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

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

February 14, 2017

Preliminary and Incomplete
Do Not Cite or Circulate

Abstract

Firms can learn about new technologies from other adopters, though they do not always exploit this form of social learning. We examine whether shale gas operators took advantage of environmentally-focused disclosure laws to learn from competitors and improve productivity; with this information, we evaluate the claim that chemical disclosure rules expose valuable trade secrets. We exploit an unusual regulatory episode in Pennsylvania that allows us to see chemical inputs prior to public disclosure. Using detailed data on well-level inputs and outputs, we study how the change in disclosure regime affected operators' chemical use and well productivity.

KEYWORDS: Information-Based Regulation, Social Learning, Disclosure, Trade Secrets, Shale Gas, Hydraulic Fracturing Fluid
JEL CODES: L24, L51, L71, O3, Q35, Q53, Q55, Q58

*T. Robert Fetter: University Program in Environmental Policy, Duke University; corresponding author: 4120 Environment Hall, Box 90328, Durham NC 27708; rob.fetter@duke.edu. Andrew L. Steck: Department of Economics, Duke University; andrew.steck@duke.edu. Christopher Timmins: 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. We are thankful for helpful comments and suggestions from conference participants at the Norwegian School of Economics and members of the Duke Public Economics Lab, Duke Environmental Policy seminar, and Duke Energy Initiative Fellows. Any errors and omissions are our own.

1 Introduction

Disclosure laws are used by regulators to disseminate information about potentially damaging activities by firms, with the hope of providing firms with incentives to reduce those activities. While disclosure policies may lead to an environmental benefit, they can also have a cost if the lack of secrecy limits firms' ability to realize the benefits of innovation. The existence of this tradeoff depends on the answers to two questions. First, does the disclosed information have value to firms? Second, does disclosure effectively enable social learning, and thereby limit firms' ability to capitalize on innovation? We study both questions in the context of hydraulic fracturing chemicals, and find affirmative evidence for each. Thus, while the use of toxic substances in hydraulic fracturing fluids makes a case for a public benefit from disclosure, policy makers should weigh that benefit against disclosure's costs on industry.

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

Similar disclosure policies designed to disseminate information about pollution releases (or the potential for such releases) have become increasingly popular in many industries. See, for example, policies regulating electricity generation (Delmas et al., 2010; Kim and Lyon, 2011), and drinking water provision (Bennear and Olmstead, 2008). By harnessing the collective power of stakeholders internal and external to the disclosing entity, information-based policies may improve public welfare via market, political, or managerial mechanisms (Fung et al., 2007; Bennear and Olmstead, 2008), as well as provide stakeholders with information that supports their "right to know" about potential environmental or health risks. For example, the regulations may reduce the use of toxic chemicals and potential threats to environmental quality or human health. Analyses within industries that are less consumer-facing,

such as manufacturing facilities that report to the Toxics Release Inventory (TRI), find that reported releases of toxic chemicals have declined over the course of the program by as much as 50 percent (Bennear and Coglianese, 2005). In non-environmental contexts, information-based policies have been used to regulate financial disclosure, nutritional labeling, restaurant hygiene, and other matters (Fung et al., 2007; Jin and Leslie, 2003). Information-based policies seem to be especially popular in settings where conventional regulatory approaches are ill-suited, such as when risks are not well understood, but are not anticipated to be extraordinary. They often represent a "pragmatic compromise" (Fung et al., 2007) that is a politically viable response to emerging risks.

In the case of hydraulic fracturing, operators are likely to have better information than regulators regarding which chemicals are most effective for releasing trapped hydrocarbons in a particular shale formation. A consequence of this asymmetric information is that direct regulations (i.e., command-and-control policies) are unlikely to be efficient. In contrast, disclosure regulations, by opening firms to the threat of public pressure or legal action, can incentivize responses that take into account the tradeoffs between productivity impacts and potential sanctions. For instance, firms may respond by using less toxic formulations, or take further actions to prevent leaks and spills.

At the same time, mandatory information disclosure may impose unintended costs on firms. In particular, these policies may limit operators' ability to monetize the value of innovations. Firms in the oil and gas industry, as in many industries, rely on secrecy as a mechanism to capture value from their investments in innovation and maintain their competitive advantage (Wang and Krupnick, 2013; Cohen et al., 2000). Indeed, the industry has made the argument that fluid formulas represent "trade secrets" in response to calls for mandatory disclosure regulations. In this paper, we shed light on whether innovations in fracturing fluids do in fact constitute a competitive advantage. In particular, we compare detailed, well-level information on inputs and outputs to test whether the chemicals used become more similar with public disclosure. We observe evidence that chemical mixes do converge for wells across different firms. We then test whether using chemicals more similar to high-performing wells appears to improve productivity for poorer-performing firms, and find an affirmative answer. These findings suggest policy-makers need to weigh the tradeoff

between right-to-know disclosure rules and industry secrecy. A further question is: to what extent does forced disclosure undermine the incentives for firms in this industry to invest in research and development for future innovations? We leave this latter question to on-going work.

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

We perform a series of empirical analyses designed to test whether regulations requiring the mandatory disclosure of chemicals created opportunities for social learning—even when those regulations included provisions that permitted operators to declare some chemicals "proprietary." As a first step, we look for evidence that the unexplained variation in production declined over time; that is, if there is a tighter distribution of residuals from a regression of output on observable well characteristics. We then test for convergence in inputs, as well as test for convergence in productivity across operators; in the latter test, we focus on evidence that low-performing firms used information revealed by the disclosure law to catch up with more successful firms. There could, naturally, be other reasons for this convergence. We therefore focus on the specific mechanism of interest, and examine the role of input similarity in the design of fracturing fluid. We find that convergence is indeed explained specifically by weaker firms adopting the chemical mixtures used by more successful wells in previously drilled wells, after the disclosure rules have gone into place.

This paper makes three primary contributions. First, we provide evidence that secrecy is valuable to hydraulic fracturing firms. This fact raises the tradeoff between the public's right

to know about fracturing chemicals used and firms' right to secrecy. Second, we study the interplay between information disclosure regulations and social learning. Third, we examine the role that chemicals play in the technology of hydraulic fracturing.

Our first contribution is to the literature on the relationship between secrecy and innovation. We provide some insight into the importance of the social tradeoffs policy makers face when considering disclosure laws. There may be a compelling social benefit from disclosing specific information, such as when production processes involve toxic chemicals in residential areas. This social benefit must be weighed against the ability of firms to realize economic returns from their innovations. Theoretically, patents solve the problem of imperfect appropriability, but in practice, evidence in many industries suggests firms realize a competitive advantage from innovation through a combination of secrecy, lead time, and investment in complementary capabilities (Cohen et al., 2000). We provide evidence that innovations in hydraulic fracturing chemicals are valuable, suggesting that this tradeoff may be important in this case.

Second, this paper expands the literature on the empirical effects of transparency or disclosure regulations in a new direction. Our analysis is the first, to our knowledge, to combine a study of the effects of information disclosure regulations with social learning. We find strong evidence that information disclosure regulations about an emerging technology can enable social learning.[1] Prior papers have documented the effects of disclosure laws on environmental, health, or other outcomes (e.g., Jin and Leslie, 2003; Fung et al., 2007; Bennear and Olmstead, 2008; Delmas et al., 2010; Fetter, 2016). In addition, some work has analyzed the effect of environmental disclosures on investor behavior (e.g. Hamilton, 1995). Separately, other authors have studied the phenomenon of social learning (Conley and Udry, 2010; Covert, 2015), without considering the role of disclosure regulations specifically.

