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

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

December 17, 2016

Preliminary and Incomplete
Do Not Cite or Circulate

Abstract

Firms can learn about new technologies from other adopters, though this social learning is not always exploited. 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: O3, L24, L51, L71, Q35, Q53, Q55, Q58

1 Introduction

Policies designed to disseminate information about pollution releases (or the potential for such releases) have become increasingly popular.[1] 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. 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.

One emerging technology in which information-based regulations have been recently used is hydraulic fracturing for shale gas and oil. This rapidly growing technology 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). Starting in 2010, policies requiring public disclosure of information about the chemicals used in the process of hydraulic fracturing have been instituted in eighteen states.

Information-based regulations, such as mandatory disclosure of hydraulic fracturing chemicals, have the potential to inspire reduced use of toxic chemicals or other reductions in potential threats to environmental quality or human health. Operators likely have better information regarding which chemicals are most effective for releasing trapped gas in a particular shale formation, meaning that direct regulations (i.e., command-and-control policies) may be inefficient. Given flexibility over specific actions, and the ability to make tradeoffs between the productivity impacts of altering specific inputs and the potential social sanction of continuing to use toxic chemicals, firms may respond (e.g. by using less toxic formulations, or by taking actions to prevent leaks and spills) to various forms of public pressure, including the threat of legal action.

At the same time, mandatory information disclosure may result in unintended

[1] See, for example, policies regulating electricity generation (Delmas et al., 2010; Kim and Lyon, 2011), and drinking water provision (Bennear and Olmstead, 2008). 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).

consequences. In particular, they may undermine the ability of operators to appropriate the results of their innovations. Firms in the oil and gas industry, as in many industries, rely on secrecy as a fundamental mechanism to capture value from their investments in innovation and maintain competitive advantage (Wang and Krupnick, 2013; Cohen et al., 2000). This may be one reason the industry has worked hard to oppose mandatory disclosure regulations, arguing for the protection of "trade secrets" with respect to the contents of fracturing fluids. In this paper, we hope to shed light on whether innovations in fracturing fluids that are the subject of these disclosure debates indeed contribute to competitive advantage. In particular, we compare detailed, well-level information on inputs and outputs to test whether operators took advantage of public disclosure (by their competitors) to copy the behaviors of industry leaders. We observe strong evidence that such copying did take place, and go on to test whether the act of copying appears to improve productivity for poorer-performing firms. We find that the answers to both questions are affirmative. This suggests policy-makers should weigh carefully the tradeoff between right-to-know disclosure rules and industry claims that secrecy is essential to preserving competitive advantage. A further question, which we leave to future work, is whether forced disclosure undermines 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, focusing on the first questions in this paper.

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 firms in previously drilled wells, after the disclosure rules have gone into place.

This paper makes several key contributions to literature on innovation, learning, and information disclosure. First, we expand the literature on the empirical effects of transparency or disclosure regulations in a new direction. A number of papers have documented the effects of disclosure laws on environmental, health, or other outcomes of interest for consumers and the general public, in both consumer-facing (Delmas et al., 2010; Bennear and Olmstead, 2008; Jin and Leslie, 2003; Fung et al., 2007) and non-consumer-facing (Fung et al., 2007; Fetter, 2016; Konar and Cohen, 1997) settings. In addition, some work has analyzed how investors respond to corporate environmental disclosures (Hamilton, 1995). Other authors, too, have explored avenues for social learning across firms or other actors (Conley and Udry, 2010; Covert, 2015). Our analysis is the first, to our knowledge, to bring together information on the effects of information disclosure regulations with the effect of social learning. We find strong evidence that information disclosure regulations do create a pathway for social learning in an emerging technology. This could also be interpreted as a form of technology diffusion facilitated by regulation.

Second, we contribute to literature on the relationship between innovation, secrecy, and disclosure. The economic returns to innovation depend upon the ability of firms to appropriate the rewards of the knowledge they create. While patents, in theory, solve the problem of imperfect appropriability, evidence suggests firms in many industries rely more often on secrecy, lead time, and investment in complementary capabilities to maintain the competitive advantage they gain by innovation (Cohen et al., 2000). At the same time, in some settings—such as when firms' production processes involve the use of toxic chemicals—there may be a compelling public interest in revealing specific process elements. In settings such as these, policy makers must weigh the social costs of requiring disclosure by private entities against the social costs of potentially reducing the ability of firms to preserve the secrecy of their innovations and thereby appropriate the rewards of their innovations. We provide some insight into to what extent this tradeoff is real or imagined.

Third, we contribute to a growing literature on how firms learn about emerging technologies in general, and hydraulic fracturing in particular. However, no prior authors have addressed the role of chemical additives in the process. We take the role of chemicals seriously.

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 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 The Sources of Innovation Behind the Rise of Shale Gas

The rise of shale gas to prominence in the US energy landscape has been well-documented.[2] Less is known, however, about the role played by public and quasi-public institutions in that rise. These institutions played a prominent role because innovations are notoriously difficult to monetize in the oil and gas industry, leading to a disincentive to undertake costly and risky research and development activities.[3] The shale industry has been, and continues to be, based on an evolving technology. Technologies developed in the 1970's and 1980's, ranging from changes in the major compounds comprising fracturing fluid to the direction in which wellbores were drilled, transformed the way shale is approached, with significant implications for the natural gas industry. 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 – along with hydraulic fracturing, the ability to drill horizontally through a shale formation made shale gas development economical. At the same time, "slick water" fracturing fluid replaced gels. Most recently, fracturing fluid has been refined in multiple dimensions for maximizing output and minimizing costs.

The early development of shale gas was largely based on advances in four key areas of technology: (1) horizontal drilling, (2) three-dimensional seismic imaging, (3) micro-seismic fracture mapping, and (4) massive hydraulic fracturing. Development of new technology in support of a nascent industry requires large, risky investments. Many of these developments were made by Mitchell Energy on the Barnett shale in north Texas (the birthplace of the modern shale gas industry, for all practical purposes). There were certain features of the Barnett shale and Mitchell Energy

[2]Shale gas grew from 5% of total US dry gas supply in 2004 to 56% in 2015 (See EIA Natural Gas Monthly data through December, STEO through May 2015 and Drilling Info; 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 to the availability of pre-disclosure data, Pennsylvania will be our area of study (see Section 3).

