Information Disclosure and Location Choices: A Study of The Marcellus Shale*

Wenshu Guo
Lingnan Univeristy

Abstract

Limited information diffusion is often a source of investment inefficiency in many emerging industries. This is particularly relevant in the context of unconventional resource extraction, where the distribution of deposits is highly uncertain, and exploration involves substantial fixed costs. This research examines the shale gas industry in Pennsylvania and explores a regulatory intervention that mandates firms to disclose their production information. Using a novel dataset on lease-taking activities, the study finds that shale gas operators are more likely to invest in locations with higher productivity following the disclosure intervention, potentially leading to more efficient resource allocation. The results underscore the signaling effect of production information on shale gas reserve quality and the social learning effect that ensues after the information's disclosure. The findings suggest that in industries where the quality of investment is uncertain, facilitating information sharing may be an effective way to enhance overall investment efficiency.

Key Words: *Information Disclosure, Uncertainty, Learning, Efficiency, Shale Gas*
JEL Code: *D2, L5, L71, Q3*

*I would like to thank John Asker, Martin Hackmann, Hugo Hopenhayn, Larry Qiu, Junji Xiao, and Daniel Xu for their advice and suggestions. I would also like to thank seminar participants at UCLA, Jinan University, SUFE, Lingnan University, NTU, and APIOC 2023 for their comments. Contact information: wenshuguo@ln.edu.hk

1 Introduction

Uncertainty regarding investment quality is often a defining characteristic of many emerging industries heavily reliant on research and exploration. In environments where information diffusion is limited, investment efficiency can be compromised due to insufficient knowledge or experience. This is particularly true in unconventional resource extraction, which involves significant fixed costs for exploration and production. This study examines the shale gas industry in Pennsylvania and investigates the impact of a regulatory intervention mandating firms to disclose their production information. By analyzing a novel dataset of lease-taking activities, this paper finds that shale gas operators are more inclined to invest in areas with higher productivity after the disclosure rule is enacted, leading to potentially more efficient resource allocation.

Vigorous shale gas production has profoundly changed the US energy landscape, yet the industry's development history is relatively short. For instance, the Marcellus Shale, the largest shale gas field in the US, was not commercially exploited until 2004. Considerable uncertainty surrounded its output potential, and estimates of its recoverable natural gas reserves were highly volatile. In 2002, the United States Geological Survey (USGS) reported an estimate of 1.9 trillion cubic feet of technically recoverable natural gas in the Marcellus Shale[1]. This figure increased dramatically to 410 trillion cubic feet in the 2011 Annual Energy Outlook published by the United States Energy Information Administration (EIA)[2]. However, just one year later, the EIA's 2012 Annual Energy Outlook revised the estimate downward by more than half to 141 trillion cubic feet[3]. With such significant uncertainty, shale gas operators faced challenges in making informed location decisions without incurring substantial exploration costs.

Pennsylvania is the epicenter of Marcellus Shale's natural gas production. In late 2010 and early 2011, the Pennsylvania Department of Environmental Protection (PADEP), which regulates shale gas production in the state, initiated a broad information transparency campaign. This campaign included a series of mandatory information disclosure requirements designed to promote more transparent regulation and facilitate public monitoring. One requirement, in particular, caught the attention of industry stakeholders: the obligation to accurately report production quantities. The first set of production data was made public in November 2010 and has since been updated periodically. Because well productivity closely correlates with the quality of shale gas reserves, the disclosure likely provided operators with a valuable source of information to refine their assessments of shale gas deposit distribution. This paper capitalizes on this regulatory shift to explore whether information sharing fosters social learning among firms and leads to shifts in their patterns of location choice.

This paper employs a dual approach, beginning with a theoretical examination followed by

[1] See the National Assessment of Oil and Gas Fact Sheet, published by the USGS in 2002, accessible at https://pubs.usgs.gov/fs/fs-009-03/FS-009-03-508.pdf

[2] See the Annual Energy Outlook 2011, published by EIA, accessible at http://www.pseudology.org/people/..%5C/gazprom/EnergyOutlook2011.pdf

[3] See the Annual Energy Outlook 2012, published by EIA, accessible at https://www.eia.gov/outlooks/aeo/pdf/0383(2012).pdf

empirical analysis. The initial step involves a simple theoretical model that serves two main objectives. Firstly, it provides a theoretical perspective on how the disclosure intervention promotes learning and enhances investment efficiency. Secondly, it underpins the subsequent empirical analysis, providing a foundation for the modeling approach. The theoretical framework takes its cues from the unique information structure within the industry: operators have the capacity to observe each other's past leasing decisions, thereby gaining indirect insights into their private information. However, it is only with the mandated disclosure of production information that they attain a comprehensive understanding of well productivity across the industry.

The second step involves a reduced-form analysis of the regulatory change's impact on firms' location choices, utilizing a unique dataset of leasing activities in Pennsylvania spanning the period from 2007 to 2014. One noteworthy feature of the data is its inclusion of all leases acquired by firms before and after the disclosure, enabling a direct comparison of location choice patterns between the two stages. The analysis comprises two main components.

The first part focuses on examining the correlation between royalty rates and productivity across different locations[4]. The study aims to identify and visualize the influence of the disclosure on firms' location choices. The estimation results indicate that the correlation between royalty rates and productivity increases by 1.54 times following the implementation of the disclosure policy. This suggests that higher productivity is associated with significantly higher lease costs when production information becomes publicly available. The findings provide compelling evidence that firms reallocate their resources toward more productive locations in response to the disclosure, leading to increased demand and subsequently higher lease costs.

The second part of the analysis focuses on directly modeling firms' leasing decisions and aims to quantify the extent to which these decisions are influenced by the disclosed production information. Building upon the theoretical model and information structure, the analysis employs an autoregressive framework to capture firms' private information based on their past leasing decisions. The estimation process yields several key findings. Firstly, a firm's current leasing decision exhibits a strong correlation with its own previous leasing decision in the same location. This robust correlation suggests that the firm's past leasing choices contain valuable private information and serve as reliable indicators for its future leasing activity. Secondly, the firm's current leasing decision also displays a significant correlation with the past leasing decisions of its competitors. This finding suggests that firms either share private knowledge or learn from their competitors' location choices to inform their own leasing decisions. Lastly, even after accounting for the private information of the firm itself and its competitors, the productivity of a location continues to exert a substantial positive impact on the firm's leasing decision following the disclosure. These results highlight the influential signaling effect of production information. Following the disclosure, social learning becomes more effective, leading to a more efficient allocation of resources.

This research is primarily related to three strands of literature. The first strand pertains to the

[4]Royalty rates represent the fraction of future gas sales revenue that the lease grantee must pay to the mineral rights owner and serve as a measure of the variable portion of a two-part tariff cost. The fixed part of the cost is seldom observed.

literature on quality disclosure. Early theoretical papers have extensively discussed the incentives and consequences associated with both voluntary and mandatory disclosure (Akerlof, 1970; Grossman and Hart, 1980; Grossman, 1981; Milgrom, 1981; Jovanovic, 1982; Matthews and Postlewaite, 1985; Shavell, 1994; Fishman and Hagerty, 2003; Board, 2009). There is a substantial body of empirical research that investigates quality disclosure in various industries (Ippolito and Mathios, 1990; Marshall et al., 2000; Jin and Leslie, 2003; Dranove, Kessler, et al., 2003; Mathios, 2000; Jin, 2005; Jin and Sorensen, 2006; Dafny and Dranove, 2008; Dranove and Sfekas, 2008; Hastings and Weinstein, 2008). Dranove and Jin (2010) provide a comprehensive review of the theoretical and empirical developments in the literature on quality disclosure. Recent empirical studies relevant to this research include Luco's (2019) examination of price disclosure in the retail gasoline market and Guo, Sriram, and Manchanda's (2021) study on the impact of transparency laws on industry payments to physicians.

The second strand of related literature focuses on learning and information externality, particularly within the oil and gas exploration industry. Several studies examining offshore oil and gas extraction have delved into the role of information in auctions of offshore leases and the exploration of wildcat tracts (Hendricks, Porter, and Boudreau, 1987; Hendricks and Porter, 1988; Hendricks and Kovenock, 1989; Hendricks, Porter, and Wilson, 1994; Porter, 1995; Hendricks and Porter, 1996; Hendricks, Pinkse, and Porter, 2003). These studies provide comprehensive investigations into the significance of information in these contexts. The importance of information has also been documented in various other industries where social learning plays a critical role in technology adoption and information spillover (Irwin and Klenow, 1994; Foster and Rosenzweig, 1995; Munshi, 2004; Conley and Udry, 2010). Additionally, this paper relates to earlier theoretical papers on learning and information externality, including Banerjee (1992), Bikhchandani, Hirshleifer, and Welch (1992), Bolton and Harris (1999), and Bessen and Maskin (2009).

The third strand of literature that this article contributes to is the recent advancements in industrial organization concerning oil and gas extraction. Several notable articles have made significant contributions to this field, including Kellogg (2014), Covert (2015), Anderson, Kellogg, and Salant (2018), Steck (2018), Bartik et al. (2019), Hodgson (2019), Compiani, Haile, and Sant'Anna (2020), Fetter et al. (2018), Agerton (2020), and Herrnstadt, Kellogg, and Lewis (2020). In particular, this article is closely related to Fetter et al. (2018). Both papers study the social learning effects following mandatory information disclosure interventions in the shale gas industry. While Fetter et al. (2018) study the disclosure of chemical inputs information, this article aims to provide a fresh perspective on the existing literature by examining the impact of production information disclosure policies, which has not been studied previously.

The rest of the article is organized as follows: Section 2 provides industry background and discusses the regulatory change. Section 3 presents the theoretical model. Section 4 describes the data. Section 5 presents the empirical models and results. Section 6 concludes the article.

