly Experiments	Experiments as % of wells	% of Other-Operator Wells Following
Pre-disclosure	1	9	1.3	7.4	0
Pre-disclosure	2	12	1.7	4.9	0.03
Pre-disclosure	3	6	0.9	2	0.18
Pre-disclosure	4	11	1.6	10.3	0.06
Post-disclosure	1	2	0.1	1	0
Post-disclosure	2	12	0.3	1.5	0.21
Post-disclosure	3	41	1.1	3	0.56
Post-disclosure	4	27	0.7	3.5	0.64

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 4.3 from Equation (2). 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.

| | 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 (5); 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.

| $\mathscr{J}$ Restriction	$\ln GPF_i$		
	180 Days	360 Days	720 Days
Log Fluid Volume	0.229***	0.228***	0.226***
	(0.0207)	(0.0208)	(0.0208)
Throttling	−0.560***	−0.561***	−0.561***
	(0.0364)	(0.0364)	(0.0364)
Northeast	0.386***	0.388***	0.389***
	(0.0434)	(0.0438)	(0.0437)
Output Price	0.0854***	0.0830***	0.0821***
	(0.0167)	(0.0168)	(0.0167)
QS^{OBS}	0.0907***	0.0849***	0.0813**
	(0.0200)	(0.0255)	(0.0270)
QS^{UNOBS}	−0.0176	−0.0135	−0.00992
	(0.0167)	(0.0189)	(0.0193)
Date Controls	Linear	Linear	Linear
Observations	4,528	4,528	4,528
RMSE	0.727	0.727	0.728

Notes: Limited to wells that are fractured by operators with at least fifty wells. Coefficients (and standard errors, in parentheses) are estimated via non-linear least squares (NLLS). Each column conducts the estimation using a different specification of $\mathscr{J}$ when constructing the QS_i^{OBS} and QS_i^{UNOBS} weights. The different columns represent $+$, $*$, $**$, and $***$ indicate statistical significance at the 10, 5, 1, and 0.1 percent levels. See text for details.
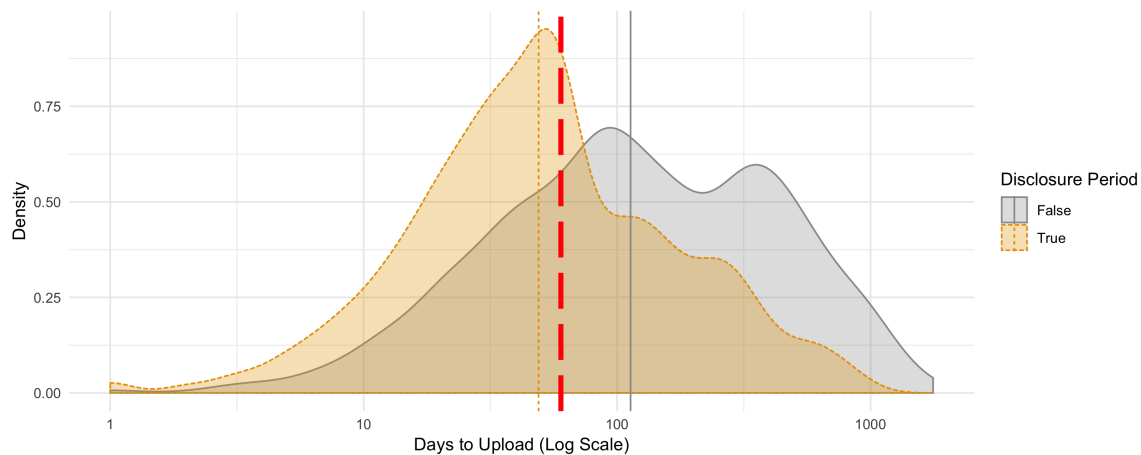
Notes: This plot contains densities of upload time in days, with the sample split between those wells stimulated before and after the disclosure regulation came into effect; vertical lines indicate sub-sample medians. The dashed red line indicates the stated deadline. See text for details.