[3] "Because private companies have difficulty monetizing and capturing all the benefits of energy technology research, it is consistently the case that federal coordination and investment is required to drive high-level technological innovation in the energy sector" (Shellenberger et al. (2012); see also Mazzucato (2014)).

that made this possible – Mitchell had secured a favorable long-term contract price at which it sold its gas to wholesalers, placing it in a secure financial position that allowed for R&D activities. At the same time, declining output from old reserves meant that Mitchell was motivated to seek new sources in order to meet long-term contractual demands. The risk of drilling experimental wells was reduced because the Barnett shale contained many overlapping formations that allowed Mitchell to hedge by extending those wells, should they fail, into traditional formations. Finally, Devon Energy (the company that eventually purchased Mitchell for $3.5 billion in 2002) brought its own particular expertise in horizontal drilling, which when combined with Mitchell's advances in slickwater fracturing, made shale gas economically competitive.

While much of the history of the shale gas industry is tied up in the history of firms like Mitchell and Devon, the size of and risk associated with the R&D investments necessary to grow the industry mean that it would not have been successful without the help of a number of public and quasi-public institutions. 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 reduces 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.[4]

Much of the technological development that took place on the Barnett 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). Mitchell also received assistance from the Gas Research Institute (GRI) in developing micro-seismic fracture mapping around 2000.[5] However, not all of the innovation was supported by federal sources. Evidence suggests that firms on Barnett developed familiarity with 3D seismic imagery, which was crucial for successful horizontal drilling, by sharing information with Republic Energy, which used the technology to adjust lateral placement on its horizontal well segments (Wang and

[4]Dan Steward, former Vice-President at Mitchell Energy, was quoted as saying in regards to these programs: "The DOE started it, and other people took the ball and ran with it... You cannot diminish DOE's involvement" (Shellenberger et al., 2012).

[5]The Gas Research Institute was a nonprofit established by the gas industry in 1976 and partially funded by a surcharge on interstate natural gas transactions until 1998. Its activities, which were focused mainly on applications and technology transfer, were justified by arguing that the benefits of R&D expenditures in natural gas were mostly unappropriated by the firm making them, and were hence underprovided by the private market.

Krupnick, 2013).

2.2 The Role of Fracturing Fluid

A wide array of chemicals are used in hydraulic fracturing fluid to enhance the productivity of the primary inputs – water and sand. In particular, additives are used to minimize the growth of bacteria that might interfere with metal casing or cause other problems, and to lower viscosity in order to allow faster pumping and higher pressures. Doing so leads to better fractures. Other chemicals are then used to increase the delivery of proppant into those fractures. Fracturing fluid is, however, a complicated solution where an additive that improves performance in one dimension may reduce performance in another. The cost of the chemicals themselves is very small in comparison to the overall cost of the well stimulation operation and is not a major factor. However, the proper choice of chemicals may have dramatic effects on the overall cost and productivity of a well (Boling, 2013).

Designing fracturing fluid for optimal performance on any particular shale play has become a major focus for oil and gas engineers, and represents a key dimension on which shale technology continues to evolve today (Robart et al., 2013; Montgomery, 2013; Gulbis and Hodge, 2000). In many cases, this leads engineers to consider the use of highly toxic chemicals (Stringfellow et al., 2014). Indeed, many of the chemicals used in fracturing fluid are known to be toxic to human health, or may cause damage to the ecosystem. These risks have raised concerns among environmental groups (Elgin et al., 2012; Haas et al., 2012).

Early concerns about shale gas development were driven by the possibility that toxic fracturing fluid might migrate to or be accidentally released into ground water or surface water. The industry responded to these concerns by pointing to the small percentage of fracturing fluid that is actually comprised of toxic substances.[6] 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.[7] Media coverage of fracturing chemicals has highlighted the toxicity of some chemicals along with the industry's desire to keep certain chemicals secret (Elgin et al., 2012; Haas et al., 2012). The federal government has largely dismissed the risk posed to water supplies by shale gas development,[8] but that debate has recently re-surfaced over suggestions that those conclusions might be politically motivated.[9]

[6]Approximately 90% of modern-day fracturing fluid is water, 9.5% is sand, and the remainder is made up of a variety of chemicals, many of which are non-toxic and found in the average home (http://www.energyfromshale.org/hydraulic-fracturing/hydraulic-fracturing-fluid). However, if even a fraction of a percentage of fracturing fluid is toxic, this could represent tens of thousands of gallons of chemicals for a typical operation that uses five million gallons of fluid.

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

[8]https://www.epa.gov/hfstudy

[9]https://stateimpact.npr.org/pennsylvania/2016/11/30/marketplace-epas-last-minute-changes-

2.3 Information-Based Regulation

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 Olmstead, 2008). One 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 risky activities off into operations with little exposed assets.[10] 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 primarily occurs deep underground, it may be hard for regulators and the general public to monitor operator activities. Providing potential victims with information reduces this cost.

Bennear and Olmstead (2008) identify an additional *political mechanism*. In particular, information may increase the ability of a concerned public to lobby for stronger regulation. Finally, 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.

Konschnik (2014) summarizes other reasons why disclosure is valuable, many of which are more direct. These are:

to-fracking-report-downplayed-risks/

[10]Insurance and bonding requirements in Pennsylvania do not do much to help. Bonding requirements are very small (either $2,500 per well or $25,000 for all wells in the state).

- In the event of an accident, disclosure to emergency medical personnel and medical staff may improve treatment and protect the staff members themselves.
- 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 help establish liability for contamination, helping victims sue, making liability a more effective regulatory tool.
- Disclosure allows regulators to enforce minimum standards.
- 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 (from which the shale gas industry is currently exempt).
- Disclosure facilitates the monitoring of environmental health impacts over time.

2.4 Disclosure Rules

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

There is less uniformity in terms of the location 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, a web-based database created by a multistate 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).[11]

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

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

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

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 data and 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 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 vertical as well as 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.[13] 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.[14] FracFocus permitted operators to upload data in

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

[14] 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

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.[15] 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 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.[16] 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 Integrated Records and Information System (IRIS), in which some chemical disclosure forms were available as scanned PDF documents.[17] 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.[18] 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

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.

[15]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.

[16]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).

[17]The IRIS system has recently been revised and rebranded as EDWIN, the Exploration and Development Well Information Network.

[18]One of the authors accessed IRIS 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.

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 operators failed to provide the length of the perforated interval for 2,076 wells, so we cannot use these in the analysis.[19] We drop an additional 97 wells because of other concerns about data quality, primarily nonsensical completion dates (the recorded completion date is after the date of initial production). Finally, we drop 20 wells with zero recorded gas production and 8 wells with exceptionally high values of gas per foot so as to ensure the analysis is not driven by outliers. This leaves us with a sample of 4,827 wells.