Third, we study the role that chemical additives have played in the development of hydraulic fracturing technology. A number of recent papers have studied the rise of hydraulic fracturing, but none have addressed the role played by chemicals. We take the role of chemicals seriously, and thereby contribute to the literature on firm learning about hydraulic fracturing specifically, and emerging technologies more generally.

[1] This could also be interpreted as a form of technology diffusion facilitated by regulation.

The paper proceeds as follows. In Section 2, we provide context and background for our study, including a review of the emergence of the technologies that have made development of shale gas economically feasible, the environmental concerns that led to the requirement for public disclosure, and the literature on the role of alternative strategies—including secrecy—that firms use to preserve competitive advantage. Section 3 describes our data. In Section 4 we present our analysis of convergence and catch-up in operators' output, as well as evidence on how the disclosure rules affected chemical input choices. Section 5 offers a discussion and concluding remarks.

2 Background

2.1 Engineering and Geology of Hydraulic Fracturing

The rise of shale gas to prominence in the US energy landscape has been well-documented.[2] The ability to profitably recover hydrocarbons from shale has been largely based on advances in four key areas of technology: horizontal drilling, three-dimensional seismic imaging, micro-seismic fracture mapping, and massive hydraulic fracturing (Wang and Krupnick, 2013). Elements of these technologies have been in development for several decades, spurred by both private and public investments in research and development. Technological advances since the 1970's have ranged from changes in the major compounds comprising fracturing fluid to greater control over directional drilling of wellbores. Foam was replaced by gels in the formulation of fracturing fluid, and order-of-magnitude changes were made in the quantity of proppant used. In the 1990's, there were important advances in the role of directional drilling; in combination with massive hydraulic fracturing, the ability to drill horizontally through a shale formation made shale gas development economical. At the same time, "slick water" fracturing fluid replaced gels. Most recently, fracturing fluid has been refined in

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

multiple dimensions for maximizing output and minimizing costs.

A wide array of chemicals are used in hydraulic fracturing fluid to enhance the productivity of the primary inputs—water and sand. In particular, firms use chemical additives to help open fractures in the rock, transport the proppant along the length of the fracture, lower viscosity in order to allow faster pumping and higher pressures, minimize fluid loss into the face of the formation, reduce scaling on the formation, reduce chemical corrosion or bacterial growth that might threaten the integrity of metal casings, facilitate breakup of other chemicals post-fracture, and serve other purposes (Stringfellow et al., 2014; Montgomery, 2013; Gulbis and Hodge, 2000). In short, fracturing fluid is a complex mixture in which an additive that improves performance in one dimension may reduce performance in another. Although the cost of the chemicals themselves is usually small in comparison to the overall cost of the well stimulation operation, the proper choice of chemicals may have dramatic effects on the overall cost and productivity of a well.[3]

As of 2014, the total estimated recovery for shale oil wells was on the order of 5%, compared to 50% for conventional oil wells.[4] As engineers seek to improve recovery from shale wells, innovation continues on several elements of the technology, including the use of longer fractures, greater use (per foot) of water and proppant, shorter stages and "micro-perforations", and improved identification of naturally existing fractures through higher-resolution micro-seismic mapping. Designing fracturing fluid for optimal performance is complementary to several of these elements, and represents a significant area of focus for oil and gas engineers, for advancement of shale production technology (Robart et al., 2013; Montgomery, 2013; Gulbis and Hodge, 2000).

In some cases, the quest for superior fracturing fluids has led engineers to consider the use of highly toxic chemicals (Stringfellow et al., 2014). Indeed, many of the chemicals used in fracturing fluid are known to be toxic to human health, or may cause damage to the ecosystem. These risks have raised concerns among environmental groups (Elgin et al., 2012; Haas et al., 2012). Early concerns about shale gas development were driven by the possibility that toxic fracturing fluid might migrate to or be accidentally released into ground

[3]Personal correspondence with M. Boling, Southwestern Energy, 2013.

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

water or surface water. The industry responded to these concerns by pointing to the small percentage of fracturing fluid that is actually comprised of substances other than water and sand.[5] Public concerns about risks to water sources were accentuated by the often close proximity of well-pads to residential and other non-industrial land uses, and by a few high-profile incidents of water pollution.[6] Media coverage of fracturing chemicals has highlighted the toxicity of some chemicals along with the industry's desire to maintain secrecy over the specific chemicals involved (Elgin et al., 2012; Haas et al., 2012).

2.2 Policy Tradeoffs of Information Disclosure

With respect to potential external impacts associated with the use of toxic chemicals, economic theory suggests a number of alternative approaches that public policy-makers might use to regulate such externalities. Command-and-control regulations require firms to undertake particular technologies or practices. When regulators have less information about production processes than do firms, this "one size fits all" approach may be suboptimal. Market-based regulations (e.g., severance taxes and impact fees) modify firms' incentives via price effects, but may affect firms only on the extensive margin, and may be difficult to implement if chemical releases are difficult to monitor. In contrast, information-based regulations require regulated entities to disclose elements of their production process that may have external impacts, but which would be difficult for outside stakeholders to ascertain without the disclosure requirement. Disclosure regulations are motivated both by the notion that the public has a "right to know" the details of firms' (and governments') production decisions, especially when those elements conceivably affect public welfare. In addition, disclosure regulations may also motivate regulated entities to change their behavior, either through pressure applied in the marketplace or the courts. The former means that information-based regulations are most commonly used in consumer-facing industries (Fung et al., 2007).

Scholars have posited alternative mechanisms to explain why mandatory disclosure laws may positively influence regulated actors even without a consumer channel. Bennear and

[5]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.

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

Olmstead (2008) identify a political mechanism: information may increase the ability of a concerned public to lobby for stronger regulation. These authors also note that disclosure may affect an organization's internal decision making, as individuals within the firm change their behavior as a result of measuring and reporting data. Another interpretation focuses on the role of liability, which is facilitated by information disclosure. Olmstead and Richardson (2014) list factors that work in favor of liability as a regulatory approach, including asymmetric information (i.e., regulated firms have better information about technologies) and limited ability to avoid payment (e.g., by spinning off risky activities into operations with little exposed assets.[7] If it is unlikely that a suit is brought, either because litigation or information costs are high, the disciplining force of liability will also be muted. Because shale gas development is a highly technical process that often occurs in areas with restricted public access, it may be difficult for the general public, and even regulators, to monitor operator activities. Providing information to nearby landowners and other interested parties reduces this cost.

Konschnik (2014) summarizes other reasons why disclosure is valuable, many of which are more direct. For instance, in the event of an accident, disclosure to emergency medical personnel and medical staff may improve treatment and protect the staff members themselves. In addition, disclosure provides information to nearby landowners and local government authorities so that they can test their water supplies, increasing their ability to bargain with operators. Disclosure can also help establish liability for contamination, making liability a more effective regulatory tool. If the information made available extends over a period of time, then disclosure can facilitate the monitoring of environmental releases, exposures, or health impacts over time. Finally, disclosure satisfies the public's "right-to-know" about possible release of, or exposure to, hazardous materials, in the spirit of the Emergency Planning and Community Right-to-Know Act (EPCRA) of 1986 (although the shale gas industry is currently exempt from some provisions of this Act).