FIGURE 4: JACCARD SIMILARITIES BY SAME / DIFFERENT OPERATOR STATUS
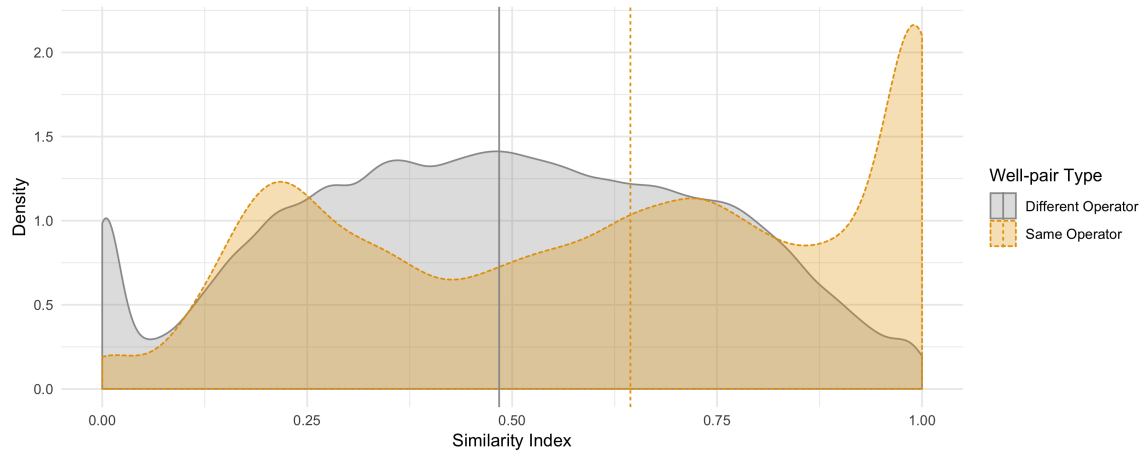

Notes: The curves plot estimated densities for well-pair similarities s_{ij}, conditional on whether the wells in the pair had the same or different operators. The vertical lines indicate the distribution medians.

FIGURE 5: PROPORTION OF EXPERIMENTS OVER TIME
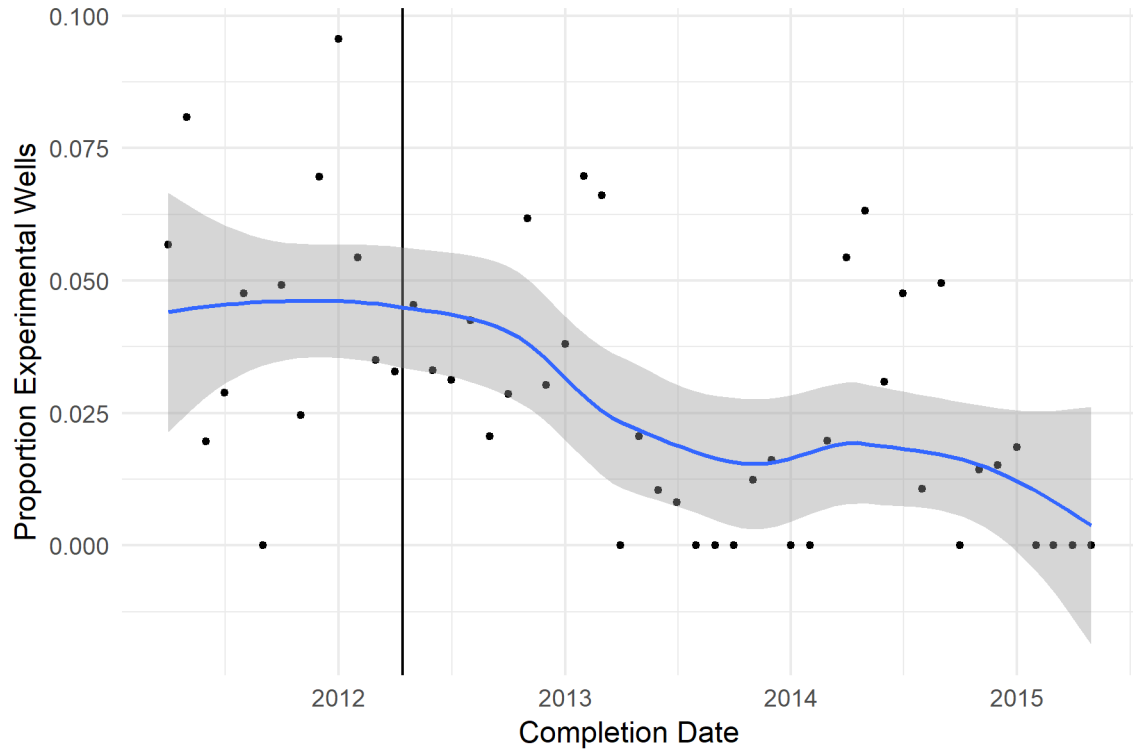
Notes: Each point represents one calendar month. The vertical axis indicates the proportion of wells hydrauli-cally fractured in that month for which operators used an experimental combination of chemicals, compared to the most recent 1,000 wells fractured in Pennsylvania. The vertical black line marks April 2012, the month in which full public disclosure became mandatory. The best-fit line is a locally weighted linear (loess) regression.