For each of these wells we observe identifying information (operator, location, and completion date) and output (per 6 months). 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 inaccessible to scraping, due 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 (including quality assurance by systematic error checking, spot checks, and comparison of duplicate entry 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 3,676 wells; proppant volume available for 2,548 wells; and both fluid and proppant quantity are available for just 1,478 wells. We have detailed chemical information for 3,334 wells, including 3,077 wells with chemical information disclosed to FracFocus and 257 wells (from the period February 2011 - April 2012) with chemical information disclosed only to PADEP. Of the 1,493 wells with no information on chemicals, 1,292 were not subject to disclosure; 201 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.[20] We were able to obtain other information on well parameters for a larger set of wells, including the number of fracturing stages for each well; this information is available for 4,611 wells.

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, note that chemicals data are available only for wells starting in 2010.[21] The

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

[20] Another possibility is that operators provided chemical information for these wells but the reports had not been digitized and uploaded to IRIS by December 2015 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 years from when DEP received the reports to entry of the report into IRIS.

[21] Mandatory disclosure started in Pennsylvania in 2011, but prior to this date some operators voluntarily released information for some of their wells.

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 initial 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	3.9	4.2	14.5	13.5	162	141.2
2011	4.5	4.6	11	11.2	151.6	124.2
2012	4.2	4.4	12	13.9	187.9	132.9
2013	5.7	5.9	15	17.1	207	163.2
2014	7.3	7.8	17	17.3	222	188
2015	9	8.9	15	15.6	171.4	166.4

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 what we observe in our sample).

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

			Chemicals		
Year	# Wells	# Operators	# New	# Retiring	# Total
2010	36	8	58	4	58
2011	816	23	87	19	134
2012	969	27	91	22	200
2013	820	30	55	77	239
2014	600	20	34	124	198
2015	67	11	1	–	80

4 Analysis

In this section, we perform a series of empirical analyses designed to test whether mandatory disclosure of chemicals, even with provisions allowing for some chemicals to be declared proprietary, led to social learning. In future work, we plan to analyze the extent to which this may have reduced operators' incentives to innovate.

Focusing on the first of these two questions, if disclosure has promoted social learning by competitors, we would expect to see convergence in productivity across operators, with less able firms using the information revealed by the disclosure laws to catch-up with more successful firms. Of course, there could be other reasons for this convergence. We therefore focus on the specific mechanism of interest, looking at the role of similarity in the design of fracturing fluid. Validating that mechanism, we find that convergence is indeed explained specifically by weaker firms adopting the chemical mixtures used by more successful firms in previously drilled wells, after the disclosure rules have gone into place.

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 marks 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.

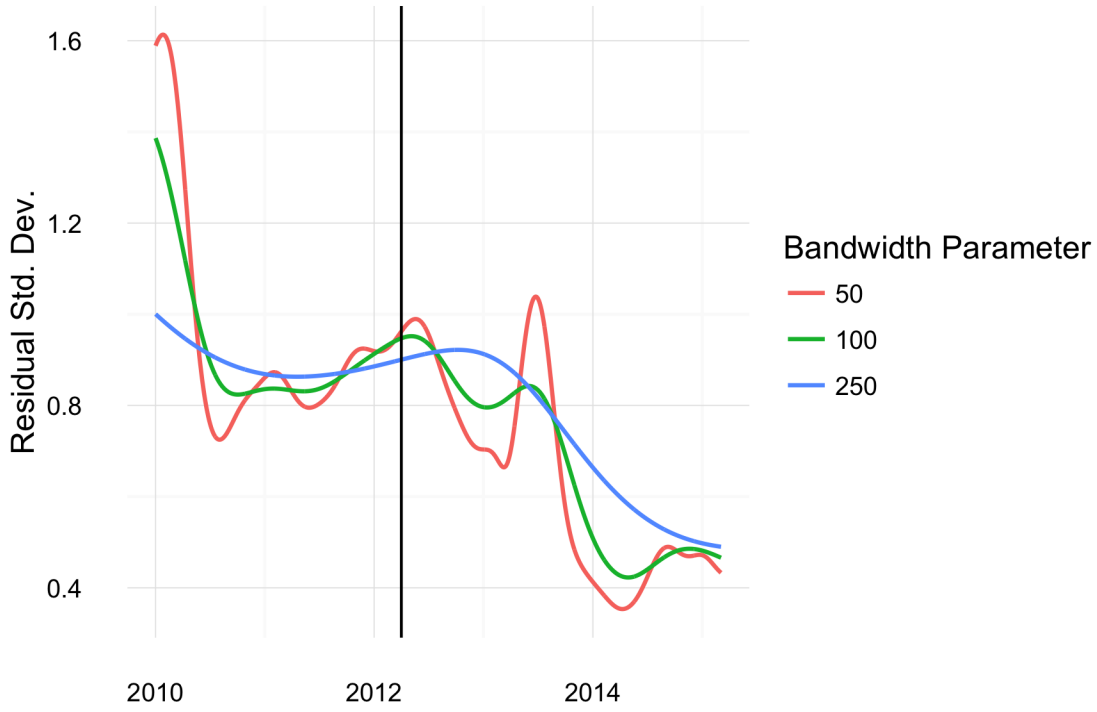
4.2 Well-to-Well Similarity Measures

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 convergence 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, define our similarity metric of choice, and provide a summary of the metric in our data.

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.

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 can analogously define distances $d_{ij} \equiv 1 - s_{ij}$.

We now introduce some notation to assist with the discussion of 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 define the following quantities for a given (i,j) pair:

- $a(i,j) = \sum_{\mathcal{C}} y_{ic}$ is the number of chemicals found in well i
- $b(i,j) = \sum_{\mathcal{C}} y_{jc}$ is the number of chemicals found in well j
- $c(i,j) = \sum_{\mathcal{C}} y_{ic} y_{jc}$ is the number of chemicals found in both wells i and j

In principal we can calculate "abundance" or "correlation" similarity metrics that take into account the more detailed information in x_{ic}. Two features of our setting prevent this 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:

$$s_{ij} \equiv \frac{2c(i,j)}{a(i,j) + b(i,j)}. \tag{2}$$

4.3 Pre- vs. Post-Similarity Measures

We calculate the similarity metric for each i,j well-pair in our data. Our chemical data on 3,334 wells gives us 5.6m such pairs. Figure 2 shows the distribution of these measures for two sub-samples of pairs: those where both wells are in the pre-disclosure period and those where both wells are in the post-disclosure period. The post-period distribution is noticeably shifted to the right of the pre-period distribution, indicating that well-pairs in the post-period tend to use more similar chemical mixes than well-pairs in the pre-period; however it gives no indication as to whether the mixes are better or worse. In the next sections we explore this fact in more detail, and provide