2 Shale Gas Production in Pennslyvania

The "shale revolution" that has unfolded over the past two decades has significantly reshaped the global energy landscape. Pioneered by groundbreaking technologies like hydraulic fracturing and horizontal drilling, the thriving shale gas industry has propelled the US, once the world's largest natural gas importer, to become a leading exporter of liquefied natural gas (LNG) since 2017.[5]. In 2007, the gross withdrawals of natural gas in the US were on par with levels observed in the early 1970s. The amount has since almost doubled, with the share of gross withdrawals from shale gas burgeoning from 8% in 2007 to 81% in 2022[6].

Among the major shale gas fields in the US, the Marcellus Shale has consistently yielded the largest amount of gas production since 2013[7][8]. The Middle Devonian-aged sedimentary formation extends throughout most of the Appalachian Basin in the Northeastern US, with its depth ranging from 5000 to 9000 feet. Range Resources completed the first Marcellus well in Pennsylvania's Washington County in 2004, kicking off the wave of massive exploration in the following decade. Currently, the Marcellus shale contributes to around 21% of all US gross natural gas production[9], of which over 90% is from Pennsylvania.

In Pennsylvania, as with many other states, shale gas companies are required to secure mineral rights prior to initiating any substantive exploration activities. These rights are acquired by executing lease agreements with local landowners, thereby granting the firms the authority to conduct seismic surveys and drill for gas. Typically, a lease's duration spans three to five years. Should the production reach or exceed a stipulated minimum threshold, the lease is eligible for extension and remains in effect for the duration of ongoing production.

The financial commitment to procure these mineral rights is bifurcated into two components. The first, often referred to as the "bonus," is an initial payment made directly to the landowner. This sum is usually kept confidential for competitive business reasons. The second component is the "royalty," which is a stipulated share of the future revenue that the gas producer agrees to pay. Unlike the bonus, the royalty rate is a matter of public record, typically accessible for review alongside lease contracts at county courthouses.

Ever since the earliest endeavors to harness natural gas from shale formations, debates over the technically recoverable resources (TRR) have persisted, and the Marcellus Shale is a prime example of this contention. McGlade, Speirs, and Sorrell (2013) highlight the significant uncertainties that have shrouded estimates of the Marcellus Shale's TRR. This remains the case even with a substantial number of wells drilled and robust production levels established. A myriad of factors likely contribute to this unpredictability, ranging from the estimation methods and underlying as-

[5] See "Natural gas explained", EIA, accessible at `https://www.eia.gov/energyexplained/natural-gas/imports-and-exports.php`

[6] See data from EIA, accessible at `https://www.eia.gov/dnav/ng/NG_PROD_SUM_DC_NUS_MMCF_A.htm`

[7] See data from EIA, accessible at `https://www.eia.gov/naturalgas/data.php`

[8] See "Top 100 U.S. Oil and Gas Fields", published by EIA, 2015, accessible at `https://www.eia.gov/naturalgas/crudeoilreserves/top100/pdf/top100.pdf`

[9] See "Natural Gas Weekly Update", EIA, accessible at `https://www.eia.gov/naturalgas/weekly/archivenew_ngwu/2022/01_27/`

sumptions to the availability of empirical data. A particularly crucial factor has been the absence of public information on well productivity, a void that persisted until long after Range Resources executed its inaugural operation.

On November 1, 2010, the PADEP commenced the release of production information, inaugurating this practice with a well-specific production report that encompassed the shale gas output from all wells across Pennsylvania for the period of July 2009 to June 2010. Subsequently, this dataset has been updated biannually, marking a watershed in transparency for the industry. For the first time, shale gas firms operating in Pennsylvania had the ability to monitor not only their production metrics but also to gain insights into the performance of every individual well within the state.

3 Theoretical Model

This section introduces a straightforward learning model designed to elucidate the manner in which access to augmented production data can enhance the efficiency of lease investments. The discourse primarily revolves around a simplified model involving just two firms, leaving the analysis of the more general multi-firm case to the appendix.

3.1 Basic Setting

N firms are assumed to sequentially choose between two locations, L_0 and L_1, to make their lease investment. Without loss of generality, the potential outputs of L_0 and L_1 are assumed to be $Q_0 = 0$ and $Q_1 = 1$. The firms have identical prior beliefs $\Pr(Q_0 = 1) = \Pr(Q_1 = 1) = 1/2$.

For a given investment order, let $d_i \in \{0, 1\}$ denote operator i's location choice. Before deciding on which location to invest in, it is assumed that firm i observes d_k, $k < i$, and a conditionally independent private signal $s_i \in \{0, 1\}$ suggesting which location is more preferable, likely resulting from past investment experiences. It is known to all firms that $\Pr(s_i = j|Q_j = 1) = p > 1/2$, $j = 0, 1$. $r(\cdot)$ is the common inverse supply function of mineral rights at both locations that maps a firm's investment order to its royalty cost. Assume $r(\cdot)$ is increasing, $r(0) = 0$, and $r(N) < 1$. Each firm i conducts Bayesian updating whenever possible and chooses d_i to maximize its posterior expected return:

$$\max_{d_i} E[\pi_{i0}(1 - d_i) + \pi_{i1} d_i | s_i, d_k, k < i], \tag{1}$$

where $\pi_{i0} = (1 - r(i - \sum_{k \leq i} d_i))Q_0$ and $\pi_{i1} = (1 - r(\sum_{k \leq i} d_i))Q_1$. The overall investment efficiency can be measured by $E[\sum_{i=1}^{N} d_i]$: for a given N, the more firms choosing location L_1, the more efficient the investment of the entire industry will be[10].

[10] There is no fixed cost assumed in this model for simplicity. Fixed cost is an essential source of inefficiency with uninformed location choices. The basic arguments of the model hold if fixed costs are similar at the two locations.

3.2 Two-Firm Case

This subsection considers a simplified special case with only two firms. The following proposition shows that inefficiency may prevail when there is a lack of information and resources are scarce.

Propsition 1. *In the two-firm setting, at least one firm chooses the less productive location L_0 when $r(1)$ is high.*

Proof. Notice that given identical priors, firm 1 always chooses L_{s_1}. If $s_1 = 0$, firm 1 chooses L_0. If $s_1 = 1$, firm 2, observing d_1, is able to infer s_1. If $s_2 = 0$, firm 2 has identical posteriors. Since firm 1 has chosen L_1 and $r(\cdot)$ is increasing, firm 2 chooses L_0.

If $s_2 = 1$, Firm 2's posterior expected returns become

$$E[\pi_{20}|s_2 = s_1 = 1] = \frac{(1-p)^2}{p^2 + (1-p)^2},$$
$$E[\pi_{21}|s_2 = s_1 = 1] = (1 - r(1))\frac{p^2}{p^2 + (1-p)^2}.$$

Thus, when $r(1) > \underline{r} \equiv 1 - (1-p)^2/p^2$, firm 2 chooses L_0 again. Overall, when $r(1) > \underline{r}$, at least one firm will choose L_0. □

The above proposition shows that when investment costs are high and investment quality is uncertain, the overall investment outcome is likely to be inefficient. The next proposition discusses how efficiency can be improved when additional information adjusts firms' prior beliefs.

Propsition 2. *Assume a pair of productivity signals $q_1 > q_0$ is disclosed that changes both firms' prior beliefs to*

$$Pr(Q_1 = 1) = q_1/(q_1 + q_0),$$
$$Pr(Q_0 = 1) = q_0/(q_1 + q_0).$$

Then there is a chance that both firms will choose L_1 if q_1/q_0 is high.

Proof. It is easy to see that when $s_1 = 1$, firm 1 chooses L_1. If $s_1 = 0$, firm 1's posterior belief is

$$\Pr(Q_1 = 1|s_1 = 0) = \frac{(1-p)q_1}{(1-p)q_1 + pq_2}.$$

Hence, when $q_1/q_0 > p/(1-p)$, firm 1 always chooses L_1, and d_1 is uninformative about s_1.

Next, if $s_2 = 1$, firm 2's posterior expected returns become

$$E[\pi_{20}|s_2 = 1] = \frac{(1-p)q_2}{pq_1 + (1-p)q_2},$$
$$E[\pi_{21}|s_2 = 1] = (1 - r(1))\frac{pq_1}{pq_1 + (1-p)q_2}.$$

When $r(1) < \overline{r} \equiv 1 - \frac{(1-p)q_2}{pq_1}$, firm 2 will choose L_1. □

The proposition above outlines the conditions under which the disclosed productivity information reshapes the likelihood of choices. It is easy to verify that when $q_1/q_0 > p/(1-p)$, $\overline{r} > \underline{r}$. Therefore, when $r(1) \in (\underline{r}, \overline{r})$, there is a positive probability that the most efficient investment decisions are made with the additional productivity signals, an unlikely outcome if such information is concealed. Moreover, when q_1/q_0 is high, the least efficient result is avoided in the sense that at least one firm chooses L_1. Without the productivity information, it is possible that both firms choose L_0 when $r(1) < \underline{r}$. Clearly, $E[\sum_{i=1}^{2} d_i]$ is greater if the disclosed productivity information provides sufficient knowledge about the qualities of the two locations.

3.3 Multiple-Firm Case

In scenarios involving multiple firms, analytical deduction becomes more complex. To address challenge, a simulation analysis is conducted to examine the quantitative impact of additional information on firms' location decisions. Detailed results of this analysis are presented in the appendix. The core insight aligns with that of the two-firm case: firms that recalibrate their prior beliefs based on productivity data are more likely to select the higher-quality location, which in turn commands a higher royalty rate.

Further analysis using the simulated data reveals a notable finding: with initially incomplete information, there is a significant shift in the correlation between firms' choices and the productivity signals they receive as their prior beliefs are updated. Conversely, when firms start with more accurate priors, this correlation exhibits minimal change. These findings inform the subsequent empirical analysis and provide a foundation for the empirical model.

4 Data

The empirical analysis draws upon various sources of information to investigate the influence of the PADEP's productivity information disclosure on firms' site selection behaviors. This section provides an overview of the datasets employed in the study, with a comprehensive account of the data refinement procedures relegated to the appendix.

The initial dataset encompasses drilling permits authorized by the PADEP spanning from January 2007 through December 2014. This dataset serves to identify operators dedicated to shale gas exploration, given that the PADEP issues unconventional permits exclusively for wells that aim at shale formations. Table 1 presents the distribution of unconventional drilling permits amongst the operators.