FIGURE 6: EXPERIMENT FOLLOWING BY OTHER OPERATORS
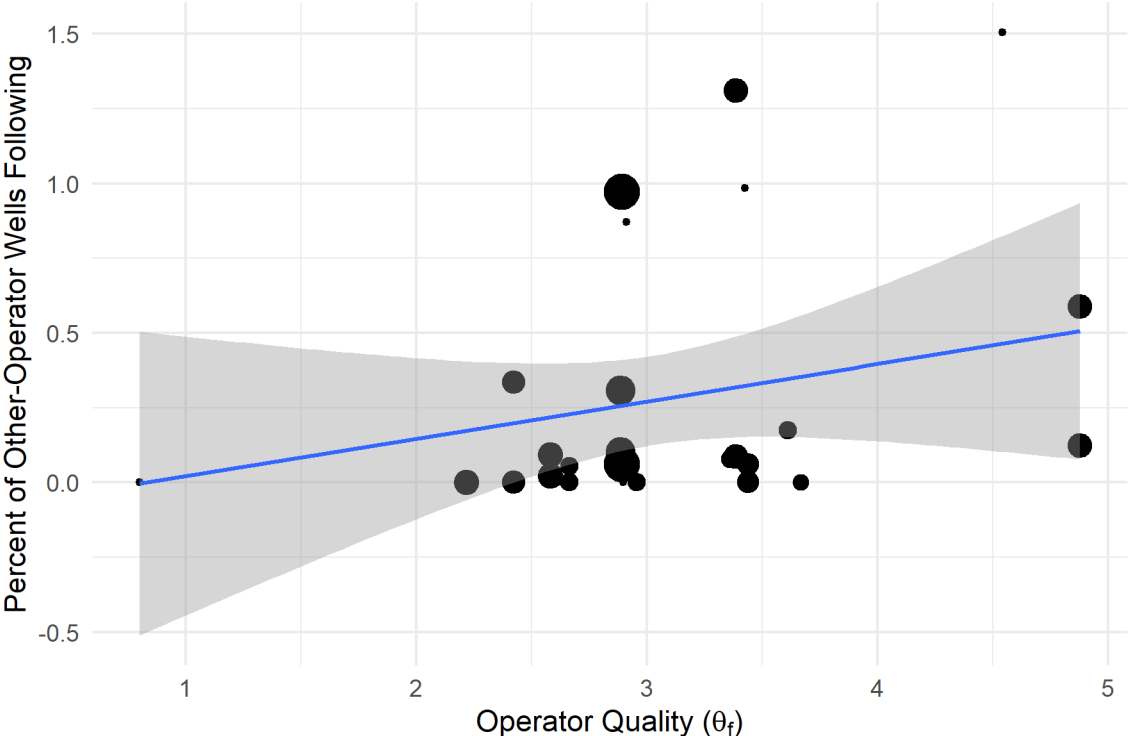
Notes: Each point represents a single operator; size of points indicates the number of wells fractured by each operator. The vertical axis measures the percent of experiments of this operator that other operators follow (see text for details); the horizontal axis indicates operator 'quality' as measured by the estimated value of theta from Equation 2. The best-fit line is a linear regression weighted by the number of wells fractured by each operator.