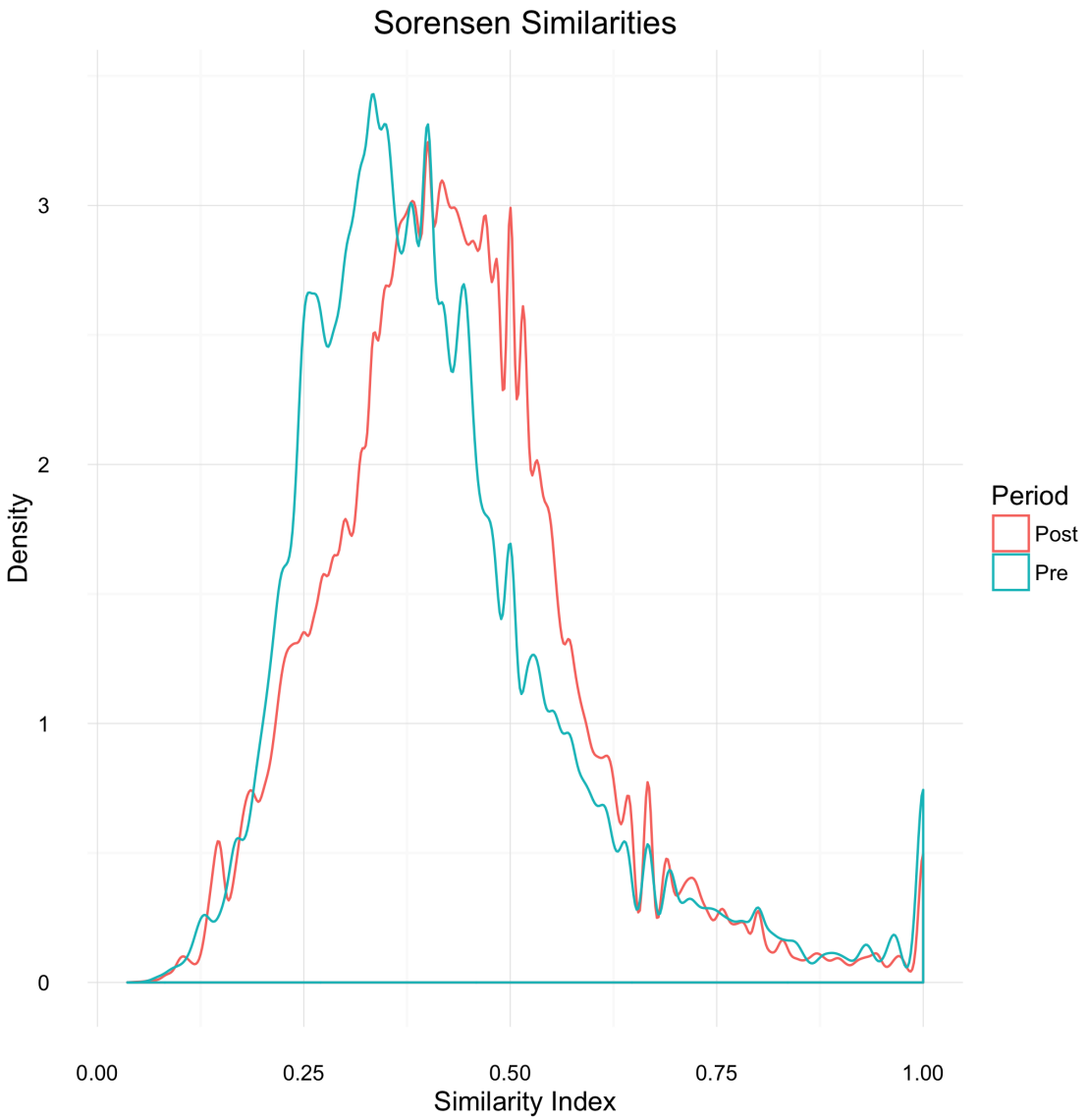
[22] Most metrics, including those we use, feature the property $s_{ij} \in [0,1]$, although other measures like correlation, with support $s_{ij} \in [-1,1]$, are possible.

[23] See Krebs (2014).

[24] The Jaccard binary metric is defined as: $s_{ij} \equiv c(i,j)/(a(i,j) + b(i,j) - c(i,j))$.

evidence that disclosure has led to learning from *better* chemical mixtures. Specifically, the next two subsections consider the evolution of similarities based on whether well-pairs are located in the same-county or belong to a big vs. small operator; both restrict attention to inter-operator well-pairs.