At the same time, to the extent that industry actors rely on secrecy to maintain competitive advantage from investments in research and development, disclosure potentially

[7]Insurance and bonding requirements in Pennsylvania are small (either $2,500 per well or $25,000 for all wells in the state).

undermines incentives to invest in innovation. It is well understood that to the extent that new knowledge generated by investments in research and development can be copied or imitated by other firms that do not pay the full cost of the investment, the social returns to investments in R&D may exceed private returns. In theory, patents solve the problem of imperfect appropriability by granting inventors the right of exclusive use for a period of time (Cohen, 2010). In practice, however, the protection afforded by patents varies substantially across industries, and would-be innovators in many industries rely on secrecy, lead time, and investments in complementary assets in order to maximize the returns to innovative activity (Cohen et al., 2000; Cohen, 2010; Teece, 1986). A survey administered in 1994 to R&D laboratories in the US manufacturing sector indicated that on average, lab managers in the petroleum and chemicals industries considered secrecy more effective than any other mechanism, including patents, for protecting both product and process innovations (Cohen et al., 2000).[8] Thus, in this context, regulations that require companies to disclose information about their use of specific inputs may result in decreased private returns to innovation.[9]

2.3 Disclosure Rules

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

[8]This does not necessarily imply patents are unimportant, as their use may be complementary in an overall strategy to maintain competitive advantage.

[9]Note that much of the technological development that facilitated the "shale revolution" was based on R&D that was directly sponsored by the government, or by government-subsidized R&D in the form of tax credits for development of unconventional shale and tight gas. In response to persistent supply shortages in the 1970s, the US government began to directly and indirectly fund R&D investment in the natural gas industry. The Natural Gas Policy Act of 1978 removed wellhead price controls and provided tax incentives for developing new natural gas resources. The Crude Oil Windfall Profit Tax Act of 1980 provided tax credits for developing unconventional fuels, which increased their financial return and reduced their risk. The Department of Energy (DOE) initiated a number of Unconventional Natural Gas Research Programs, including the Eastern Gas Shales Program, the Western Gas Sands Program, the Methane Recovery from Coalbeds Program, the Seismic Technology Program, and the Drilling, Completion and Stimulation Program. While these programs did not have a role in the development of horizontal drilling or 3D seismic imaging technologies, they did play an important role in high-volume fracturing and micro-seismic fracture mapping (Shellenberger et al., 2012; Wang and Krupnick, 2013).

There is less uniformity in terms of where the information must be registered. Five of the eighteen states, including several of the states that passed the earliest disclosure rules, require operators to report to a state regulatory agency or commission. Six require that operators report to FracFocus, an online database created by a multi-state commission in partnership with a non-profit organization (Council and Commission, 2015). When uploading information to FracFocus, operators are also asked to provide information about well location and characteristics including vertical depth, volume of water used, latitude and longitude, and well name. The seven remaining states allow operators to choose their reporting location (i.e., to FracFocus or the state) (Fetter, 2016).[10]

States have adopted similar approaches to accommodating the need for trade secrets, partly due to the Uniform Trade Secrets Act.[11] In particular, all states allow exemptions for the disclosure of additives considered to be confidential business information that firms believe gives them a competitive advantage. Operators must declare an exemption for individual chemical ingredients for which they claim trade secret status. This is accommodated by FracFocus, which allows for uploaded information to include the concentration of the chemical used but not its name or chemical identification number. Some states also require operators to report the chemical family to which the proprietary substance belongs.

3 Data

We obtained production data for oil and gas wells in Pennsylvania from DrillingInfo, a national provider of information on the oil and gas industry. We limit our analysis to wells in unconventional reservoirs, as identified by the Pennsylvania Department of Environmental Protection (PADEP). This is consistent with our focus on innovation within unconventional shale development: 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

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

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

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.

We identified 7,028 unconventional wells in Pennsylvania that reported initial production between January 2007 and December 2015. We collect data on inputs and operating parameters from two sources: Well Completion Reports and Stimulation Fluid Additive reports from PADEP, and the FracFocus database. Well Completion Reports, which operators must submit within 30 calendar days following completion, contain firm identifying information, well location, and information about the perforation and stimulation process. The requirement to submit Well Completion Reports dates back to 1989.[12] Effective in February 2011, operators were also required to submit information on chemicals used in the stimulation process, including the name and concentration of each chemical additive in the fracturing fluid. Operators were instructed to submit information about chemicals either along with the Well Completion Report or on a separate DEP form, the Stimulation Fluid Additive report.

Some Pennsylvania operators elected to submit chemical additive information to the national FracFocus registry. Operators from other states (not just Pennsylvania) were uploading chemical additive information to FracFocus at the same time, typically on a voluntary basis.[13] FracFocus permitted operators to upload data in a standardized format, and the template contained the same information that operators had to report under Pennsylvania's 2011 disclosure law. Thus the operators who uploaded chemical additive information to FracFocus were in compliance with the reporting law.[14] In April 2012, Pennsylvania amended its reporting regulation to require operators to upload chemical additive information to FracFocus, in place of the requirement to submit information to the DEP on Well Completion

[12]http://www.pacode.com/secure/data/025/chapter78/s78.122.html

[13]Operators in Wyoming and Arkansas were required to disclose chemical additive information as of September 2010 and January 2011, respectively; however, both states provided their own reporting websites for this purpose. Later in 2011 and 2012 several other states passed laws requiring chemical disclosure. Some of those laws required operators to upload reports to FracFocus, and others offered operators the option to use FracFocus or a state registry.

[14]There may have been confusion on this point at the time; some operators, evidently uncertain how DEP would enforce the requirement, provided printouts of their FracFocus disclosure forms along with their Well Completion Reports. Other operators submitted disclosures to FracFocus but not in the Well Completion Reports.

Reports.

Thus, we obtained chemical additive information from both Well Completion Reports and associated Stimulation Fluid Additive reports, or from the FracFocus database. Notably, information that operators submitted to FracFocus was more readily observable by competitors and the public.[15] By contrast, Well Completion Reports and Stimulation Fluid Additive reports submitted to the DEP were available for review by a subscription service or by in-person review at regional DEP offices. Subscribers could view Well Completion Reports through the Exploration and Development Well Information Network (EDWIN), in which some chemical disclosure forms were available as scanned PDF documents.[16] However, there was a long wait time for reports to be scanned and uploaded to this system, especially in the height of the fracturing boom.[17] Members of the public could also review reports in person, but would have had to identify the permit number of a specific well, contact the appropriate regional DEP office, file a request, schedule an appointment to visit in person (typically three to four weeks in advance), and would then be allowed to review a limited number of hard copy documents onsite (on the order of 25 per day). Therefore we do not believe operators would have expected others to observe their fluid contents, prior to the April 2012 rule that required public disclosure of chemical information on FracFocus.

To capture productivity, we use the standard industry metric of initial gas output per foot of wellbore. Total output is highly correlated with initial output, 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 1,158 wells, so we cannot use these in the analysis.[18] We drop an additional 94 wells that have nonsensical completion dates (the recorded completion date

[15]When FracFocus was launched, GWPC and IOGCC provided fracturing fluid chemical reports as individual PDF files that were easy to download individually but challenging to compile en masse. At the time, GWPC and IOGCC specifically stated their intent to provide a forum where the public could view individual reports but not look at many reports at once. At least two entities had successfully scraped the entire FracFocus database by late 2012. One was a for-profit consulting firm motivated by commercial interest in the database; the other was an environmental NGO that claimed, on right-to-know grounds, that the public should be able to review and compare information across wells (Skytruth, 2013).

[16]Until 2015, EDWIN was known as the Integrated Records and Information System (IRIS).

[17]One of the authors accessed the system several times in 2012-13 and found that the wait time for stimulation fluid additive information to be available in the system was highly variable, and the delay for its addition could be up to 18 months or more after well completion.