The second dataset comprises lease contracts executed in Pennsylvania during the period from January 2007 to December 2014. The data, courtesy of Enverus (formerly known as Drillinginfo), an energy analytics firm, encompass a wealth of details. Each lease contract records the acreage, associated operator, precise location (county and municipality), commencement date, royalty rate, and lease duration. A notable limitation is the absence of bonus value information, which remains confidential and undisclosed by operators. To ensure relevance to shale gas extraction, leases attributed to operators with fewer than 10 unconventional permits within the timeframe are excluded from the analysis. Table 2 summarizes the annual quantity of leases, their average sizes, the durations of the contractual terms, and the size-weighted average royalty rates. Lease agreements surged to over 20,000 in 2008, subsequently declined to below 5,000 by 2012, and then increased again to above 10,000 in 2014. The average lease size commenced at approximately 70 acres in 2007, diminishing to under 30 acres from 2009 onwards, while the average royalty rate remained in the vicinity of 15%. The details of constructing the lease sample are given in the appendix.

The third dataset encompasses nine production reports released by the PADEP from 2010 through 2014. Table 3 outlines the publication dates of these reports alongside the production intervals they encompass. The inaugural report, issued in November 2010, detailed production data from July 2009 to June 2010. Subsequent reports were disseminated biannually, in February and August, from 2011 to 2014, each encapsulating a six-month production timeframe.

Table 4 offers summary statistics derived from these production reports. It is important to differentiate between the two primary classifications of natural gas wells: vertical and horizontal. These categories are aligned with conventional and unconventional drilling permits, respectively, and are distinguished by their technical specifications and extraction objectives. Vertical wells predominantly target conventional natural gas reservoirs, whereas horizontal wells are designed for shale gas exploitation. To maintain clarity and focus, the empirical analysis will exclusively consider the productivity data of horizontal wells. The spatial correlation of well productivity, along with additional methodological considerations, is addressed in the appendix.

Municipalities are the smallest geographic units in the empirical analysis. The gazetteer of Pennsylvania is obtained from the US Census Bureau, which provides manageable indices for the counties and municipalities in the state, along with their land areas, longitudes, latitudes, and populations in the 2010 Census[11]. Additionally, monthly natural gas spot prices from the Henry Hub are utilized as proxies for wellhead prices within Pennsylvania. A detailed exploration of the interplay between gas prices and leasing decisions can be found in the appendix.

5 Empirical Analysis

This section adopts a reduced-form approach to empirically examine the impact of production information disclosure on shale gas operators' location choices. It contains two segments. As a motivating exercise, it begins by inspecting the correlation between productivity and the cost of acquiring leases. The underlying hypothesis posits that if the dissemination of production information enhances the industry's collective understanding of shale gas reserve distribution, the correlation should become stronger. To obtain a quantitative understanding of the specific impact, an autoregression model is proposed and estimated in the second part of the analysis, inspired by and aimed

[11] There are three types of municipalities in Pennsylvania: cities, boroughs, and townships. It is possible for a city and a township or a city and a borough to be in the same county and have the same name. For these cases, the pair of municipalities is treated as a single unit with their population sizes and land areas added up. The midpoint of the longitudes and latitudes of the two municipalities is used as the coordinates of the combined unit.

at verifying predictions of the theoretical model.

5.1 Royalty Rates and Productivity

Figure 1 plots the acreage-weighted average royalty rate against the average log daily production of each municipality in Pennsylvania where the values of both variables are observed. The productivity is calculated using the ninth report covering production from January to June 2014, which can be understood as the "real" productivity of a municipality. The royalty rates of the circular dots are calculated using the leases granted between January 2007 and October 2010, the sample period before the information disclosure. The crosses represent royalties calculated using the post-disclosure leases granted between November 2010 and December 2014. The dashed and the solid curves are the fitted linear predictions, with 95% confidence intervals given by the shaded regions. The figure shows that royalty rates are more responsive to productivity when the production information is publicly available. The possible mechanism behind the comparison is that if the production information generates additional knowledge about the quality of the shale gas reserves, locations with higher productivity will attract greater lease investment, leading to higher royalty rates.

5.1.1 Model and Estimation

More formally, consider the following linear regression model:

$$Royalty_{jt} = \beta_0 + \beta_1 \log(Prod_j^9) + \beta_2 D_t + \beta_3(\log(Prod_j^9) \times D_t) + \varepsilon_{jt}, \tag{2}$$

where t stands for either "pre-disclosure" or "post-disclosure", while j stands for the municipality. $Royalty_{jt}$ is the average royalty rate of municipality j in period t, weighted by the lease acreages. $\log(Prod_j^9)$ is the average log daily production (Mcf./d) of municipality j, calculated using the ninth production report, and is fixed for each municipality. D_t is the disclosure indicator variable that takes a value of 0 for "pre-disclosure" and 1 for "post-disclosure". ε_{jt} is an unobserved error term and is assumed to be i.i.d. normal across municipalities and periods. The coefficient β_3 captures the impact of the production information disclosure on the correlation between lease cost and productivity. A positively significant β_3 suggests an informative disclosure.

The estimation results are shown in Table 5. The coefficients are all statistically significant, though economically their values are relatively small. As clearly shown in Figure 1, there is tremendous heterogeneity in average productivity across municipalities, and some places have a rate of production over ten times higher than others. The considerable variation in production rates explains the small scales of the estimated coefficients. According to the estimation, a 100% increase in daily production is accompanied by around 0.26%[12] increase in the absolute value of the royalty rate prior to the information disclosure, or slightly lower than 2% of the average royalty rate at that time. This marginal effect increases dramatically by 0.40%[13] or 1.54 times its original value in

[12]Since productivity is measured in natural log forms, a 100% increase in productivity leads to $\ln(2) \times 0.37\% \approx 0.26\%$ increase in the royalty rate.

[13]$\ln(2) \times 0.58\% \approx 0.40\%$.

the post-disclosure period. The sharp contrast indicates that the disclosed production information does make shale gas operators pour greater resources into more productive areas.

5.1.2 Robustness Analysis

To evaluate the robustness of the above result, this subsection first studies the following regression model that measures the correlation between productivity and royalty rate on a yearly basis:

$$Royalty_{jt} = \beta_0 + \beta_{1t} \log(Prod_j^1) + \beta_t + \varepsilon_{jt}, \tag{3}$$

where $Royalty_{jt}$ is the weighted average royalty rate of municipality j in year t, and $\log(Prod_j^1)$ is the average log daily production (Mcf./d) of municipality j, calculated using the first production report. The coefficient of productivity β_{1t} is allowed to be year-specific. β_t represents the year fixed effects, and ε_{jt} is an unobserved i.i.d. normal random term. The purpose of using the first report is to examine the impact of the initial production report disclosure on the cost of acquiring leases. The hypothesis is that the coefficient β_{1t} of the years 2010 and 2011 shall be positively significant if the first production report generates shocks to the information sets of the shale gas operators, whereas β_{1t} of the previous few years may not be positive or significant.

For the year-by-year regression model above, the estimated productivity coefficient in 2007 is -0.0012 with a standard error of 0.0027. This suggests that there is no statistically significant correlation between the royalty rate and productivity in that year. Figure 2 displays the coefficient estimates of the interaction between productivity and the year fixed effects. The round dots connected by the solid line represent the changes in the values of the productivity coefficient estimate in subsequent years relative to the 2007 estimate. The vertical line represents the time of the initial production data disclosure.

As the figure shows, the productivity coefficients in 2008 and 2009 are slightly greater than the coefficient in 2007, though the increments are not statistically significant. The coefficient in 2010 is greater than the 2007 value by 0.0088, and the increase jumps to 0.0137 in 2011. Both estimates are statistically significant. The productivity coefficients in 2012, 2013, and 2014 are not significantly different from the value in 2007. The results suggest that right after the first production report was disclosed, there was a strong enough adjustment in shale gas operators' lease location choices that the correlation between royalty rate and productivity disclosed by the report surged. The change is significant both statistically and economically. The jump in the correlation does not continue into the years after 2011. A likely explanation is that the disclosed production information introduces a shock to the firms' understanding of the shale gas reserve distribution, which was completely digested after 2011.

To further assess the robustness of the outcome, a difference-in-difference analysis is conducted. The sample of municipalities is divided into subgroups that are "shale-gas-oriented" and "non-shale-gas-oriented", based on the number of active horizontal wells[14]. Because the disclosed production

[14] There are 32 counties in which at least one horizontal well is active from January to June 2014. The municipalities in the 12 counties, each with more than 50 active horizontal wells, are classified as shale gas oriented.

information is shale-gas-specific, it generates additional knowledge about the shale gas reserve distribution only and should only affect leases taken for shale gas production. By comparing the correlation between royalty rates and productivity across the two sub-samples, the impact of information disclosure can be identified.

Specifically, the following regression model is considered:

$$Royalty_{jt} = \beta_0 + [\beta_1 + \beta_2 S_j + \beta_3 D_t + \beta_4 (S_j \times D_t)] \log(Prod_j^9) + \varepsilon_{jt}, \quad (4)$$

where t is either "pre-disclosure" or "post-disclosure", and j stands for the municipality. $Royalty_{jt}$, $Prod_j^9$, and D_t are defined as before. S_j is the municipality type indicator and is equal to 1 if municipality j is classified as "shale-gas-oriented". ε_{jt} is assumed to be i.i.d. normal[15]. The coefficient of productivity is allowed to vary both over time and across groups of municipalities. The critical parameter in this model is β_4, which captures the distinction in the impacts the production information disclosure had on the locations of leases taken for different purposes.