Figure 2: Sorensen Similarity Measures

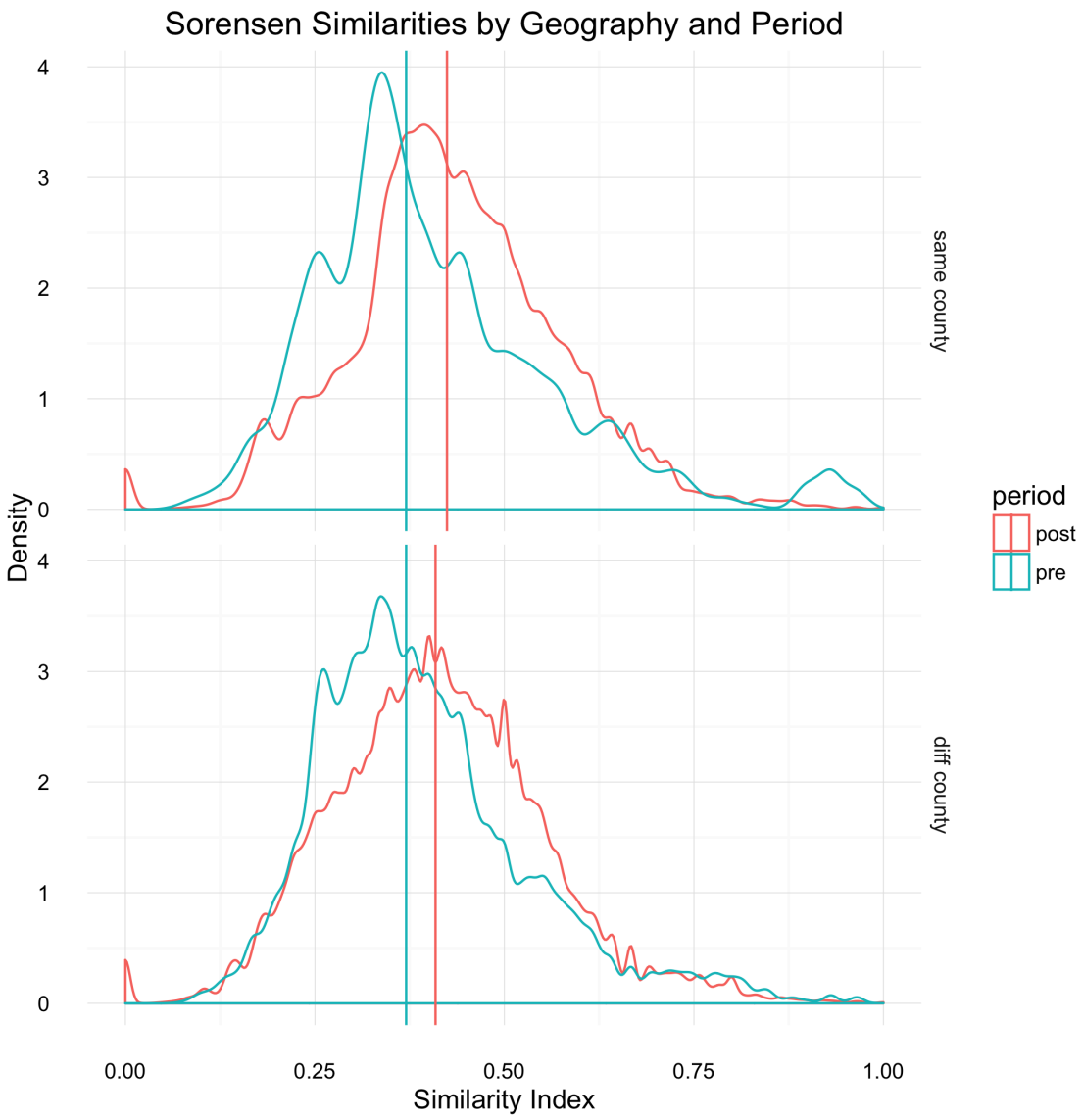


4.3.1 By County

We first investigate the development of similarities between same-county and different-county well-pairs. Figure 3 shows four distributions on two plots: the top plot has the

pre- and post-disclosure distributions of same-county well-pairs; the bottom plot has the pre- and post-disclosure distributions of different-county well-pairs. Both plots also feature distribution medians as vertical lines. It can be seen from Figure 3 that both the same-county and different-county well-pairs become more similar on average in the post-disclosure period. The medians further suggest that the same-county well-pairs similarity distribution is shifted further to the right in the post period - this is consistent with operators using disclosure to imitate chemicals from those wells on the same part of the formation.

Figure 3: Sorensen Similarity Measures - by County



This idea is further tested in Table 3, where we adopt a difference-in-differences

regression approach. We test a shift in the mean with OLS, and at three different quantiles with quantile regression. In each case, the regression equation is:

$$s_{ij} = \beta_0 + \beta_1 same_county_{ij} + \beta_2 post_{ij} + \beta_3 same_county_{ij} * post_{ij} + \epsilon_{ij} \quad (3)$$

The parameter of interest is "same county x post", which if positive would indicate that inputs for same-county well-pairs become more similar than inputs of different-county well-pairs moving from the pre- to the post-period. As can be seen in the seventh and eighth rows of Table 3, this is the case at the mean, 25th percentile, and median, where the treatment effect associated with being in the same county is significant and positive, except for the 75th percentile.

Table 3: Sorensen Similarity Regressions - by County

	OLS	25th Quantile	Median	75th Quantile
Intercept	0.392^{***}	0.296^{***}	0.370^{***}	0.462^{***}
	0.0003	0.0003	0.0003	0.0005
Same County	0.014^{***}	0.011^{***}	0.000	0.021^{***}
	0.001	0.001	0.001	0.002
Post	0.020^{***}	0.024^{***}	0.039^{***}	0.038^{***}
	0.0003	0.0003	0.0003	0.0005
Same County x Post	0.007^{***}	0.019^{***}	0.015^{***}	-0.005^{***}
	0.001	0.001	0.001	0.002
N	3,196,755	3,196,755	3,196,755	3,196,755
(Pseudo) R2	0.004	0.004	0.005	0.004
Notes:	***Significant at the 1 percent level.			
	**Significant at the 5 percent level.			
	*Significant at the 10 percent level.			

4.3.2 By Operator Size

We next perform a similar analysis, but split the well-pairs by operator size rather than whether or not the wells are in the same county. We define a variable "big operator" that takes the value 1 if the operator drills more than 100 wells in our data, and zero otherwise.[25] Figure 4 shows the evolution of similarities between those well-pairs where both operators are "big" and those well-pairs where both operators are "small". It is evident from the figure that "big" well-pairs exhibit much larger increases in similarity in the post period. This is consistent with the expected impact of disclosure laws, as larger firms are more likely to have the resources to analyze disclosed chemical mixes and respond accordingly.

[25] 100 wells is just above the median of 93 in our data.