Appendices

Appendix A presents details about the algorithm used to identify experimental wells in chemical input space. Appendix B presents additional tables with results from robustness checks in Section 4.3, and a table with additional results of the experimentation analysis.

# A   Illustration of DBSCAN Algorithm

This section describes the DBSCAN algorithm, which we use to sort wells into clusters in chemical input space and identify experimental wells.

The DBSCAN clustering algorithm (Ester et al., 1996) provides an intuitive approach to identifying clusters of similar items (i.e., chemical combinations used for well stimulation) and items that do not fit any cluster. The algorithm groups together points that have many nearby neighbors, and also identifies points that are 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 points within that ball. If one or more observations lies within the radius of the ε-ball, the algorithm then groups the collection together, constructs an ε-ball around the perimeter, and continues in iterative fashion, searching for nearby data points and adding to the collection any that lie within the new (larger) ball. To be designated a cluster, a collection of points must have cardinality over a threshold that is set as a tuning parameter. An illustrative example is provided in Figure A.1.

Observations that fall outside any cluster — because there is no other observation within an ε-ball, or because the observations within a candidate cluster are too few to meet the minimum points threshold — are deemed 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 a pre-specified number of clusters (e.g., k-means clustering or finite mixture models), it identifies the number of clusters based on the observed data. Second, it is effective in discovering clusters that do not form convex sets, because points are placed into clusters based on their proximity to one another rather than on their location relative to space-

65

FIGURE A.1: DBSCAN ALGORITHM ILLUSTRATION

First Step

Second Step

Third Step

Final
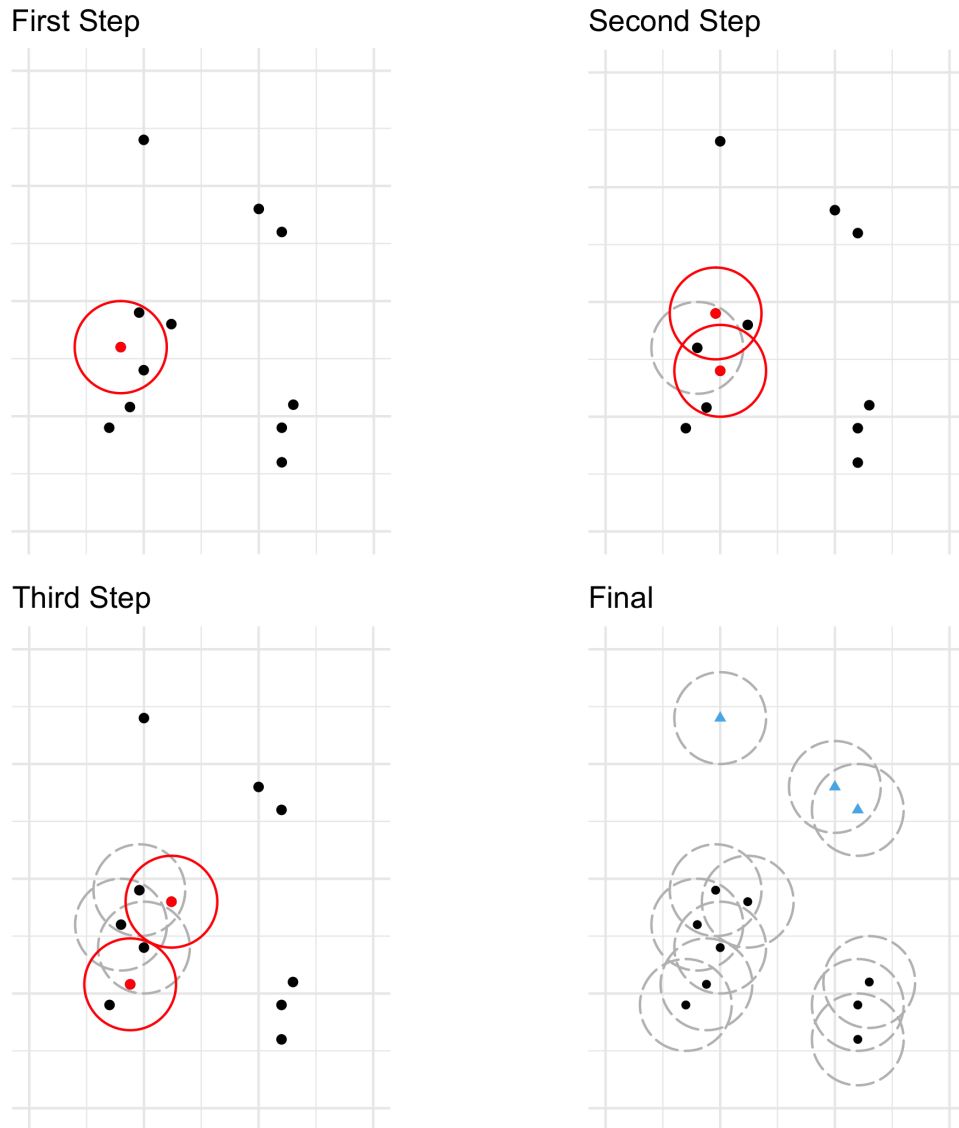
Notes: An example of the density-based spatial clustering of applications with noise (DBSCAN) algorithm, with the minimum-points parameter set at 3. In the first step, a point is selected and an ε-ball is drawn around it (marked in red). In the next step, ε-balls are drawn around the encompassed points, and the process repeats in the third step. The fourth pane shows the outcome of the algorithm: two clusters have been identified, along with three noise points (marked as blue triangles).