Table 6 reports the coefficient estimates of the above model. The coefficient of the interactive variable between the municipality type indicator and productivity is so small and insignificant that there is no discernible difference in how royalty rates respond to productivity between the two types of municipalities before the production information is disclosed. The coefficient of the interaction between the disclosure dummy and productivity is also not significant, which shows that in the "non-shale-gas-oriented" municipalities, the disclosure of the production information has no significant impact on lease location choices. By contrast, the coefficient of the triple interactive variable that multiplies productivity with both indicators is positive and significant. The result shows that compared to "non-shale-gas-oriented" municipalities, in those municipalities devoted to shale gas production, the production information disclosure makes royalty rates one-third higher when productivity doubles. This finding is consistent with the conjecture that the production information disclosure only affects land leasing decisions targeting shale gas production.

Finally, there may be an endogeneity concern if production decisions are affected by royalty costs in reverse. To address this issue, model (3) is re-estimated using production information from the first report to minimize the impact of higher lease rates on production decisions. The results are given in Table 7. Figure 3a presents the analogous plot as Figure 1, using productivity calculated from the first report. The first thing to note is that there are considerably fewer observations than when using the ninth report because there are many municipalities where no production ever took place by the time the first report was released. Additionally, there is no significant difference in the correlation between royalty rates and productivity over time, and Column 1 of Table 7 confirms this observation.

A closer look at the figure, however, reveals that this insignificance is mainly due to the set of

[15] There may be a concern that the classification of "shale-gas-oriented" locations is endogenous. One fact that alleviates this concern is that after the first few years of exploration and production (2004-2007), geologists and industry participants have had a broad picture of the shale gas reserve distribution in Pennsylvania. What is unclear is how the deposit's quality varies within this broadly defined "shale gas-oriented" region. Viewed in this way, the classification of the treatment group can be regarded as exogenous in general.

municipalities where the royalty rates are set at or close to the lowest possible values. If the same analysis is conducted for the subset of municipalities where the average royalty rates are higher than 13.5%, the difference becomes significant again, as shown by Figure 3b. Column 3 of Table 7 reports the estimation results using this selected subsample. The results are comparable to those in Table 5. What may be happening here is that the first production report does not show a comprehensive picture of the state's reserve distribution, and there is a minimum royalty requirement of 12.5%.

5.2 Location Choices and Productivity

The last subsection provides preliminary evidence of the potential influence of the production information disclosure on shale gas firms' leasing decisions. This subsection takes a closer look at such effects by modeling firms' leasing decisions directly. The empirical model resembles the theoretical framework and asks how the disclosed information might change firms' beliefs about shale gas reserve distribution, represented by their past leasing experience. The crucial information structure here is that firms have always been perfectly informed about the lease-taking histories of each other,[16] but the production information is not available until around the middle of the sample period.

5.2.1 Model

This subsection presents an autoregression model that seeks to directly measure the quantitative impacts of the information disclosure. Let $\mathcal{I}$ denote the set of shale gas operators, $\mathcal{J}$ denote the set of municipalities, and $t = 1, 2, ..., 96$ denote the months between January 2007 and December 2014. Let a_{ijt} be the acreages of the leases granted to operator i in municipality j and month t, measured in acres. Define $\mathcal{O}_t = \{i \in \mathcal{I} | \sum_{j \in \mathcal{J}} a_{ijt} > 0\}$ as the set of operators that acquire positive lease acreages in month t. Let d_{jk} be the distance in miles between municipalities j and k. For each municipality $j \in \mathcal{J}$, define $\mathcal{N}_j = \{k \in \mathcal{J} | d_{jk} \leq 5\}$ to be the set of its neighboring municipalities[17]. The production information is disclosed at $t = 47$.

In each month t, each operator $i \in \mathcal{O}_t$ decides a_{ijt}, the acreages of the leases it shall take in each municipality $j \in \mathcal{J}$, which is assumed to be determined by the following model:

$$\log a_{ijt} = \alpha_i + \alpha_t + \xi r_{jt} + \boldsymbol{S}_{ijt}\boldsymbol{\delta} + \theta \overline{\log q}_{jt} + \left(\overline{\log q}_{jt} \times \boldsymbol{S}_{ijt}\right)\boldsymbol{\gamma} + \boldsymbol{X}_{ijt}\boldsymbol{\beta} + \varepsilon_{ijt}. \tag{5}$$

In the above model[18], r_{jt} represents the average royalty rate of the location-time pair (j, t). $\boldsymbol{S}_{ijt}$ is a vector of signals regarding the shale gas reserve quality of location j. $\overline{\log q}_{jt}$ is the average log daily production of the municipalities in $\mathcal{N}_j$, calculated using the most recent production report available in month t. Productivities in the neighboring municipalities are included because of their strong positive spatial correlation to better characterize the output potential of location j. $\boldsymbol{X}_{ijt}$ is

[16] Copies of lease contracts are stored in county courthouses and hence can be regarded as public information. Alternatively, firms can turn to energy consulting firms like Everus to obtain information about leases.

[17] The construction of the neighboring municipality set is meant to explore the spatial correlation of productivity and build up a more robust productivity signal.

[18] To accommodate for zero values of a_{ijt}, $\log(a_{ijt} + 1)$ is used in estimation. Likewise, $\overline{\log(q_{jt} + 1)}$ is used in place of $\overline{\log q}_{jt}$.

a vector of control variables, including natural gas price p_t, population size Pop_j, and land area $Aland_j$ of municipality j, and an indicator variable Pub_i that takes the value of 1 if operator i is public. α_i is the operator fixed effect, and α_t is the time fixed effect. ε_{ijt} is an unobserved random term that is assumed to be i.i.d. normal both cross-sectionally and over time.

Inspired by the theoretical model, past leasing decisions are used as a proxy for the quality signals. The vector of signals $\boldsymbol{S}_{ijt} = \left(\log A_{ijt}^{Recent}, \log A_{-i,jt}^{Recent}\right)$ includes the following variables:

1. Own recent lease acreages A_{ijt}^{Recent} : the lease acreages operator i obtained in municipality j in the past three months[19], defined by $A_{ijt}^{Recent} = \sum_{\tau=t-3}^{t-1} a_{ij\tau}$
2. Peer's recent lease acreages $A_{-i,jt}^{Recent}$: the lease acreages obtained by operators other than i in municipality j in the past three months, defined by $A_{-i,jt}^{Recent} = \sum_{\tau=t-3}^{t-1} \sum_{k\in\mathcal{I}\setminus\{i\}} a_{kj\tau}$

The construction of the signal vector is based on the assumption that leasing decisions reflect operators' beliefs regarding a location's output potential and should be correlated over time. The components represent contemporary signals of the shale gas reserve quality. A larger recent lease acreage indicates that the operator itself or its competitors are more likely to have private information that lease investment in the municipality is promising. The two types of leases are separated to distinguish between the effects of signals from different sources.

The above model examines whether the disclosed production reports contain information about reserve quality that was not fully considered in earlier leasing decisions represented by $\boldsymbol{S}_{ijt}$. If firms have perfect knowledge about the shale gas reserve distribution through learning by doing and observing the location choices of each other, additional production information will not add much value to their leasing decisions. A positive coefficient of productivity would thus indicate the opposite: when production reports are released, firms reallocate their leasing investment to locations with higher productivity, despite the cost likely being higher. In other words, the production reports update the knowledge sets of shale gas firms and adjust their location choices[20].

The inclusion of the interaction between productivity and the signal vector is motivated by the estimation results obtained from the simulated sample described in the appendix. This addition aims to capture the firms' level of knowledge about the deposit distribution in the absence of production information. As discussed in the appendix, when firms have limited initial information, the correlation between their location choices and the signals they receive will be significantly influenced by introducing the productivity signal, leading to a non-zero coefficient for the interaction term. Conversely, if firms possess relatively accurate prior beliefs, the coefficient of the interaction term is likely to be insignificant.

5.2.2 Identifiation

A critical identifying assumption of model 5 is that the unobserved term is uncorrelated with the included independent variables. This assumption is strong because it overlooks the potential

[19] For notational simplicity, a_{ijt} is defined to be 0 for $t = -2, -1, 0$.

[20] An operator can observe its own production even when $t < 47$. However, this information is not available in the data and, therefore, is not considered in the model.

for serial correlation in the unobserved term, given the autoregressive nature of the model. This assumption may be questionable if the contemporary shock persists into future periods. To address this concern, in the robustness analysis, the past lease acreages used to formulate the signal vector $\boldsymbol{S}_{ijt}$ are redefined to exclude leases from the most recent months. This adjustment aims to avoid possible correlation with the current period's shock.

Another possible source of endogeneity is that the unobserved term includes technology improvement, which may affect a firm's resources to acquire more leases and can be correlated with productivity at the same time. However, two facts help alleviate this concern. Firstly, the productivity variable, calculated using the most recently disclosed production report, reflects the average production rate of a location over the past six months[21]. Therefore, by definition, it is not correlated with the technology improvement in the current period. Secondly, if the technology improvement occurs only within a particular firm, it would not generate a strong correlation with the *average* productivity of a location, given the presence of many other firms. On the other hand, if the innovation is industry-wide, it would affect the productivity of not only one location but the entire state. In either case, the extra financial resources resulting from the technology improvement would not be allocated solely to a specific location but would be spread across all municipalities based on the proposed decision-making process. Therefore, even if the endogeneity of productivity exists, it may not be a significant concern in this model.

Finally, the royalty rate may become endogenous if the unobserved term contains location-specific reserve quality information that is known to firms but not to researchers. To address this endogeneity concern, the per capita income of the county where the municipality is located is utilized as an instrument for the royalty rate. The rationale behind this choice is that the per capita income reflects the opportunity cost of the landowner's bargaining time and is expected to be correlated with the cost of acquiring leases. Specifically, a higher per capita income suggests that the landowner is likely to be less demanding in splitting future revenue, resulting in a lower royalty rate being charged.

5.2.3 Estimation Results

Table 8 presents the estimation results of model (5) . Column 1 represents the first-stage result, where the dependent variable is the royalty rate. Column 2 displays the benchmark result. In the third and fourth columns, either the operator or the time-fixed effect is excluded. Finally, Column 5 shows the estimation result without instrumenting for the royalty rate variable. All standard errors are robust and clustered at the municipality level.