Figure 4: Sorensen Similarity Measures - by Operator Size

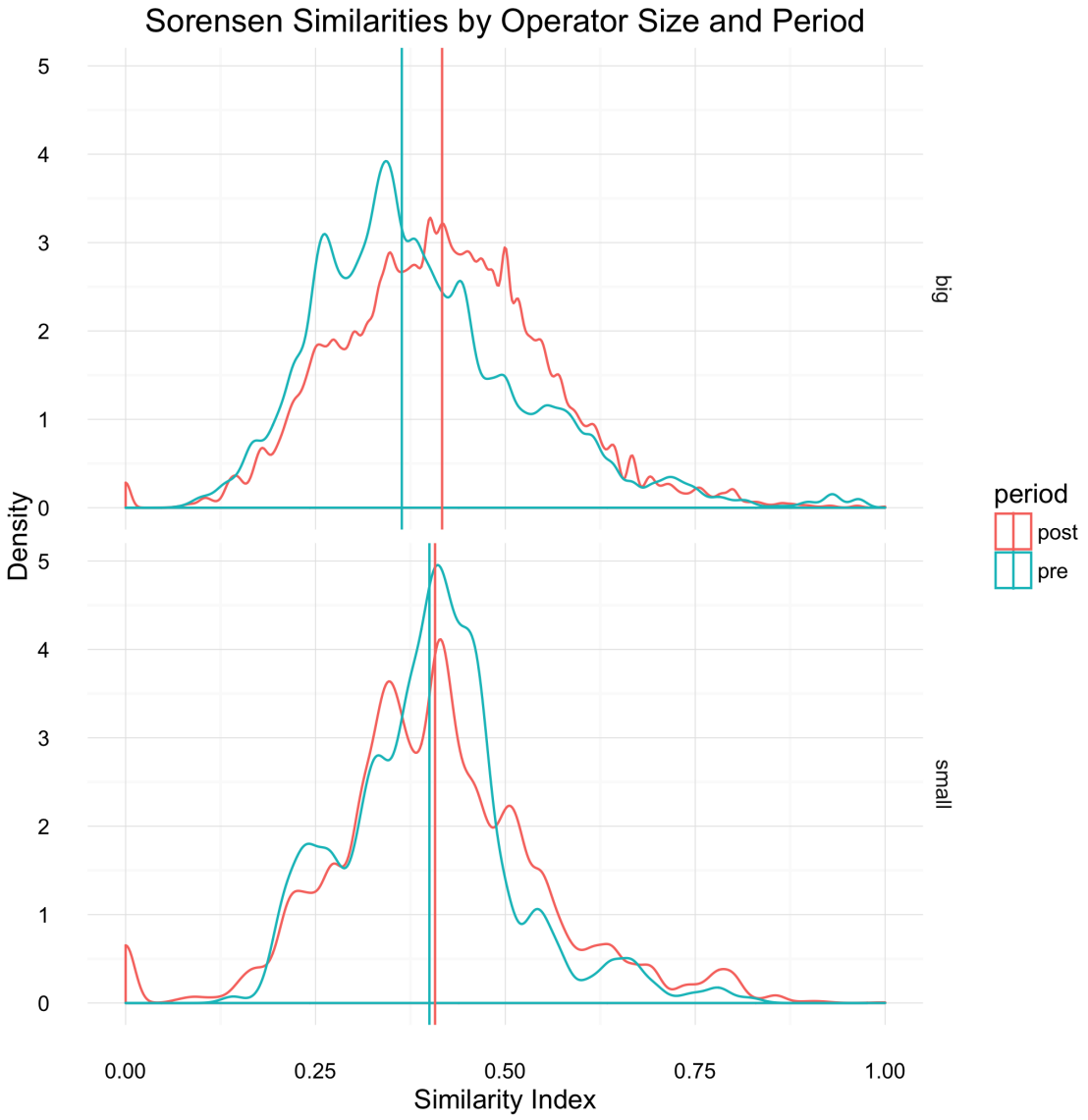


Table 4 tests the changes in the pre- and post-distributions in a difference-in-differences framework, with the regression:

$$s_{ij} = \beta_0 + \beta_1 big_{ij} + \beta_2 post_{ij} + \beta_3 big_{ij} * post_{ij} + \epsilon_{ij} \quad (4)$$

The "big operator x post" coefficient is estimated to be positive and statistically significant at the mean, as well as the 25th, 50th, and 75th percentiles. This is in contrast to the "post" estimate which is significantly negative at the lower percentile and median, indicating that small operator well-pairs became less similar at those parts of the distribution. Again, this suggests that the chemical mixes of big operators respond to disclosure differently from those of small operators, and is consistent with the use of disclosed chemicals as a source of information.

Table 4: Sorensen Similarity Regressions - by Operator Size

	OLS	25th Quantile	Median	75th Quantile
Intercept	0.402***	0.333***	0.400***	0.462***
	0.001	0.001	0.001	0.001
Big Operator	−0.013***	−0.048***	−0.036***	0.000
	0.002	0.001	0.001	0.001
Post	0.007***	−0.000	0.007***	0.027***
	0.002	0.001	0.001	0.002
Big Operator x Post	0.022***	0.041***	0.046***	0.011***
	0.002	0.001	0.001	0.002
N	2,235,523	2,235,523	2,235,523	2,235,523
(Pseudo) R2	0.004	0.014	0.006	0.004

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

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 initial (first-6-months) gas production per foot on a set of observables and fixed effects, including operator fixed effects that differ in each of two periods (before and after the mandatory disclosure regulation). These operator fixed effects 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 improvement (or reduction) 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 operators.

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

$$y = F \times g(X) \tag{5}$$

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

$$lny = lnF + X'\beta \tag{6}$$

In our preferred specification, the function $g(X)$ includes a county fixed effect, a year fixed effect, the density of unconventional wells previously completed in the county (as a quadratic), and the current output price.[26] The year fixed effect should absorb secular technological change, and the county 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 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 six 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

[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.

contribution of the firm – a scale parameter, or multiplicative effect – that is positive by construction.

Table 5 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). The analysis also shows productivity increasing with prior well density, up to a point (about 0.13 wells per square km, which is approximately the mean of the distribution). This is consistent with the notion that as operators gain familiarity with an area they develop relatively more promising sites, but eventually only less promising sites are available.

Table 5: Results of first stage regression

Variable	**Coefficient** (Std. Err.)
Gas Price	0.069**
	(0.021)
Prior Well Density	1.548*
	(0.663)
Prior Well Density (squared)	-5.842**
	(1.513)
Constant	4.810**
	(0.286)
County FE	✓
Year FE	✓
Operator FE	✓
N	4827
R^2	0.134
$F_{(143,4683)}$	16.849

4.4.2 QS Metric

To investigate the relationship between changes in productivity and the convergence in chemical mixes shown in Figure 2, 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:

$$QS_f \equiv \frac{1}{|M_f|} \sum_{m \in M_f} s_m \theta_m^{PRE} \tag{7}$$

$$M_f \equiv \{i, j \mid i \text{ is } f \text{ POST well, } j \text{ is non-}f \text{ POST well, } j \text{ fractured before } i\} \tag{8}$$

$$s_m \equiv \text{similarity index for well-pair } m \tag{9}$$

$$\theta_m^{PRE} = \text{estimated PRE period fixed effect for well } j\text{'s firm.} \tag{10}$$

QS_f is thus a quality-weighted measure of the 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 QS DD Test

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

$$\theta_f^{POST} - \theta_f^{PRE} = \beta_0 + \beta_1 \theta_f^{PRE} + \beta_2 QS_f + \beta_3 QS_f \theta_f^{PRE} + \epsilon_f. \tag{11}$$

The results can be seen in Table 6. 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 11, with the addition of QS_f and the interaction term. With these additions, the point estimate of β_1 has flipped to be positive and is no longer statistically significant: if the point estimate is taken seriously on its own however, it suggests divergence in θ_f over time. The point estimate of β_2 is also positive and insignificant, suggesting that a higher QS_f is associated with higher growth in θ_f. Finally, the estimate of 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.