partitioning hyperplanes. 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). The algorithm is computationally efficient, and has been used in a wide variety of applications (Ester et al., 1996). Finally, DBSCAN does not require the assignment of every point to a cluster; the fact that it identifies some observations as "noise points" is exactly why it is appealing in our setting.

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 dense matrices is to use the number of dimensions in the data or, for sparse matrices, the intrinsic dimensionality (Bruske and Sommer, 1998; Johnsson, 2016). 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 chemical inputs for 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 applies principal components analysis to local subspaces of the observed data to identify the minimum number of parameters needed to describe the data we observe. The resulting estimate of intrinsic dimensionality varies from 10 to 16 depending on the number of local subspaces chosen; accordingly, for the results that follow, we set the minimum-points parameter at 10.[42]

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 A.2 shows these distances for our data, based on a 1000-well lookback and a threshold cluster size of 10. (The figure includes two curves: one that includes all wells in the dataset, and one that omits the first 1,000 wells in the dataset, which do not have a full 1,000-well comparison set.) Reading the sorted kNN distance plot from right to left, the knee of the plot shows where the 10th-nearest-neighbor distance increases dramatically; this corresponds to a value of ε that would result

---

[42]Results using alternative values are similar to those reported here; details are available on request.

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 10th-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 A.2: JACCARD DISTANCE TO 10TH-NEAREST NEIGHBOR
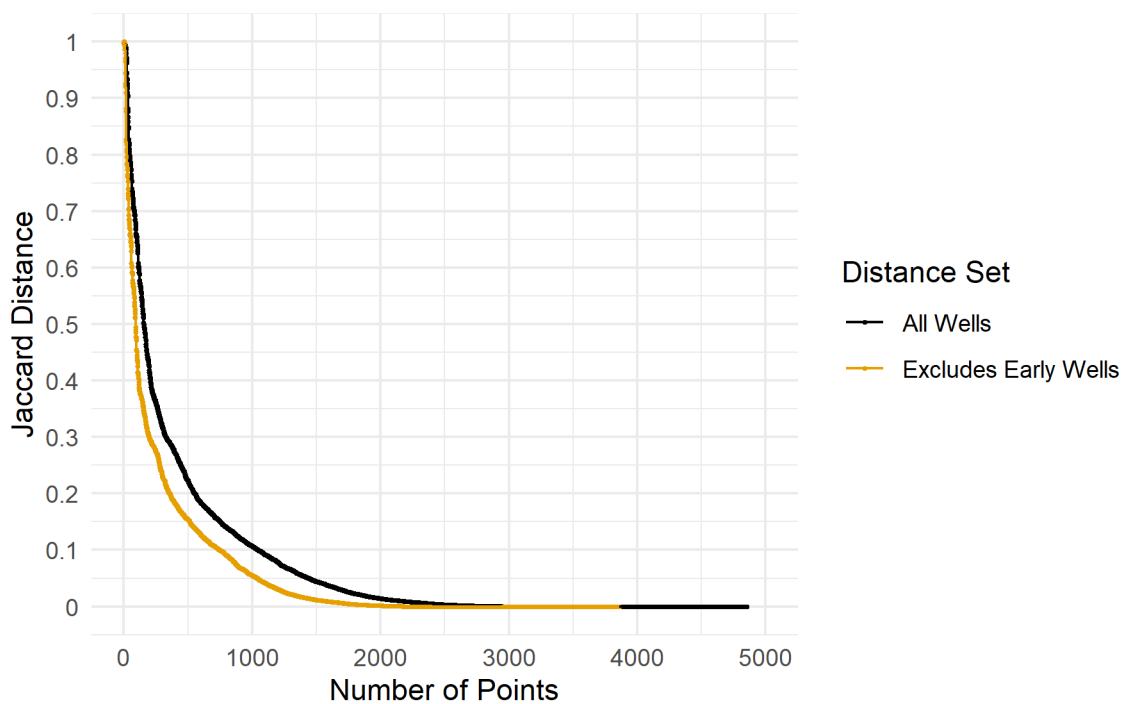
Notes: Sorted Jaccard distance indicating the proximity of each well's chemical mix to its 10th-nearest neighbor, within a comparison set of the 1,000 most recent prior wells. The outer (black) curve includes all wells, while the inner (yellow) curve excludes the earliest 1,000 wells ranked by completion 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.