In all of the estimation results, the coefficients for the firm's own recent lease acreages in the same municipality are consistently positive and statistically significant, ranging from 0.4270 to 0.4330. Given the lease acreages are in their natural log forms, the estimated values suggest that a 1%

[21] In fact, the productivity value measures a location's average production condition *more* than six months ago since the disclosure time of a production report is two months after the production period it covers. For example, the report disclosed in August covers production from January to June in the current year, and the report disclosed in February covers the production from July to December last year.

increase in the firm's lease acreage in the past three months leads to about 0.43% increase in the firm's current month's lease acreage in the same municipality. This significant positive correlation is not unexpected, as a firm's recent leasing decisions encompass valuable private information about its perception of a location's reserve quality, along with other considerations such as economies of space. Consequently, these leasing decisions provide informative signals regarding the firm's decision-making in the current period.

In comparison, the coefficients of the competitors' recent lease acreages are relatively smaller in value, although they remain consistently positive and statistically significant. However, it is important to note that the magnitude of all other firms' lease acreages combined is much larger than the firm's own lease acreage alone. Therefore, the contrast in coefficients should not be interpreted too literally. The results show that firms do gain insights from the location choices of other firms, although the impact of this external information is weaker compared to the influence of their own experience and decision-making.

The most compelling evidence of the impact of production information disclosure on firms' location choices is the consistently positive and significant coefficients of productivity in all scenarios. While the positive effect is somewhat less pronounced when time-fixed effects are not considered and diminishes further if the endogeneity of royalty rates is not addressed, the significance of the results remains. The results indicate that even after accounting for the private signals conveyed by a firm's own recent leasing decisions and the less precise signals derived from other firms' actions, the production reports still provide meaningful insights into the distribution of reserve quality. This information is significant enough to influence firms' leasing decisions.

On average, the results suggest that a doubling of productivity revealed by the most recent production report leads to approximately a 0.24%[22] increase in lease acreages for each operator in a specific municipality. While the magnitude may not appear substantial, it is important to remember that in many periods and locations, the dependent variable has a value of 0. Therefore, the impact is significant and represents more than half of the marginal effect of the recent lease acreage of competing firms.

In contrast, the lack of significance in the coefficient of the interaction between productivity and a firm's own recent lease acreage, both economically and statistically, suggests that firms' reliance on their own private signals remains largely unchanged following the disclosure of production information. This implies that firms' prior beliefs concerning a location's potential output fall within a knowledgeable range, as depicted in Figure 5.

Lastly, the coefficients of the interaction between productivity and peers' recent lease acreage indicate that, if anything, the production information disclosure reduces firms' reliance on other firms' decisions as an indirect signal for making their own choices, although the effect size is relatively small.

The estimation results of other coefficients align with intuitive expectations. For instance, the benchmark outcome in the second column reveals that, on average, leases acquired by public firms

[22] $\ln 2 \times 0.35\% \approx 0.24\%$

are 0.43% larger than those obtained by private firms. This outcome can be attributed to the fact that public firms are typically larger and possess more abundant financial resources.

Similarly, the results indicate that a 1% increase in royalty rates leads to a decrease of 2.2% in lease areas. It is crucial to address the endogeneity of the royalty rate, as demonstrated by the positive coefficient in the last column when two-stage least square estimation is not utilized.

An unexpected finding is that the coefficient of natural gas price, although positive, is statistically insignificant. One possible explanation for this finding is that a firm's expectation of the current period's natural gas price is considered part of its private information and has already been incorporated into its recent leasing decisions. Thus, when controlling for these decisions, the natural gas price itself does not have a significant impact.

5.2.4 Robustness Analysis

The remaining part of the section discusses a series of robustness tests of the estimation results in the last subsection. The first set of tests seeks to deal with the validity and potential endogeneity of the signal vector. Table 9 reports the estimation results of modified versions of model (5) where the signal vector $\boldsymbol{S}_{ijt}$ is adjusted. As part of the effort to address the endogeneity concern when using recent lease acreages as the signal, vector $\boldsymbol{S}_{ijt}$ is redefined to include earlier lease acreages. The first column of the coefficient estimates is the benchmark result and is the same as Column 2 of Table 8. In the second column, the recent lease acreages refer to the areas of leases acquired in the three-month period preceding the past three months. Specifically, $\boldsymbol{S}_{ijt} = \left(\log A_{ijt}^{Recent}, \log A_{-i,jt}^{Recent}\right)$ is redefined to include:

1. Own recent lease acreages A_{ijt}^{Recent} : the lease acreages operator i obtains in municipality j within the time frame of the past six and past three months, defined by $A_{ijt}^{Recent} = \sum_{\tau=t-6}^{t-4} a_{ij\tau}$
2. Peer's recent lease acreages $A_{-i,jt}^{Recent}$: the lease acreages operators other than i obtain in municipality j within the time frame of the past six and past three months, defined by $A_{-i,jt}^{Recent} = \sum_{\tau=t-6}^{t-4} \sum_{k\in\mathcal{I}\setminus\{i\}} a_{kj\tau}$

The restructuring of the signal vector has minimal impact on the estimation results, indicating that the potential endogeneity issue, if any, is not a major concern. The only considerable change comes from the coefficient of the interaction between productivity and the firm's own recent lease acreage, which is now positive and significant. This change is likely due to the fact that the new set of signals, which includes earlier lease acreages, is weaker, and firms are less well-informed standing in earlier periods.

The last column of Table 9 presents the estimation results of a model with a more drastic alteration of the signal vector $\boldsymbol{S}_{ijt}$. In this case, instead of using recent lease acreages as a proxy for private signals, the model incorporates cumulative acreages up to the past three months. This modification allows for a broader consideration of the leasing history and provides a different perspective on the impact of private signals on firms' location choices. Specifically, $\boldsymbol{S}_{ijt} = \left(\log A_{ijt}^{Cumulative}, \log A_{-i,jt}^{Cumulative}\right)$ is redefined to include:

1. Own cumulative lease acreages $A_{ijt}^{Cumulative}$: the cumulative lease acreages operator i obtains in municipality j up to the past three months, defined by $A_{ijt}^{Cumulative} = \sum_{\tau=-3}^{t-4} a_{ij\tau}$
2. Peer's cumulative lease acreages $A_{-i,jt}^{Cumulative}$: the lease acreages operators other than i obtain in municipality j up to the past three months, defined by $A_{-i,jt}^{Cumulative} = \sum_{\tau=-3}^{t-4} \sum_{k\in\mathcal{I}\setminus\{i\}} a_{kj\tau}$

As shown in the last column, the coefficients of productivity and the firm's own cumulative lease acreage remain positive and significant. Moreover, the coefficient of productivity becomes even greater, indicating a stronger influence of the production information disclosure. This result highlights the robustness of the positive impact of productivity information on firms' leasing decisions across various configurations of the signal vector.

However, the coefficient of lease acreages acquired by other firms is no longer significant, and the coefficient of the interaction between productivity and the firm's own cumulative lease acreage turns negative. Ane possible explanation for this change is that when the signal vector includes leases acquired a long time ago, the information contained within becomes so outdated that it either becomes less relevant to the current period's decision-making or needs to be updated by the disclosed production information.

The second set of tests examines the robustness of the results when the sample size is reduced. Column 1 of Table 10 presents the benchmark estimation results obtained from the full sample. Columns 2 to 4 in the table display the estimation results from subsamples, with observations from the first and last one, two, and three years excluded iteratively.

The purpose of this test is to determine whether the positive impact of productivity on the firm's leasing decision is primarily driven by the disclosure of production information, which occurs in the middle of the sample period, rather than being influenced by outliers or irregular data patterns at the tails of the sample. The consistently positive and significant coefficients of productivity indicate that even when the sample is trimmed to include only observations one year before and one year after the production information disclosure, the impact of productivity information remains significant.

The last set of tests involves dividing the sample into pre- and post-disclosure periods and estimating model (5) separately using the two subsamples. In the pre-disclosure observations, the values of the productivity variable are constructed based on the information later revealed by production reports. This experiment aims to investigate whether productivity information has an impact on firms' leasing decisions *before* it is disclosed. The conjecture behind this experiment is that since productivity is not observed prior to its disclosure[23], it should not influence firms' location choices once their private signals, represented by their past leasing decisions, are taken into account.

Table 11 presents the estimation results. Column 1 represents the benchmark result obtained from the full sample. In Column 2, the estimation result for the post-disclosure subsample is provided. The coefficient of productivity remains positive and significant, with its value being

[23]Firms do observe the productivity of their own wells even before the disclosure. Hence the test targets on the *extra* information of reserve quality revealed by the production reports.

nearly double that of the full-sample case. The obtained result is intuitive as it aligns with the notion that firms only start actively learning from the production information during the post-disclosure period. This suggests that once the production information becomes available, firms can effectively incorporate it into their decision-making process, leading to a stronger impact on their leasing outcomes.

Columns 3 and 4 in Table 11 display the estimation results for the pre-disclosure subsample, where productivity is computed using the first and last production reports, respectively. As anticipated, in both cases, productivity loses its significance as a determining factor in firms' leasing decisions. This finding suggests that during the pre-disclosure period, firms either lack a comprehensive understanding of the reserve distribution or do not heavily rely on productivity when selecting lease locations. Once again, this comparison reinforces the importance of the production information disclosure in reshaping firms' leasing decisions. It highlights that the availability of production information plays a crucial role in influencing firms' decision-making processes and subsequent location choices.

6 Conclusion

Information sharing is a highly debated topic and is a feature of many markets and regulatory structures. Existing literature suggests a variety of models to inform the impacts of information sharing, with its consequences ranging from improving efficiency to facilitating collusion. It remains an empirical question as to how information disclosure affects market outcomes in a specific industry. This article addresses this fundamental issue within the context of shale gas exploration. Through an investigation of a transformative intervention that mandates the disclosure of production data, the study reveals that firms actively learn from the production information, leading to more informed location choices and more efficient investment decisions. The findings indicate that in the early stages of an industry where exploration and experimentation are crucial, mandatory information sharing is likely to enhance the efficiency of resource allocation.