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. 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?

In an attempt to address this concern, we make several alternative assumptions about how operators would have performed in the post period had they continued to operate in the state. We make four alternative assumptions about firm performance:

Table 6: Second Stage Results

	(1)	(2)
θ^{PRE}	-0.638***	3.262*
	(0.18)	(1.62)
QS_f		3.328
		(4.92)
$\theta^{PRE}QS_f$		-8.392**
		(3.53)
Constant	0.649**	-1.022
	(0.26)	(2.25)
R^2	0.291	0.573
N	31	27

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

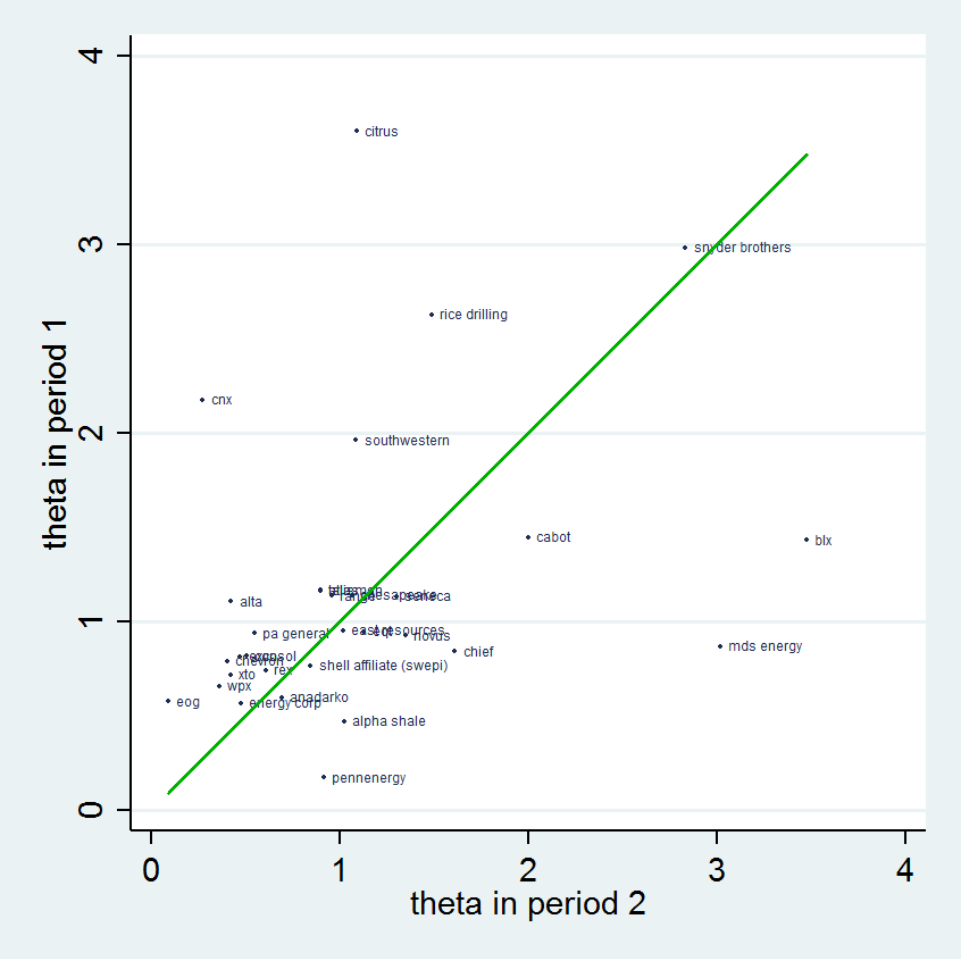
2. Assume exiting firms would have performed at the median of the firms that did continue to operate

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 the top half in the pre period would have performed at the mean of the distribution in the post period

Note that in the context of testing for simple convergence (without considering the role of quality-similarity), an assumption that all exiting firms would have performed at the top of the distribution in the post period does not provide for a conservative test of the convergence hypothesis. Such an assumption would imply 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 also be seen in Figure 5, which shows the relative magnitudes of the first-period and second-period operator fixed effects. 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.

Figure 5: Operator Fixed Effects in Period 1 and Period 2



We must also assign values for QS_f for the firms that exited. In this case, we make two 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.

This range of assumptions also provides for a conservative test of the hypothesis that convergence is achieved primarily through ex-ante low-performing firms using input vectors that are more similar to those of ex-ante higher-performing firms. 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. Table 7 provides a summary of results evaluating convergence (with and without inclusion of the QS_f variable) under each combination of alternative assumptions. Note that the table contains the results of 16 regressions; each column and set of rows (between the horizontal lines) represents a distinct regression.

As can be seen from the table, the results of this analysis are consistent with the results of the main analysis, supporting the finding that convergence is driven primarily through the interaction of the QS_f term and the θ^{PRE} term.

4.5 Placebo Tests

In addition to our attempt to combat potential sample selection in the prior section, we recreate this analysis with three 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. The second alternative goes further, and *only* includes those well-pairs where firm f's well was stimulated first. Finally, we create a version that includes a geographic limitation in order to rule out spatial spillovers.

In the first two alternatives, the estimated coefficients are not significant which supports the hypothesis that disclosure aided learning. The alternative with a geographic restriction has very similar results to the original, suggesting that the convergence is not due to spatial proximity.

Table 7: 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}	-0.66^{***}	1.3	2.4^{*}	3.4^{***}
		(0.1)	(1.2)	(1.3)	(1.2)
	QS_f		2.2	1.9	2.2
			(3.7)	(4.2)	(3.4)
	$QS_f \times \theta^{PRE}$		-4.5^{*}	-6.8^{**}	-8.8^{***}
			(2.7)	(3.0)	(2.5)
Median	θ^{PRE}	-0.7^{***}	0.6	1.8	3.7^{***}
		(0.1)	(1.0)	(1.2)	(1.1)
	QS_f		-1.4	-0.0	5.1
			(3.2)	(3.7)	(3.3)
	$QS_f \times \theta^{PRE}$		-3.0	-5.6^{**}	-9.7^{***}
			(2.4)	(2.6)	(2.4)
Bottom Or Top	θ^{PRE}	-0.1	5.1	4.3	2.8
		(0.6)	(6.2)	(8.0)	(8.6)
	QS_f		-11.2	0.4	21.6
			(19.3)	(25.0)	(24.8)
	$QS_f \times \theta^{PRE}$		-11.7	-9.8	-6.1
			(14.1)	(17.9)	(18.3)
Bottom Or Mean	θ^{PRE}	0.4^{***}	3.2^{***}	3.7^{***}	2.6^{*}
		(0.1)	(1.1)	(1.3)	(1.4)
	QS_f		6.2^{*}	5.0	0.4
			(3.6)	(4.2)	(4.0)
	$QS_f \times \theta^{PRE}$		-8.3^{***}	-9.3^{***}	-6.8^{**}
			(2.6)	(3.0)	(2.9)
N		57	57	57	57
N (exit)		26	26	26	26