B Additional Tables

TABLE B.1: EFFECTS OF INPUT VISIBILITY ON PRODUCTIVITY (WITH ALTERNATIVE RE-
GIONAL CONTROLS)

		$\ln GPF_i$	
$\mathscr{J}$ Restriction	180 Days	360 Days	720 Days
Log Fluid Volume	0.211***	0.211***	0.210***
	(0.0189)	(0.0189)	(0.0189)
Throttling	−0.595***	−0.596***	−0.596***
	(0.0340)	(0.0340)	(0.0340)
Gas-in-place	0.00500***	0.00499***	0.00500***
	(0.000245)	(0.000246)	(0.000246)
Output Price	0.0641***	0.0620***	0.0608***
	(0.0136)	(0.0145)	(0.0143)
QS^{OBS}	0.0463**	0.0385*	0.0369*
	(0.0157)	(0.0162)	(0.0159)
QS^{UNOBS}	0.00259	0.00667	0.00680
	(0.0155)	(0.0151)	(0.0155)
Date Controls	Linear	Linear	Linear
Observations	4,794	4,794	4,794
RMSE	0.692	0.692	0.692

Notes: This table differs from Table 7 by the use of regional gas-in-place estimates from
Ikonnikova et al. (2018) rather than a binary variable for the northeast region. Coefficients
are estimated via non-linear least squares (NLLS); standard errors are in parentheses. Each
column conducts the estimation using a different specification of $\mathscr{J}$ when constructing
the QS_i^{OBS} and QS_i^{UNOBS} weights. The superscripts $^+$, *, **, and *** indicate statistical
significance at the 10, 5, 1, and 0.1 percent levels. See text for details.

TABLE B.2: EFFECTS OF INPUT VISIBILITY ON PRODUCTIVITY (WITH YEAR FES)

$\mathscr{J}$ Restriction	$\ln GPF_i$		
	180 Days	360 Days	720 Days
Log Fluid Volume	0.210***	0.208***	0.208***
	(0.0191)	(0.0195)	(0.0196)
Throttling	−0.572***	−0.572***	−0.573***
	(0.0351)	(0.0351)	(0.0351)
Northeast	0.386***	0.386***	0.385***
	(0.0420)	(0.0421)	(0.0422)
Output Price	0.0759***	0.0757***	0.0712***
	(0.0153)	(0.0158)	(0.0157)
QS^{OBS}	0.0779**	0.0625*	0.0498+
	(0.0251)	(0.0262)	(0.0266)
QS^{UNOBS}	−0.00796	0.000956	0.000886
	(0.0211)	(0.0215)	(0.0200)
Date Controls	Year FEs	Year FEs	Year FEs
Observations	4,794	4,794	4,794
RMSE	0.715	0.715	0.715

Notes: This table differs from Table 7 by the use of year fixed effects rather than a linear time trend. Coefficients are estimated via non-linear least squares (NLLS); standard errors are in parentheses. Each column conducts the estimation using a different specification of $\mathscr{J}$ when constructing the QS_i^{OBS} and QS_i^{UNOBS} weights. The superscripts $+$, $*$, $**$, and $***$ indicate statistical significance at the 10, 5, 1, and 0.1 percent levels. See text for details.