The current research does not address two fundamental questions in the shale gas exploration industry. The first question pertains to how royalty rates are determined. The theoretical model proposed in this article assumes a simplified upward-sloping supply curve. However, in practice, royalty rates and bonus payments are outcomes of the bargaining process between operators and landowners. The results from Myerson and Satterthwaite (1983) imply that when disclosed production information alleviates information asymmetry, negotiations may be facilitated. Addressing the negotiation problem requires a comprehensive structural model, which goes beyond the scope of this research.

The second question relates to the presence of strategic concerns within a dynamic framework. The theoretical framework presented in the article assumes one-shot information sharing. However, in reality, information sharing may be persistent or periodic over time. In such cases, firms may have incentives to strategically delay their decision-making in order to gain additional information.

Exploring these strategic considerations within a dynamic framework is crucial.

Both of these questions are important as they directly impact the measurement of the welfare consequences of production information disclosure and the utility gains or losses for the operators and landowners. Answering these questions requires further research and investigation.

Declaration of generative AI and AI-assisted technologies in the writing process:
During the preparation of this work the author used ChatGPT 3.5 in order to improve the language. After using this tool/service, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

References

Agerton, Mark (2020). "Learning where to drill: Drilling decisions and geological quality in the haynesville shale". In.

Akerlof, George A. (1970). "The Market for "Lemons": Quality Uncertainty and the Market Mechanism". In: *The Quarterly Journal of Economics* 84(3), pp. 488–500. ISSN: 00335533, 15314650. URL: http://www.jstor.org/stable/1879431.

Anderson, Soren T, Ryan Kellogg, and Stephen W Salant (2018). "Hotelling under pressure". In: *Journal of Political Economy* 126(3), pp. 984–1026.

Banerjee, Abhijit V. (1992). "A Simple Model of Herd Behavior". In: *The Quarterly Journal of Economics* 107(3), pp. 797–817. ISSN: 00335533, 15314650. URL: http://www.jstor.org/stable/2118364.

Bartik, Alexander W et al. (2019). "The local economic and welfare consequences of hydraulic fracturing". In: *American Economic Journal: Applied Economics* 11(4), pp. 105–55.

Bessen, James and Eric Maskin (2009). "Sequential Innovation, Patents, and Imitation". In: *The RAND Journal of Economics* 40(4), pp. 611–635. ISSN: 07416261. URL: http://www.jstor.org/stable/25593730.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch (1992). "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades". In: *Journal of Political Economy* 100(5), pp. 992–1026. ISSN: 00223808, 1537534X. URL: http://www.jstor.org/stable/2138632.

Board, Oliver (2009). "Competition and disclosure". In: *The Journal of Industrial Economics* 57(1), pp. 197–213.

Bolton, Patrick and Christopher Harris (1999). "Strategic Experimentation". In: *Econometrica* 67(2), pp. 349–374. ISSN: 1468-0262. DOI: 10.1111/1468-0262.00022. URL: http://dx.doi.org/10.1111/1468-0262.00022.

Compiani, Giovanni, Philip Haile, and Marcelo Sant'Anna (2020). "Common values, unobserved heterogeneity, and endogenous entry in US offshore oil lease auctions". In: *Journal of Political Economy* 128(10), pp. 3872–3912.

Conley, Timothy G. and Christopher R. Udry (Mar. 2010). "Learning about a New Technology: Pineapple in Ghana". In: *American Economic Review* 100(1), pp. 35–69. DOI: 10.1257/aer.100.1.35. URL: http://www.aeaweb.org/articles?id=10.1257/aer.100.1.35.

Covert, Thomas R. (2015). *Experiential and Social Learning in Firms: The Case of Hydraulic Fracturing in the Bakken Shale*. Discussion Paper. Cambridge, Massachusetts, USA.

Dafny, Leemore and David Dranove (2008). "Do Report Cards Tell Consumers Anything They Don't Already Know? The Case of Medicare HMOs". In: *The RAND Journal of Economics* 39(3), pp. 790–821. ISSN: 07416261. URL: http://www.jstor.org/stable/25474397.

Dranove, David and Ginger Zhe Jin (2010). "Quality Disclosure and Certification: Theory and Practice". In: *Journal of Economic Literature* 48(4), pp. 935–963. ISSN: 00220515. URL: http://www.jstor.org/stable/29779704.

Dranove, David, Daniel Kessler, et al. (2003). "Is More Information Better? The Effects of "Report Cards" on Health Care Providers". In: *Journal of Political Economy* 111(3), pp. 555–588. ISSN: 00223808, 1537534X. URL: http://www.jstor.org/stable/10.1086/374180.

Dranove, David and Andrew Sfekas (2008). "Start spreading the news: a structural estimate of the effects of New York hospital report cards". In: *Journal of health economics* 27(5), pp. 1201–1207.

Fetter, T Robert et al. (2018). *Learning by viewing? Social learning, regulatory disclosure, and firm productivity in shale gas*. Tech. rep. National Bureau of Economic Research.

Fishman, Michael J and Kathleen M Hagerty (2003). "Mandatory versus voluntary disclosure in markets with informed and uninformed customers". In: *Journal of Law, Economics, and organization* 19(1), pp. 45–63.

Foster, Andrew D. and Mark R. Rosenzweig (1995). "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture". In: *Journal of Political Economy* 103(6), pp. 1176–1209. ISSN: 00223808, 1537534X. URL: http://www.jstor.org/stable/2138708.

Grossman, S. J. and O. D. Hart (1980). "Disclosure Laws and Takeover Bids". In: *The Journal of Finance* 35(2), pp. 323–334. ISSN: 00221082, 15406261. URL: http://www.jstor.org/stable/2327390.

Grossman, Sanford J. (1981). "The Informational Role of Warranties and Private Disclosure about Product Quality". In: *The Journal of Law & Economics* 24(3), pp. 461–483. ISSN: 00222186, 15375285. URL: http://www.jstor.org/stable/725273.

Guo, Tong, S Sriram, and Puneet Manchanda (2021). "The effect of information disclosure on industry payments to physicians". In: *Journal of Marketing Research* 58(1), pp. 115–140.

Hastings, Justine S and Jeffrey M Weinstein (2008). "Information, school choice, and academic achievement: Evidence from two experiments". In: *The Quarterly journal of economics* 123(4), pp. 1373–1414.

Hendricks, Kenneth and Dan Kovenock (1989). "Asymmetric Information, Information Externalities, and Efficiency: The Case of Oil Exploration". In: *RAND Journal of Economics* 20(2), pp. 164–182. URL: https://ideas.repec.org/a/rje/randje/v20y1989isummerp164-182.html.

Hendricks, Kenneth, Joris Pinkse, and Robert H Porter (2003). "Empirical implications of equilibrium bidding in first-price, symmetric, common value auctions". In: *The Review of Economic Studies* 70(1), pp. 115–145.

Hendricks, Kenneth and Robert H. Porter (1988). "An Empirical Study of an Auction with Asymmetric Information". In: *The American Economic Review* 78(5), pp. 865–883. ISSN: 00028282. URL: http://www.jstor.org/stable/1807154.

Hendricks, Kenneth and Robert H. Porter (1996). "The Timing and Incidence of Exploratory Drilling on Offshore Wildcat Tracts". In: *The American Economic Review* 86(3), pp. 388–407. ISSN: 00028282. URL: http://www.jstor.org/stable/2118203.

Hendricks, Kenneth, Robert H. Porter, and Bryan Boudreau (1987). "Information, Returns, and Bidding Behavior in Ocs Auctions: 1954-1969". In: *The Journal of Industrial Economics* 35(4), pp. 517–542. ISSN: 00221821, 14676451. URL: http://www.jstor.org/stable/2098586.

Hendricks, Kenneth, Robert H. Porter, and Charles A. Wilson (1994). "Auctions for Oil and Gas Leases with an Informed Bidder and a Random Reservation Price". In: *Econometrica* 62(6), pp. 1415–1444. ISSN: 00129682, 14680262. URL: http://www.jstor.org/stable/2951754.

Herrnstadt, Evan M, Ryan Kellogg, and Eric Lewis (2020). *The economics of time-limited development options: The case of oil and gas leases*. Tech. rep. National Bureau of Economic Research.

Hodgson, Charles (2019). *Information externalities, free riding, and optimal exploration in the uk oil industry*. SIEPR, Stanford Institute for Economic Policy Research.

Ippolito, Pauline M. and Alan D. Mathios (1990). "Information, Advertising and Health Choices: A Study of the Cereal Market". In: *The RAND Journal of Economics* 21(3), pp. 459–480. ISSN: 07416261. URL: http://www.jstor.org/stable/2555620.

Irwin, Douglas A. and Peter J. Klenow (1994). "Learning-by-Doing Spillovers in the Semiconductor Industry". In: *Journal of Political Economy* 102(6), pp. 1200–1227. ISSN: 00223808, 1537534X. URL: http://www.jstor.org/stable/2138784.

Jin, Ginger Zhe (2005). "Competition and Disclosure Incentives: An Empirical Study of HMOs". In: *The RAND Journal of Economics* 36(1), pp. 93–112. ISSN: 07416261. URL: http://www.jstor.org/stable/1593756.

Jin, Ginger Zhe and Phillip Leslie (2003). "The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards". In: *The Quarterly Journal of Economics* 118(2), pp. 409–451. ISSN: 00335533, 15314650. URL: http://www.jstor.org/stable/25053911.

Jin, Ginger Zhe and Alan T Sorensen (2006). "Information and consumer choice: the value of publicized health plan ratings". In: *Journal of health economics* 25(2), pp. 248–275.

Jovanovic, Boyan (1982). "Truthful Disclosure of Information". In: *The Bell Journal of Economics* 13(1), pp. 36–44. ISSN: 0361915X. URL: http://www.jstor.org/stable/3003428.

Kellogg, Ryan (2014). "The Effect of Uncertainty on Investment: Evidence from Texas Oil Drilling". In: *The American Economic Review* 104(6), pp. 1698–1734. ISSN: 00028282. URL: http://www.jstor.org/stable/42920863.