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

is after the date of initial production). Finally, we drop 21 wells with zero recorded gas production. This leaves us with a sample of 5,755 wells.[19]

For each of these wells we observe identifying information (operator, location, and completion date) and output. We collected data on inputs, including volume of water, quantity of proppant, and chemical additives to the stimulation fluid, from DEP reports and FracFocus. The DEP reports proved unamenable to optical character recognition scraping: they featured to a wide variety of formats (over ten different formats with different page headers), the use of numbers that were sometimes handwritten or crossed out and overwritten, overlaid date stamps, and raster images. Thus the relevant information was digitized manually, by a team of data entry contractors. This effort took about 1,800 person-hours over 4 months, and involved the entry of about 200 data items per report. We verified the quality of data entry by using standard procedures including systematic checks for consistency and reasonableness, spot checks comparing hand entries to the original reports, and comparison of duplicate entries by different contractors.

Despite our careful collection and entry of input data, not all variables are available for each well. Fluid volume is available for 4,353 wells (76%); proppant volume is available for 4,316 wells (75%); and both fluid and proppant quantity are available for 3,184 wells (55%). We have detailed chemical information for 4,290 wells, including 3,982 wells with chemical information disclosed to FracFocus and 308 wells (from the period February 2011 - April 2012) with chemical information disclosed only to PADEP.[20] Of the 1,465 wells with no information on chemicals, 1,250 were not subject to disclosure; 215 wells appear to be out of compliance, as they were fractured after February 2011 but chemical additive information was not released to PADEP nor FracFocus.[21] We were able to obtain other information on

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

[20] Mandatory disclosure started in Pennsylvania in 2011, but prior to this date some operators voluntarily submitted chemical information for some wells.

[21] 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. Assuming the reports were provided to DEP within 30 days of well completion, this would imply a delay of 4 to 5 years from when DEP received the reports to entry of the report into the system. While such a delay may seem unlikely, we did note that approximately 800 Well Completion Reports, some pertaining to wells completed in 2011 or early 2012, were uploaded into the system between December 2015 and January 2017.

other well parameters for a larger set of wells; we have information on the number of stages, for instance, for 4,699 wells (82%).

Table 1 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 available only for wells starting in 2010. 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 1: Summary Information for Wells in Sample with Chemicals Data

	Fluid Volume		# Chemicals / Well		First 6 Month Gas per foot	
Year	Median	Mean	Median	Mean	Mean	Median
2010	3.8	4.1	11.5	10.7	160.2	141.2
2011	4.4	4.5	9	10	154.7	122.8
2012	4.2	4.4	11	12.5	186.5	131.5
2013	5.6	5.9	13	15.4	198.1	153.9
2014	7.8	8.3	13	15.6	200.6	159.9
2015	8.9	9.1	13	14.3	151.3	128.1

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 UNITS HERE

Table 2 provides additional information for wells in the sample for which we observe chemical use. This table demonstrates that operators continued to innovate by introducing new chemicals each year, with the largest innovation (in terms of number of new chemicals) occurring in 2011-2012—right around the time that the state was issuing laws requiring mandatory disclosure. Companies also retired some chemicals from use (at least within our sample's timespan).

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, enable social learning. Second, did the learning enabled by

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

Year	# Wells	# Operators	Chemicals # New	Chemicals # Retiring	Chemicals # Total
2010	38	8	51	3	51
2011	881	24	87	17	132
2012	1055	27	92	20	199
2013	961	31	62	66	249
2014	897	27	43	121	223
2015	148	17	4	–	112

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.

disclosures have value, helping less productive operators to "catch up" to more productive operators. In future work, we plan to also analyze the extent to which this may have reduced operators' incentives to innovate.

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

If the answer to both questions is affirmative, we would expect to see convergence in productivity across operators, with less able firms using information revealed by the disclosure laws to catch up with more successful firms. After confirming that this is the case, we examine each of the two questions more closely to rule out alternative explanations. To answer the first question, we examine how the design of fracturing fluids changed after disclosure. After finding a convergence in the chemicals used consistent with copying from disclosure, we turn to the second question. To answer it, we construct a two-stage estimator that combines information on chemical formulas and well productivity. We find evidence that convergence is

indeed explained specifically by weaker firms adopting the chemical mixtures used by more successful firms in previously drilled wells. This finding is consistent with firms learning through chemical disclosure. Finally, we conduct several robustness and placebo tests to rule out alternative explanations.

4.1 Change in Well Productivity

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

$$g_w = \beta_0 + \beta_1 s_w + \beta_2 f_w + \epsilon_w, \tag{1}$$

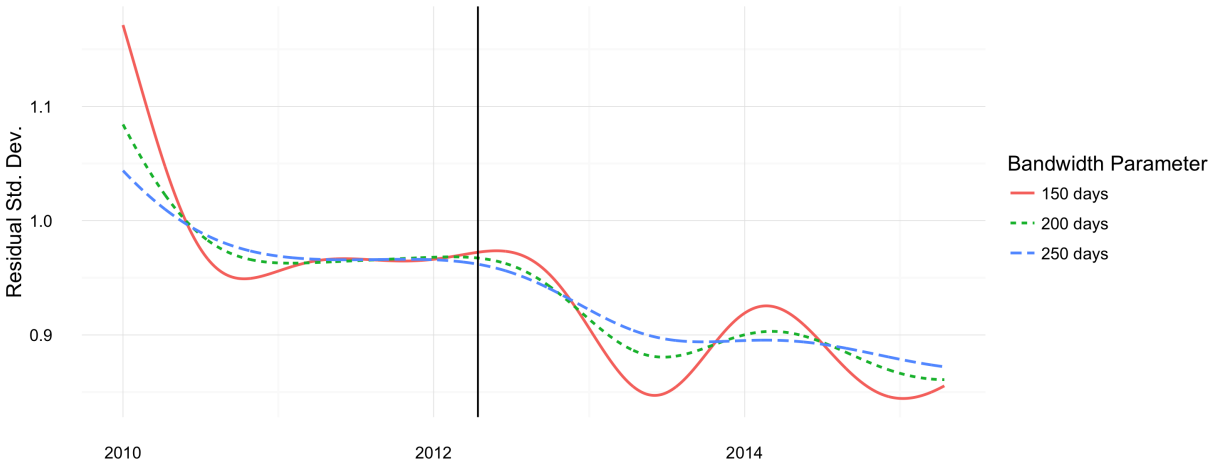
where the w subscript indicates a well, g denotes per-foot first-six-month gas production, and s and f are fracturing proppant and fluid volumes. We then take the estimated residuals $\hat{\epsilon}_w$ and calculate a Gaussian kernel-weighted standard deviation over time. The results are shown in Figure 1. The curves plot the estimated standard deviation, for various bandwidth choices; the vertical black line marks April 2012, the date that disclosure came into effect. All three bandwidths show a marked fall in the standard deviation of the estimated residuals after disclosure. This fact motivates the rest of our analysis that follows.

4.2 Well-to-Well Similarity Measures

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

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

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



The curves indicate the kernel-weighted estimated residual standard deviation over time, with different bandwidths. The vertical line indicates the date that public disclosure came into effect.

we observe.

We begin by introducing some notation, defining our similarity metric of choice, and providing a summary of the metric in our data.

Define $s_{ij} \in [0, 1]$ as the pairwise similarity between wells i and j.[22] A value $s_{ij} = 1$ implies that the hydraulic fracturing fluids used in wells i and j are indistinguishable according to the chosen metric; $i = j \Rightarrow s_{ij} = 1$.