TABLE B.3: EFFECTS OF INPUT VISIBILITY ON PRODUCTIVITY (WITH TEMPORAL RESTRICTIONS)

$\mathscr{J}$ Restriction	$\ln GPF_i$		
	180 Days	360 Days	720 Days
Log Fluid Volume	0.197***	0.195***	0.196***
	(0.0200)	(0.0200)	(0.0201)
Throttling	−0.610***	−0.610***	−0.610***
	(0.0386)	(0.0386)	(0.0386)
Northeast	0.361***	0.360***	0.359***
	(0.0449)	(0.0450)	(0.0450)
Output Price	0.0586***	0.0581***	0.0574***
	(0.0145)	(0.0149)	(0.0155)
QS^{OBS}	0.0591**	0.0544**	0.0499*
	(0.0186)	(0.0201)	(0.0204)
QS^{UNOBS}	0.00459	0.00455	0.00932
	(0.0174)	(0.0179)	(0.0180)
Date Controls	None	None	None
Observations	4,111	4,111	4,111
RMSE	0.720	0.720	0.720

Notes: Observations include only those wells that are fractured within 14 months before or after the mandatory disclosure law took effect. Coefficients are estimated via non-linear least squares (NLLS); standard errors are in parentheses. Each column conducts the estimation using a different specification of $\mathscr{J}$ when constructing the QS_i^{OBS} and QS_i^{UNOBS} weights. The superscripts $^+$, *, **, and *** indicate statistical significance at the 10, 5, 1, and 0.1 percent levels. See text for details.

TABLE B.4: EFFECT OF DISCLOSURE ON INCIDENCE OF EXPERIMENTATION

	Experimental Well	
	(1)	(2)
Intercept	−2.642***	−2.296***
	(0.154)	(0.408)
Post-Disclosure	−0.581***	−0.551***
	(0.180)	(0.182)
Price		−0.120
		(0.131)
Observations	3,861	3,861

Notes:	***Significant at the 1 percent level.
	**Significant at the 5 percent level.
	*Significant at the 10 percent level.

TABLE B.5: EFFECTS OF INPUT VISIBILITY ON PRODUCTIVITY (OPERATORS WITH AT LEAST 10 WELLS)

$\mathscr{J}$ Restriction	180 Days	360 Days	720 Days
		$\ln GPF_i$	
Log Fluid Volume	0.219***	0.219***	0.218***
	(0.0200)	(0.0201)	(0.0201)
Throttling	−0.565***	−0.567***	−0.566***
	(0.0353)	(0.0353)	(0.0353)
Northeast	0.389***	0.389***	0.388***
	(0.0424)	(0.0426)	(0.0426)
Output Price	0.0828***	0.0797***	0.0785***
	(0.0155)	(0.0159)	(0.0160)
QS^{OBS}	0.0822***	0.0702***	0.0684**
	(0.0171)	(0.0212)	(0.0253)
QS^{UNOBS}	−0.0180	−0.0140	−0.0134
	(0.0158)	(0.0161)	(0.0156)
Date Controls	Linear	Linear	Linear
Observations	4,737	4,737	4,737
RMSE	0.717	0.717	0.717

Notes: Limited to wells that are fractured by operators with at least ten wells. Coefficients (and standard errors, in parentheses) are estimated via non-linear least squares (NLLS). Each column conducts the estimation using a different specification of $\mathscr{J}$ when constructing the QS_i^{OBS} and QS_i^{UNOBS} weights. The superscripts $^+$, *, **, and *** indicate statistical significance at the 10, 5, 1, and 0.1 percent levels. See text for details.