Luco, Fernando (2019). "Who benefits from information disclosure? the case of retail gasoline". In: *American Economic Journal: Microeconomics* 11(2), pp. 277–305.

Marshall, Martin N et al. (2000). "The public release of performance data: what do we expect to gain? A review of the evidence". In: *Jama* 283(14), pp. 1866–1874.

Mathios, Alan D. (2000). "The Impact of Mandatory Disclosure Laws on Product Choices: An Analysis of the Salad Dressing Market". In: *The Journal of Law and Economics* 43(2), pp. 651–678. ISSN: 00222186, 15375285. URL: http://www.jstor.org/stable/10.1086/467468.

Matthews, Steven and Andrew Postlewaite (1985). "Quality testing and disclosure". In: *The RAND Journal of Economics*, pp. 328–340.

Milgrom, Paul R. (1981). "Good News and Bad News: Representation Theorems and Applications". In: *The Bell Journal of Economics* 12(2), pp. 380–391. ISSN: 0361915X. URL: http://www.jstor.org/stable/3003562.

Munshi, Kaivan (2004). "Social learning in a heterogeneous population: technology diffusion in the Indian Green Revolution". In: *Journal of Development Economics* 73(1), pp. 185–213. ISSN: 0304-3878. DOI: http://doi.org/10.1016/j.jdeveco.2003.03.003. URL: http://www.sciencedirect.com/science/article/pii/S0304387803001342.

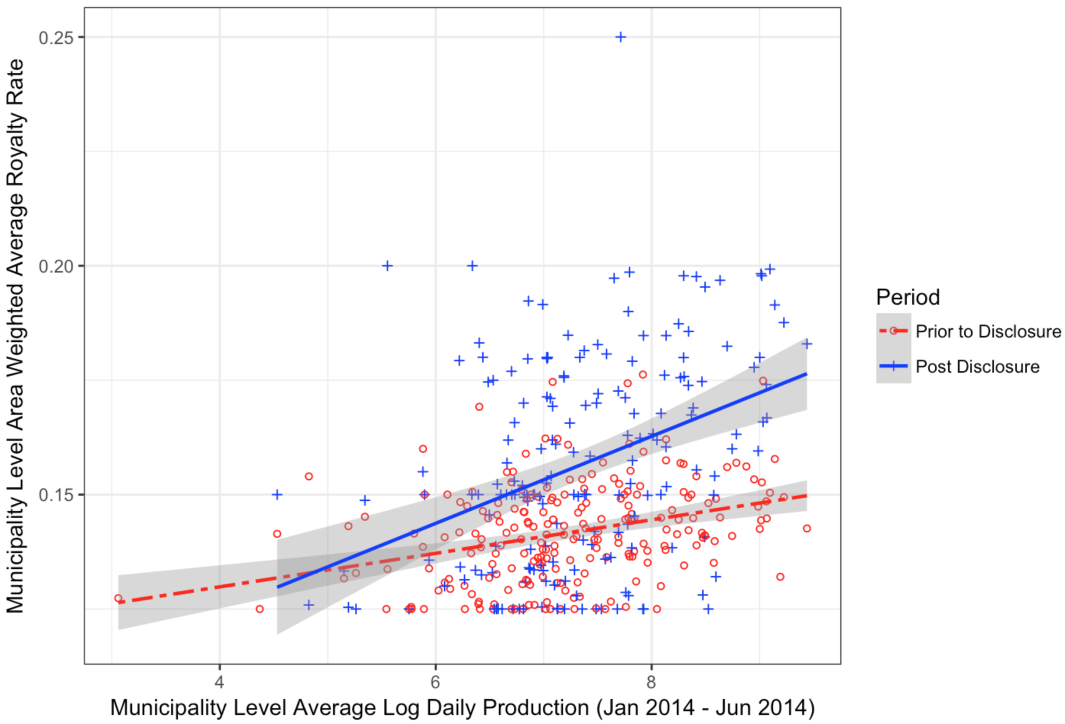
Myerson, Roger B and Mark A Satterthwaite (1983). "Efficient mechanisms for bilateral trading". In: *Journal of economic theory* 29(2), pp. 265–281.

Porter, Robert H. (1995). "The Role of Information in U.S. Offshore Oil and Gas Lease Auction". In: *Econometrica* 63(1), pp. 1–27. ISSN: 00129682, 14680262. URL: http://www.jstor.org/stable/2951695.

Shavell, Steven (1994). "Acquisition and disclosure of information prior to sale". In: *The RAND Journal of Economics*, pp. 20–36.

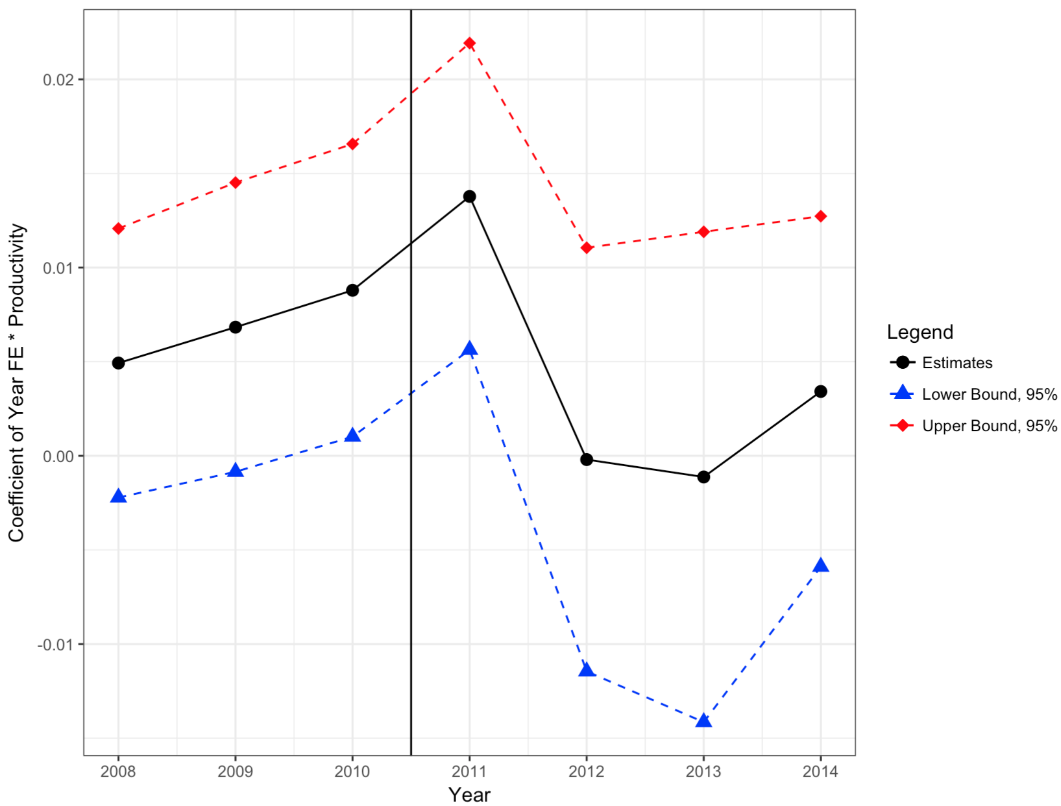
Steck, Andrew L (2018). "Industry dynamics with social learning: Evidence from hydraulic fracturing". In: *Department of Economics*.

Figure 1: Plot of Royalty Rate against Productivity Prior and Post to The Production Information Disclosure



Note: the figure plots the acreage weighted average royalty rate against the average log daily gas production for each municipality. Productivity is in thousand cubic feet per day and is calculated using the production report from January 2014 to June 2014. The royalty rates of the red dots are the weighted averages of the leases taken before the disclosure, and the blue dots are the averages after the disclosure.

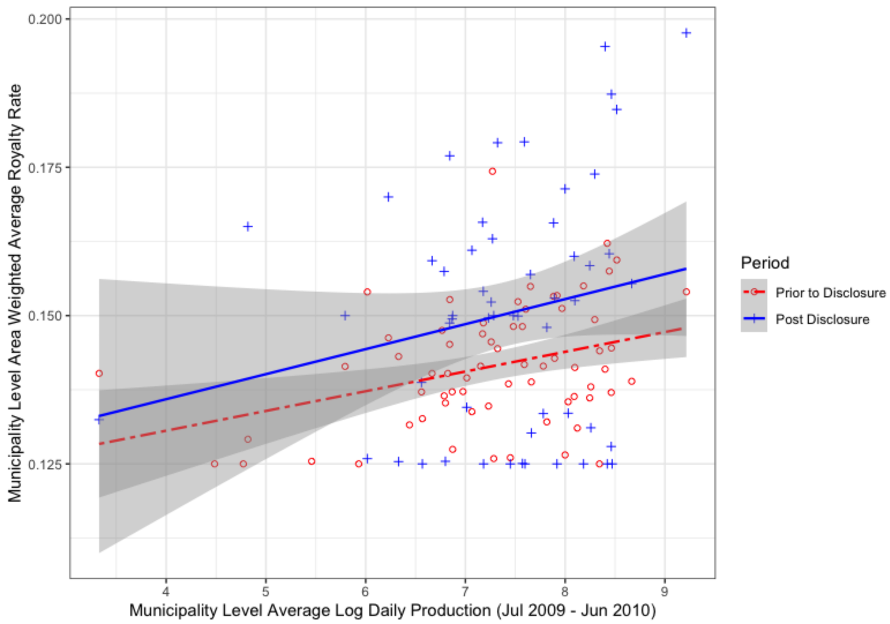
Figure 2: Plot of The Yearly Coefficient Estimates of Productivity in Model (3)