We now introduce some notation to define the similarity metrics. Let $\mathcal{C}$ denote the set of all possible chemicals, and let $x_{ic} \in [0, 1]$ be the concentration of chemical $c \in \mathcal{C}$ in well i. So the sum of all such contributions $\sum_{\mathcal{C}} x_{ic} = 1$. We also define the binary variable $y_{ic} = \mathbb{1}\{x_{ic} > 0\}$. Then we can define the following quantities for a given (i, j) pair:

- $a_{ij} = \sum_{\mathcal{C}} y_{ic}$ is the number of chemicals found in well i
- $b_{ij} = \sum_{\mathcal{C}} y_{jc}$ is the number of chemicals found in well j
- $c_{ij} = \sum_{\mathcal{C}} y_{ic} y_{jc}$ is the number of chemicals found in both wells i and j

In principle we can calculate "abundance" or "correlation" similarity metrics that take

[22] Most metrics, including those we use, feature the property $s_{ij} \in [0, 1]$, although other measures like correlation, with support $s_{ij} \in [-1, 1]$, are possible.

into account the more detailed information in x_{ic}. Two features of our setting argue against taking this approach in practice. First, chemical quantities are reported as maxima rather than actual concentrations; second, FracFocus and the PADEP reports differ in terms of recording concentrations in terms of mass and volume. Due to these limitations, we focus on "binary" metrics that consider only the absence or presence y_{ic} of chemical c from well i. We proceed with the Sorensen binary metric, but find similar results with the alternative Jaccard binary metric.[23,24]

The Sorensen binary metric is defined as:

$$s_{ij} \equiv \frac{2c_{ij}}{a_{ij} + b_{ij}}.$$

4.3 Disclosure and Similarities

We calculate the similarity metric for each i, j well-pair in our data. Our chemical data on 4,015 wells gives us 8 million such pairs. Figure 2 shows the distribution of these measures for two sub-samples of pairs: those where both wells are drilled by the same operator, and those where the wells are drilled by different operators. The figure also plots the medians of the two distributions. A few observations are noteworthy: the different-operator distribution has a mass point at $s_{ij} = 0$, with most of the mass roughly centered around $s_{ij} = 0.25$. While the same-operator distribution also has most of its mass below $s_{ij} = 0.5$, it is noticeably more uniform and has a mass point at $s_{ij} = 1$. This confirms what we might have expected: intra-operator well-pairs tend to use more similar chemical mixes than inter-operator well-pairs.

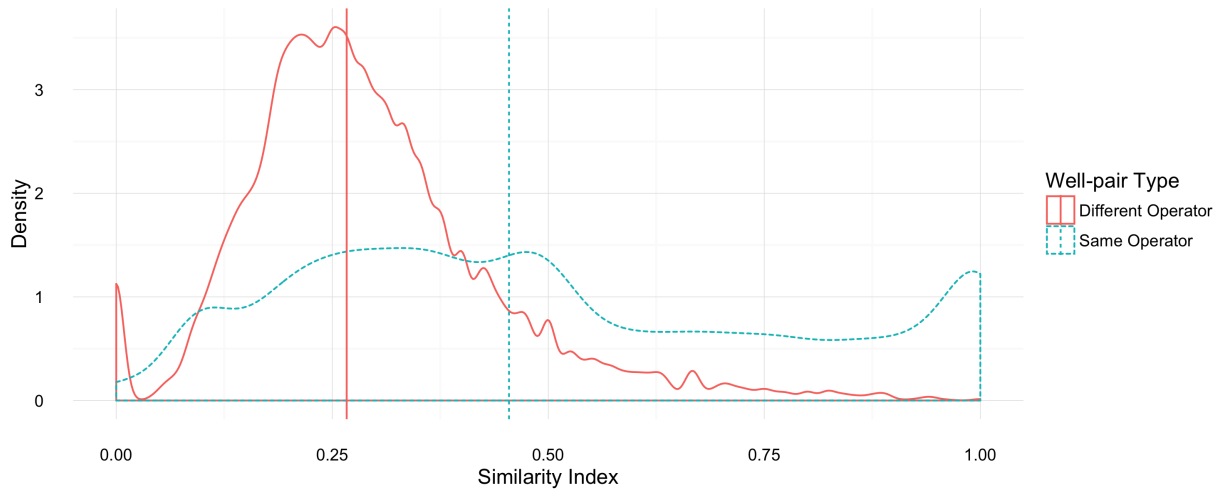
We next consider how well-pair similarities have changed with the advent of disclosure. We do this by running a series of regressions of the following form:

$$s_{ij} = \beta_0 + \beta_1 oper_{ij} + \beta_2 post_{ij} + \beta_3 dist25_{ij} + \beta_4 oper_{ij} * post_{ij} + \beta_5 dist25_{ij} * post_{ij} + \alpha_i + \alpha_j + \epsilon_{ij}, \quad (2)$$

[23] See chapter 12 of Krebs (2014).

[24] The Jaccard binary metric is defined as: $s_{ij} \equiv c_{ij}/(a_{ij} + b_{ij} - c_{ij})$.

Figure 2: Sorensen Similarities by Same / Different Operator Status



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

where s_{ij} is the Sorensen similarity metric defined above, $oper_{ij}$ is a binary variable equal to 1 if wells i and j were drilled by different operators, $post_{ij}$ is a binary variable equal to 1 if either well i or j was fractured in the post-disclosure period, and $dist25_{ij}$ is a binary variable indicating whether the wells are located more than 25km apart. Operator fixed effects are included with α_i and α_j. The results of regressions featuring subsets of these variables are shown in Table 3.

A few things are worth noting from Table 3. First, all columns show a significantly negative estimate of the coefficient on $oper_{ij}$: this indicates that well-pairs with different operators tend to use less-similar chemical mixes, confirming the graph in Figure 2. Second, the coefficient of $post_{ij}$ is also always estimated to be significantly negative, indicating that well-pairs use less-similar chemical mixes on average if at least one of the wells is drilled in the post-disclosure period. Third, the coefficient on $dist25_{ij}$ is estimated to be negative, indicating that more geographically distant well-pairs use less-similar chemical mixes on average.[25]

The interaction terms are more interesting. The interaction between $dist25_{ij}$ and $post_{ij}$ is estimated to be positive and significant, indicating that after disclosure, distant wells

[25]Similar results hold for other cut-off distances, or for letting distance enter the regression directly.

use more similar chemical mixes than before disclosure. Similarly, the interaction between $oper_{ij}$ and $post_{ij}$ is estimated to be positive and statistically significant. This indicates that different-operator well-pairs use more similar chemical mixes post-disclosure than pre-disclosure. This is consistent with disclosure facilitating the transfer of knowledge about chemical mixes, and motivates our further analysis in the next section.

Table 3: Disclosure and well-pair similarities

	Similarity Index				
	(1)	(2)	(3)	(4)	(5)
Different Operator	−0.227***	−0.220***	−0.322***	−0.314***	−0.302***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
Post	−0.024***	−0.023***	−0.117***	−0.115***	−0.123***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
Different Operator x Post			0.101***	0.099***	0.086***
			(0.001)	(0.001)	(0.001)
Dist > 25km		−0.014***		−0.014***	−0.035***
		(0.000)		(0.000)	(0.001)
Dist > 25km x Post					0.022***
					(0.001)
Operator FEs	✓	✓	✓	✓	✓
Observations	8,058,105	8,058,105	8,058,105	8,058,105	8,058,105
R^2	0.183	0.183	0.184	0.184	0.185

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.
The similarity index is the Sorensen index, described in detail in Section 4.2. Post is a binary variable, indicating whether *at least* one of the wells in the well-pair was fractured after disclosure.