*** p$<$ 0.01, ** p$<$ 0.05, * p$<$ 0.10.
All QS is standard Sorensen. All regressions include a constant term. Growth rate is $\theta^{POST} - \theta^{PRE}$

4.5.1 "Time-fluid QS"

For the first alternative, we define a new quality-similarity index QS_f^{a1}:

$$QS_f^{a1} \equiv \frac{1}{|M_f^{a1}|} \sum_{m \in M_f^{a1}} s_m \theta_m^{PRE} \tag{12}$$

$$M_f^{a1} \equiv \{i, j \mid i \text{ is } f \text{ POST well, } j \text{ is non-}f \text{ POST well}\} \tag{13}$$

$$s_m \equiv \text{similarity index for well-pair } m \tag{14}$$

$$\theta_m^{PRE} = \text{estimated PRE period fixed effect for well } j\text{'s firm.} \tag{15}$$

Note that the only change in this definition is the set of well-pairs used, M_f^{a1}. 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 the first alternative, we do away with this limitation and construct QS_f^{a1} 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^{a1} will result in less statistical significance of the estimates of β_2 and β_3.

The results of the new second stage regression are shown in column 3 of Table 8. Unlike the original QS_f results (shown in column 2), none of the estimated coefficients are statistically significant. Furthermore, the signs have all flipped and are no longer consistent with convergence being due to the quality-weighted similarity measured by QS_f^{a1}.

4.5.2 "Time-Inverse QS"

The second alternative goes further, and defines the new quality-similarity index QS_f^{a2}:

$$QS_f^{a2} \equiv \frac{1}{|M_f^{a2}|} \sum_{m \in M_f^{a2}} s_m \theta_m^{PRE} \tag{16}$$

$$M_f^{a2} \equiv \{i, j \mid i \text{ is } f \text{ POST well, } j \text{ is non-}f \text{ POST well, } j \text{ fractured } \textit{after } i\} \tag{17}$$

$$s_m \equiv \text{similarity index for well-pair } m \tag{18}$$

$$\theta_m^{PRE} = \text{estimated PRE period fixed effect for well } j\text{'s firm.} \tag{19}$$

This definition of M_f^{a2} now only includes wells for which f could not have seen chemical disclosures. As in the prior case, statistical insignificance will support the hypothesis that chemical learning is occurring through disclosures.

The second stages results using QS_f^{a2} are shown in column 4 of Table 8. The signs of the point estimates remain the same as in the original version, but all shrink toward zero and none are statistically significant.

Table 8: Second Stage Results with Alternative QS Definitions

	(1)	(2)	(3)	(4)	(5)
θ^{PRE}	-0.638***	3.262*	-1.232	0.821	3.247*
	(0.18)	(1.62)	(1.26)	(2.00)	(1.68)
$\theta^{PRE}QS_f$		-8.392**			
		(3.53)			
QS_f		3.328			
		(4.92)			
$\theta^{PRE}QS_f^{a1}$			1.210		
			(2.85)		
QS_f^{a1}			-6.593		
			(4.86)		
$\theta^{PRE}QS_f^{a2}$				-3.087	
				(4.28)	
QS_f^{a2}				-1.234	
				(5.45)	
$\theta^{PRE}QS_f^{a3}$					-8.447**
					(3.69)
QS_f^{a3}					3.209
					(4.98)
Constant	0.649**	-1.022	3.707	1.223	-0.945
	(0.26)	(2.25)	(2.24)	(2.59)	(2.27)
R^2	0.291	0.573	0.375	0.395	0.577
N	31	27	27	27	27

4.5.3 Distance Buffer

The final 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 quality-similarity index QS_f^{a3}:

$$QS_f^{a3} \equiv \frac{1}{|M_f^{a3}|} \sum_{m \in M_f^{a3}} s_m \theta_m^{PRE} \tag{20}$$

$$M_f^{a3} \equiv \{i, j \mid i \text{ is } f \text{ POST well, } j \text{ is non-}f \text{ POST well, } i \text{ and } j \text{ at least 25km apart}\} \tag{21}$$

$$s_m \equiv \text{similarity index for well-pair } m \tag{22}$$

$$\theta_m^{PRE} = \text{estimated PRE period fixed effect for well } j\text{'s firm.} \tag{23}$$

In this alternative, M_f^{a3} is defined to include only more distant well-pairs. The second stage results are in the final column of Table 8. 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^{a3} and QS_f argues against the learning being spatial, and thereby for learning from disclosures.

4.6 Mechanism

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 is 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 one of the wells is in the post-disclosure period.

$$s_{ij} = \beta_0 + \beta_1 contractor_{ij} + \beta_2 post_{ij} + \beta_3 contractor_{ij} * post_{ij} + \epsilon_{ij} \tag{24}$$

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 9.

When it is the only non-constant regressor, the dummy for sharing a contractor is associated with an increase in the similarity index of 0.217. In the second column, the dummy for post is included: it leaves the contractor coefficient unchanged, and post enters negatively and significantly, but is quite small in magnitude. [28] 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 period.

Table 9: Determinants of Well-pair Similarities

	Sorensen Similarity Index		
Same Contractor	0.217***	0.217***	0.332***
	(0.000)	(0.000)	(0.002)
Post		−0.025***	−0.006***
		(0.001)	(0.001)
Contractor x Post			−0.118***
			(0.002)
Observations	3,499,335	3,499,335	3,499,335
R^2	0.155	0.156	0.157

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

5 Discussions and Conclusions

[28] (discuss how this relates to earlier findings)

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.

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.

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.

Konar, Shameek and Mark A. Cohen, "Information As Regulation: The Effect of Community Right to Know Laws on Toxic Emissions," *Journal of Environmental Economics and Management*, jan 1997, *32* (1), 109–124.

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.

Mazzucato, Mariana, *The entrepreneurial state: debunking public vs. private sector myths*, New York: Anthem Press, 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.

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.