4.4 Convergence in Output

4.4.1 First stage regressions

To evaluate the effect of changes in chemicals and other inputs on productivity, we set up a two-stage regression framework. In the first stage, we regress gas production on a set of observables and fixed effects. The fixed effects include operator by period fixed effects, which represent the relative productivity of each operator in each period, conditional on observables. The difference between the first-period and second-period fixed effects, in turn,

represents an operator-level measure of change in average productivity from the first to the second period. In the second stage, we calculate the difference between first- and second-period fixed effects and regress this difference on the first-period fixed effect, along with a "quality-similarity" measure that is explained in section 4.4.2 below. The second-stage regression is similar in spirit to simple tests of convergence in the economic growth literature (e.g., Bernard and Durlauf (1996)). As we shall see, our results suggest there is convergence among operators from the first to the second period, but this convergence is driven strongly by the fact that after disclosure, ex-ante lower-performing operators use chemical formulas that are more similar to those used by ex-ante higher-performing wells.

In our first stage regression, we regress per-foot initial gas production on a set of observables and fixed effects, using the form

$$y = F \times g(X)$$

where F is an operator fixed effect and $g(X)$ is a function of variables X that affect the well in question. Suppose $g(X) = e^{X'\beta}$. Taking logs on both sides, this implies that

$$\log y = \log F + X'\beta$$

In our preferred specification, the function $g(X)$ includes a township fixed effect, a year fixed effect, the density of unconventional wells previously completed in the township (as a quadratic), and the current output price.[26] The year fixed effect should absorb secular technological change, and the township fixed effect helps to control for spatial differences in resource quality.[27] We explored several alternative specifications, including specifications without price, without prior well density, with alternative measures of prior well density, and

[26]For output prices, we use prices for the Henry Hub exchange for the month in which well i was completed. City gate prices for Marcellus production have often been lower than Henry Hub prices (http://www.eia.gov/todayinenergy/detail.php?id=24712), but we have not been able to obtain a continuous price series for Dominion South or other exchanges that may be more appropriate to the Marcellus. In any case the Henry Hub price is likely highly correlated with these other exchanges; thus, to the extent that the main point of our first stage analysis is to recover operator-period fixed effects, our use of the Henry Hub prices likely does not introduce significant error.

[27]Alternatively, we could perform a semiparametric regression that conditions flexibly over space. However, Covert (2015) implemented such a procedure in a related setting—shale oil production in North Dakota—and found the results were not substantially different from those using fixed effects for administrative jurisdictions.

with well density (or alternative measures) in linear form rather than quadratic; all of these alternatives produced results similar to those reported here.

We estimate the first stage regression using the log of first 18 month gas output per foot for the dependent variable, and then exponentiate the fixed effects (to transform them from lnF back into F). We interpret these exponentiated effects as the contribution of the firm—a scale parameter, or multiplicative effect—that is positive by construction.

Table 4 shows the results of the first stage regression, with initial gas per foot as the dependent variable. Gas price appears to be positively correlated with initial productivity, consistent with other research showing a positive elasticity of production with respect to price (Newell et al., 2016).

Table 4: Results of first stage regression

	Log First 18 Months Gas Production
Gas Price	0.078***
	(0.017)
Prior Well Density	0.000
	(0.001)
Prior Well Density (squared)	−0.000
	(0.000)
Township FE	✓
Year FE	✓
Operator-period FE	✓
Observations	5,755
R^2	0.988

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.
This regression omits the constant in order to calculate a fixed effect for each operator-period; therefore, the R^2 value should not be given the usual interpretation.

4.4.2 Quality-Similarity Metric

To investigate the relationship between changes in productivity and the convergence in chemical mixes across different-operator well-pairs shown in Table 3, we define a quality-similarity index, QS_f for each firm f. It is designed to capture the quality of an operator's chemical

matching, where quality is defined by estimated pre-disclosure fixed effects. It is constructed as follows:

$$
\begin{aligned}
QS_f &\equiv \frac{1}{|M_f|} \sum_{m \in M_f} s_m(\theta_m^{PRE} + \hat{\epsilon}_m) \\
M_f &\equiv \{i, j \mid i \text{ is } f \text{ POST well, } j \text{ is non-}f \text{ POST well, } j \text{ fractured before } i\} \\
s_m &\equiv \text{similarity index for well-pair } m \\
\theta_m^{PRE} &= \text{estimated PRE period fixed effect for well } j\text{'s firm.} \\
\hat{\epsilon}_m &= \text{(exponentiated) first-stage residual for well } j.
\end{aligned}
$$

QS_f is thus a quality-weighted measure of the post-period similarity between f's wells and other firms' wells. We require that well j was stimulated before well i to ensure the possibility that firm f would have had the opportunity to view the chemicals used in well j and adjust its mix for well i if it chose.

4.4.3 Quality-Similarity Difference-in-Differences Test

With calculated values of QS, we proceed to the second stage regression:

$$\theta_f^{POST} - \theta_f^{PRE} = \beta_0 + \beta_1 \theta_f^{PRE} + \beta_2 QS_f + \beta_3 QS_f \theta_f^{PRE} + \epsilon_f. \tag{3}$$

Because the "data" for this regression features estimation error from the first stage, we calculate standard errors via a bootstrap. We use a modified bootstrap to ensure that we do not lose power in the second stage, which takes place at the operator-period level. To get our bootstrap sample, we first calculate the number of wells fractured within each operator-period. Then we draw, with replacement, that number of observations from that operator-period. This method, which is in the spirit of a block or panel bootstrap (Cameron and Trivedi, 2005), ensures that we do not lose any second stage observations due to the random draws inherent in the bootstrap routine. Unless noted otherwise, all second-stage results presented below use this bootstrap routine with 200 bootstrap replications.

The results can be seen in Table 5. The first column shows the regression when only θ_f^{PRE}

is included; the coefficient is found to be significantly negative, which suggests those with lower values of θ_f in the first period experience greater growth in their θ_f with the advent of disclosure – i.e., there is convergence in the values of θ_f over time. The second column shows estimates from Equation 3, with the addition of QS_f and the interaction term. With these additions, the point estimate of β_1 has flipped to be positive, suggesting divergence in θ_f over time. The point estimate of β_2 is also positive and statistically insignificant, suggesting that a higher QS_f is associated with higher growth in θ_f. Finally, β_3 is estimated to be significantly negative. This suggests that when well similarities are accounted for, the convergence of θ_f over time is due not to θ_f^{PRE}, but rather to the interaction term.

Table 5: Second Stage Results

	$\theta^{POST} - \theta^{PRE}$	
	(1)	(2)
θ^{PRE}	−0.289	2.582**
	(0.243)	(1.018)
QS		2.335
		(2.597)
θ^{PRE} x QS		−4.501**
		(2.177)
Constant	0.881***	−1.473
	(0.237)	(1.431)
Observations	28	24
R^2	0.002	0.346

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.
Standard errors are calculated using 200 bootstrapped samples, collected with binning at the operator-period level. See text for details.

4.4.4 Selection and Exiting Firms

The results reported above support the notion that convergence occurred over the period we study, but that convergence is driven primarily by the interaction of θ_f^{PRE} with QS_f. However, a substantial number of operators exited Pennsylvania in the POST period (i.e., did not drill any new wellbores). We must therefore consider the extent to which selection

affects the results. Would we find a different relationship if these firms had continued to operate in the Pennsylvania shale fields?

To address this concern, we assign values of θ_f^{POST} and QS_f to the exiting firms, re-run the second-stage regression in Equation 3, and check for stability of results. We use several alternative assumptions in an effort to test for robustness under a range of plausible but conservative scenarios.

At first glance, it might appear that an assumption that all exiting firms would have performed at the top of the distribution in the POST period would provide a conservative test. However, this turns out not to be the case. Instead, such an assumption implies that a number of poor performers (in the PRE period) all improved their performance in the POST period, which would lend support to the convergence hypothesis. This can be seen visually in Figure 3, which shows the relative magnitudes of the first-period and second-period operator fixed effects (abstracting for a moment from the role of QS_f). Firms that appear below and to the right of the 45-degree line improved their average performance from period 1 to period 2 ("laggards caught up"), whereas firms that appear above and to the left suffered decreases in performance ("leaders fell behind"). If all exiting firms were assigned very high values of θ^{POST} regardless of their θ^{PRE} value, this would place all exiting firms below and to the right of the 45-degree line, which would stack the scales in favor of convergence.

Instead, we consider the following four alternative scenarios:

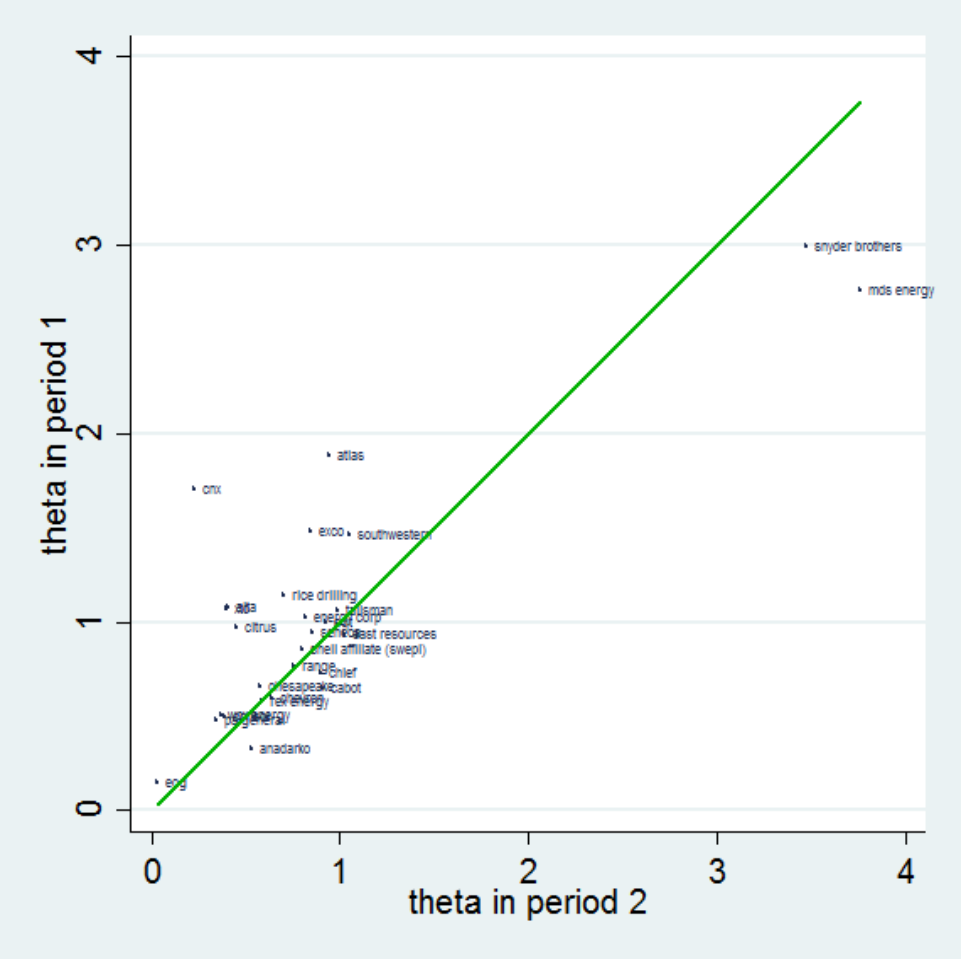
1. Assume exiting firms would have performed at the 25th percentile of the firms that did continue to operate

2. Assume exiting firms would have performed at the median of the firms that did continue to operate

3. Assume exiting firms in the bottom quartile of the distribution in the pre period would have performed at the bottom of the distribution in the post period; assume firms in the top three quartiles in the PRE period would have performed at the top of the distribution in the POST period

4. Assume exiting firms in the bottom half of the distribution in the PRE period would have performed at the bottom of the distribution in the POST period; assume firms in

Figure 3: Operator Fixed Effects in Period 1 and Period 2



the top half of the distribution in the PRE period would have performed at the mean of the distribution in the POST period

We must also assign values for QS_f for the firms that exited. We make three alternative assumptions:

1. Assume exiting firms would have had a QS_f at the 25th percentile of the firms that remained.

2. Assume exiting firms would have had a QS_f at the median of the firms that remained.

3. Assume exiting firms would have had a QS_f at the 75th percentile of the firms that remained.

It is not clear *a priori* whether a relatively high or low assignment of QS_f should make for a conservative test of its role in achieving convergence, so using a range of alternative assumptions allows us to test across various alternative states of the world. We run distinct regressions that match each of the four assumptions for θ_f^{POST} against each of the three assumptions for QS_f (plus an additional set of regressions that considers a simple convergence without QS_f). Table 6 provides a summary of the results for these tests. Note that the table contains the results of 16 regressions; each column and set of rows (between the horizontal lines) represents a distinct regression.

The table shows that this analysis is consistent with the main analysis, implying that selection on surviving firms is not driving our results. This supports our earlier finding that convergence is driven primarily through the interaction of the QS_f term and the θ^{PRE} term, which is consistent with disclosure-enabled "catching up".

4.5 Placebo Tests

In addition to our attempt to combat potential sample selection in the prior section, we recreate this analysis with two alternative definitions of QS_f to check the robustness of our findings. In the first alternative, QS_f is generated using all post well-pairs with firm f's wells, regardless of the chronological order of the stimulation. In the second alternative, we create a version that includes a geographic limitation in order to rule out spatial spillovers.

Table 6: Results of QS-DD Tests with Alternative Assumptions for Exiting Firms

Assumption for θ^{POST}		No QS (simple convergence)	25th Percentile QS	Median QS	75th Percentile QS
25th Percentile	θ^{PRE}	-0.254^{**}	1.522	3.397^{***}	4.351^{***}
		(0.119)	(1.231)	(1.135)	(0.812)
	QS		2.955	5.23^{**}	6.014^{***}
			(2.188)	(2.014)	(1.460)
	$QS \times \theta^{PRE}$		-3.160	-6.462^{***}	-7.944^{***}
			(2.243)	(2.023)	(1.402)
Median	θ^{PRE}	-0.287^{**}	1.409	3.177^{***}	4.136^{***}
		(0.112)	(1.119)	(1.022)	(0.727)
	QS		1.961	4.584^{**}	6.082^{***}
			(1.989)	(1.814)	(1.306)
	$QS \times \theta^{PRE}$		-3.014	-6.124^{***}	-7.610^{***}
			(2.040)	(1.822)	(1.255)
Bottom Or Top	θ^{PRE}	-0.178	2.950	0.591	-1.613
		(0.340)	(3.367)	(3.609)	(3.213)
	QS		-2.959	-3.700	-2.654
			(5.982)	(6.407)	(5.771)
	$QS \times \theta^{PRE}$		-4.966	-0.621	3.310
			(6.135)	(6.436)	(5.546)
Bottom Or Mean	θ^{PRE}	-0.069	2.014^{*}	3.006^{***}	3.188^{***}
		(0.110)	(1.062)	(1.027)	(0.844)
	QS		4.151^{**}	5.087^{***}	4.562^{***}
			(1.887)	(1.823)	(1.516)
	$QS \times \theta^{PRE}$		-3.704^{*}	-5.407^{***}	-5.577^{***}
			(1.935)	(1.831)	(1.456)
N		44	44	44	44
N (exit)		20	20	20	20

*** p$<$ 0.01, ** p$<$ 0.05, * p$<$ 0.10.
All QS is standard Sorensen. All regressions include a constant term. Growth rate is $\theta^{POST} - \theta^{PRE}$. Significance tests and standard errors do not reflect the bootstrap procedure.

In the first alternative, the estimated coefficient on the interaction of θ^{PRE} and QS_f is not significant, which is in contrast to our finding with the proper definition of QS_f. This contrast supports the hypothesis that disclosure in fact aided learning. The alternative with a geographic restriction has similar results to the original, suggesting that convergence is not due to spatial proximity.

4.5.1 "Time-Fluid QS"

For the first alternative, we define a new quality-similarity index QS_f^{TF}:

$$QS_f^{TF} \equiv \frac{1}{|M_f^{TF}|} \sum_{m \in M_f^{TF}} s_m \theta_m^{PRE}$$

$$M_f^{TF} \equiv \{i, j \mid i \text{ is } f \text{ POST well, } j \text{ is non-}f \text{ POST well}\}$$

Note that s_m and θ_m^{PRE} are defined as before, so the only change is the set of well-pairs used, M_f^{TF}. In the original definition, we imposed the requirement that f's well was completed after the other operator's well, to allow for the possibility that f was able to view the chemical disclosures of the non-f well. In this alternative, we do away with this limitation and construct QS_f^{TF} as though we had no information on the relative timing of the wells. If the learning through disclosure mechanism is operative, we expect that the analysis with the new measure QS_f^{TF} will result in less significantly negative estimate of β_3.

The results of the new second stage regression are shown in column 3 of Table 7. While the estimates are significantly different from zero, the coefficient on the interaction term is positive, suggesting that the QS^{TF} is contributing to divergence. This result supports the conclusion that convergence is driven by learning through disclosure: including matches with wells drilled after f's wells in M_f^{TF} weakens the explanatory power of the regression.

4.5.2 Distance Buffer

The second alternative is designed to rule out an alternative hypothesis: that well-pairs are becoming more similar due to learning spillovers that are limited to, or primarily driven by, spatial proximity (Conley and Udry, 2010). To test this possibility, we construct a final

Table 7: Second Stage Results with Alternative QS Definitions

	$\theta^{POST} - \theta^{PRE}$			
	No QS	QS	Time-Fluid QS	Distance QS
	(1)	(2)	(3)	(4)
θ^{PRE}	−0.289	2.582**	−11.137***	2.775***
	(0.243)	(1.018)	(1.013)	(0.934)
QS		2.335	−34.060***	2.495
		(2.597)	(2.483)	(2.381)
θ^{PRE} x QS		−4.501**	23.032***	−4.974**
		(2.177)	(2.280)	(2.100)
Constant	0.881***	−1.473	16.505***	−1.505
	(0.237)	(1.431)	(1.333)	(1.405)
Observations	28	24	28	24
R^2	0.002	0.346	0.318	0.461

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.
Standard errors are calculated using 200 bootstrapped samples, collected with binning at the operator-period level. See text for details.

quality-similarity index QS_f^{DIST}:

$$QS_f^{DIST} \equiv \frac{1}{|M_f^{DIST}|} \sum_{m \in M_f^{DIST}} s_m \theta_m^{PRE}$$

$$M_f^{DIST} \equiv \{i, j \mid i \text{ is } f \text{ POST well, } j \text{ is non-}f \text{ POST well, } i \text{ and } j \text{ at least 25km apart}\}$$

In this alternative, M_f^{DIST} is defined to include only more distant well-pairs. The second stage results are in the final column of Table 7. The point estimates are almost identical to those with the original QS_f measure, and the interaction term is still significantly negative. The similarity of the estimates using QS_f^{DIST} and QS_f rules out spatial spillovers as the source of learning, and thereby lends support to learning through disclosures.

4.6 Mechanism: Learning Through Contractors

One possible channel for the transfer of information about chemical mixtures is the contractors who are hired to hydraulically fracture the wells. Contractors perform a variety of roles;

operators hire them to assist with jobs that may include drilling, cementing, well logging operations, and other tasks, as well as designing and conducting the fracturing job. The PADEP requires operators to provide information about contractors on Well Completion Reports. According to these reports, operators hired 8 contractors on the median well in our sample to assist in various roles; for some wells, operators hired up to 40 contractors. However, the roles these contractors play are not always specified. We were able to identify the fracturing contractor for 3,153 wells. This includes 2,558 wells for which the contractor job was specified as involving stimulation or fracturing (distinct from related tasks such as perforation), and 595 wells in which an operator hired a contractor that exclusively performs such services.

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 at least one of the wells is in the post-disclosure period. We consider only inter-operator well-pairs.

$$s_{ij} = \beta_0 + \beta_1 contractor_{ij} + \beta_2 post_{ij} + \beta_3 contractor_{ij} * post_{ij} + \epsilon_{ij} \tag{4}$$

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). The results of these regressions are shown in Table 8.

When it is the only non-constant regressor, the dummy for sharing a contractor is associated with an increase in the similarity index of 0.272. In the second column, the dummy for post-disclosure is included: this dummy leaves the contractor coefficient statistically unchanged, enters negatively and significantly, but is smaller in magnitude. The third column adds an interaction between these two dummies. The coefficient for same contractor increases, the coefficient on POST becomes even smaller, 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 third in the post-disclosure period, suggesting that operators are not as reliant on contractors as a source of 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 effective in the post-disclosure period.

Table 8: Determinants of Well-pair Similarities

	Similarity Index		
	(1)	(2)	(3)
Same Contractor	0.272***	0.271***	0.394***
	(0.005)	(0.005)	(0.020)
Post		−0.077***	−0.057***
		(0.007)	(0.006)
Contractor x Post			−0.126***
			(0.020)
Constant	0.289***	0.366***	0.345***
	(0.002)	(0.007)	(0.006)
Observations	3,544,453	3,544,453	3,544,453
R^2	0.216	0.219	0.220

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.
Results are from the regression described in 4; similarities are from the subsample where the fracturing contractor can be identified with confidence, see Section 4.6 for details.
Standard errors are clustered at the well-index level, i and j of s_{ij}.

5 Discussions and Conclusions

References

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.

Bennear, Lori Snyder and Cary Coglianese, "Measuring Progress: Program Evaluation of Environmental Policies," *Environment: Science and Policy for Sustainable Development*, 2005, *47* (2), 22–39.

Bernard, Andrew B. and Steven N. Durlauf, "Interpreting tests of the convergence hypothesis," *Journal of Econometrics*, March 1996, *71* (1-2), 161–173.

Cameron, A. Colin and Pravin K. Trivedi, *Microeconometrics: Methods and Applications*, Cambridge University Press, 2005.

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

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

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

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

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

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

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

Fetter, T. Robert, "Fracking, Toxics, and Disclosure," 2016. Unpublished manuscript.

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

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

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

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

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

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.

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

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

Krebs, Charles, *Ecological Methodology*, Menlo Park, California: Addison Wesley Longman, 2014.

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.

Olmstead, Sheila M. and Nathan Richardson, "Managing the Risks of Shale Gas Development Using Innovative Legal and Regulatory Approaches," RFF Discussion Paper 2014.

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

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

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

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

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

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

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