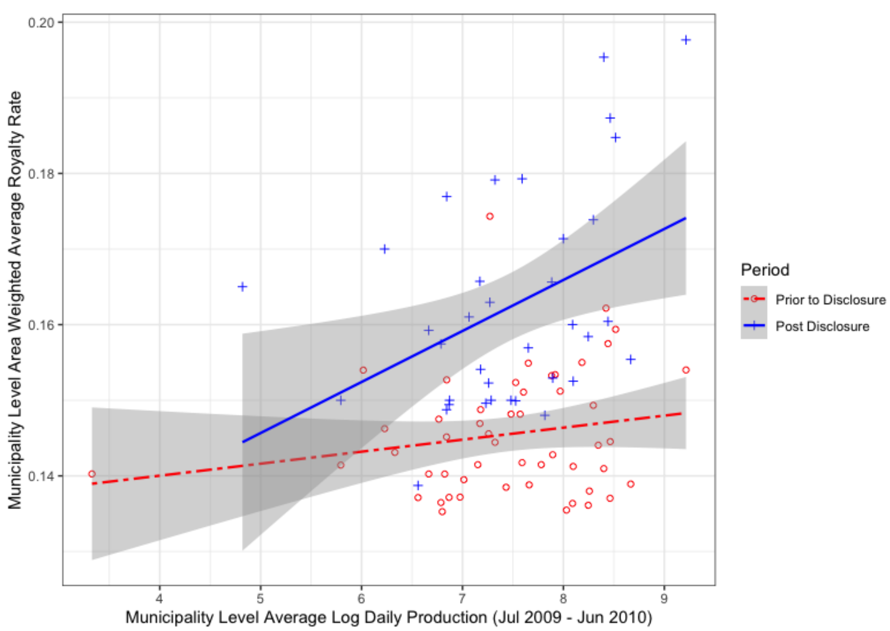
Note: the solid line in the plot connects the coefficient estimates of the interacting variables between productivity and the year fixed effects. The values of the round dots represent changes in the estimates of productivity's coefficients in the corresponding years relative to productivity's coefficient estimate in 2007. The dashed curves represent the upper and lower bounds of the 95% confidence intervals.

Figure 3: Plot of Royalty Rate against Productivity Prior and Post to The Production Information Disclosure (Using The 1st Production Report)

(a) Full Sample



(b) Royalty Rate ≥ 0.135



Note: the figures plot the acreage weighted average royalty rate against the average log daily gas production for each municipality. Productivity is in thousand cubic feet per day and is calculated using the production report from July 2009 to June 2010. The royalty rates of the red dots are the weighted averages of the leases taken before the disclosure, and the blue dots are the averages after the disclosure. The upper figure is the plot of the full sample. The lower figure is the plot of the subsample of leases where royalty rates are higher than or equal to 0.135.

Table 1: No. of Unconventional Permits Issued between Jan 2007 and Dec 2014

Operator	No. of Uncon. Permits	Accumulated Percentage
Chesapeake	3062	16.72%
Range Resources	1772	26.40%
SWEPI	1317	33.58%
EQT Production	1180	40.03%
Talisman	1116	46.12%
Operators with 100 - 999 Uncon. Permits (22)	8605	93.10%
Operators with 10 - 99 Uncon. Permits (31)	1105	99.13%
Total No. of Uncon. Permits	18316	100.00%

Note: The permit data is from PADEP. There are 390 operators receiving at least one drilling permit from Jan 2007 to Dec 2014, among which 107 received at least one unconventional permit. The percentage in the third column is taken with respect to the total number of unconventional permits.

Table 2: Summary Statistics of The Selected Lease Sample

Year	No. of Leases	Ave. Acres	Ave. Term (M)	Ave. Royalty
2007	5524	70.03	66.07	12.56%
2008	20700	37.39	63.86	14.26%
2009	14167	25.17	60.35	15.53%
2010	17358	22.16	60.57	16.31%
2011	9355	22.97	60.54	15.17%
2012	4668	23.39	62.44	16.73%
2013	7216	29.61	59.50	14.73%
2014	11253	17.68	64.16	14.43%

Note: The lease data is from Enverus. Operators receiving fewer than 10 unconventional permits between January 2007 and December 2014 are dropped. Duplicated observations due to multiple grantors and subplots are eliminated using the algorithm described in the appendix. For leases with both the original contracts and the lease memos, only the original contracts are kept. The average royalty rates in the fifth column are taken over the observed values given by the original contracts and are weighted by acreages.

Table 3: Release Time of The Production Reports

Report No.	Release Time	Period Covered
1	Nov 2010	07/2009 – 06/2010
2	Feb 2011	07/2010 – 12/2010
3	Aug 2011	01/2011 – 06/2011
4	Feb 2012	07/2011 – 12/2011
5	Aug 2012	01/2012 – 06/2012
6	Feb 2013	07/2012 – 12/2012
7	Aug 2013	01/2013 – 06/2013
8	Feb 2014	07/2013 – 12/2013
9	Aug 2014	01/2014 – 06/2014

Note: The contents of the table are based on the production reports from PADEP. The first report covers the production of twelve months from July 2009 to June 2010. Each other report covers six months of production. Each reports is assumed to be released on the first day of the month.

Table 4: Summary Statistics of The Production Data

Report	No. of Active Wells		Mean Producing Days		Mean Quant./Day (MMcf.)		S.D. of Quant./Day	
	Horizonal	Vertical	Horizontal	Vertical	Horizontal	Vertical	Horizontal	Vertical
1	438	411	181.55	304.17	2.74	0.21	2.17	0.33
2	768	451	133.18	167.83	2.81	0.15	2.33	0.23
3	1180	463	139.22	166.18	2.79	0.12	2.49	0.17
4	1750	507	140.36	159.83	2.72	0.14	2.26	0.68
5	2376	506	151.05	166.44	2.64	0.14	2.40	0.53
6	3036	520	158.73	155.89	2.56	0.16	2.53	0.53
7	3694	521	157.37	165.12	2.71	0.18	2.91	0.71
8	4369	565	161.26	153.19	2.67	0.27	3.18	1.22
9	4862	541	160.76	164.64	2.69	0.28	2.96	1.08

Note: the production data is from PADEP. The maximum number of producing days is 365 for the first report, and 184 (or 183) for the other reports. The unit of the daily production quantity is million cubic feet (MMcf.)

Table 5: Estimation Results of Model(2)

	Coef.	S.E.
Const.	0.1152^{**}	(0.0082)
$\log(Prod_j^9)$	0.0037^{**}	(0.0011)
D_t	-0.0285^{**}	(0.0129)
$D_t \times \log(Prod_j^9)$	0.0058^{**}	(0.0017)

Significance codes: 0.05 '**', 0.1 '*'

Note: $\log(Prod_j^9)$ represents the average log daily production (Mcf./d) of municipality j calculated using the 9th production report. D_t is the disclosure indicator.

Table 6: Estimation Results of Model(4)

	Coef.	SE.
$\log(Prod_j^9)$	0.0047^{**}	(0.0011)
$S_j \times \log(Prod_j^9)$	-1.37×10^{-5}	(0.0004)
$D_t \times \log(Prod_j^9)$	0.0009	(0.0006)
$(S_j \times D_t) \times \log(Prod_j^9)$	0.0014^{**}	(0.0006)

Significance codes: 0.05 '**', 0.1 '*'

Note: $P\log(Prod_j^9)$ is the average log daily production (Mcf./d) of municipality j calculated using the 9th production report. D_t is the information disclosure indicator variable. S_j is the municipality type indicator variable that takes value 1 if municipality j is classified as shale gas oriented.

Table 7: Estimation Results of Model(2) Using The 1st Production Report

	Full Sample		Royalty Rate ≥ 0.135	
	Coef.	S.E.	Coef.	S.E.
	(1)	(2)	(3)	(4)
Const.	0.1173**	(0.0128)	0.1337**	(0.0116)
$\log(Prod_j^1)$	0.0033*	(0.0017)	0.0016	(0.0015)
D_t	0.0017	(0.0202)	0.0217	(0.0191)
$D_t \times \log(Prod_j^1)$	0.0009	(0.0027)	0.0052**	(0.0025)

Significance codes: 0.05 '**', 0.1 '*'

Note: $Prod_j^1$ represents the average log daily production (Mcf./d) of municipality j calculated using the 1st production report. D_t is the information disclosure indicator variable.

Table 8: Estimation Results of Model (5)

Dep. Var.	Royalty Rate	Current Period's Acreage			
	(1)	(2)	(3)	(4)	(5)
Own Recent Acreage	0.0011**	0.4295**	0.4330**	0.4283**	0.4270**
	(0.0000)	(0.0091)	(0.0092)	(0.0091)	(0.0090)
Peer's Recent Acreage	0.0018**	0.0052**	0.0049**	0.00346*	0.0012**
	(0.0000)	(0.0023)	(0.0023)	(0.0023)	(0.0003)
Productivity	0.0011**	0.0035**	0.0035**	0.0024**	0.0011**
	(0.0000)	(0.0012)	(0.0012)	(0.0012)	(0.0002)
Productivity × Own Recent Acreage	−0.0001**	0.0011	0.0009	0.0012	0.0013
	(0.0000)	(0.0023)	(0.0024)	(0.0023)	(0.0024)
Productivity × Peer's Recent Acreage	−0.0001**	−0.0005**	−0.0005**	-0.0004	-0.0002
	(0.0000)	(0.0002)	(0.0002)	(0.0002)	(0.0001)
Income	−4.52e-07**				
	(9.67e-09)				
Royalty Rate		−2.2304*	−2.2382*	-1.5990	0.0516
		(1.2891)	(1.2917)	(1.02637)	(0.0346)
Natural Gas Price	−0.0048**	0.0033	0.0017	0.0023	0.0118**
	(0.0001)	(0.0054)	(0.0056)	(0.0024)	(0.0026)
Public	−0.0000	0.0043*	0.0101**	0.0067**	0.0044*
	(0.0002)	(0.0022)	(0.0011)	(0.0026)	(0.0022)
Population	−0.0009**	-0.0018	-0.0018	-0.0009	0.0015**
	(0.0000)	(0.0025)	(0.0025)	(0.0021)	(0.0006)
Land Area	−0.0006**	0.0026**	0.0026**	0.0029**	0.0034**
	(0.0000)	(0.0011)	(0.0011)	(0.0009)	(0.0003)
Operator FE	Y	Y	N	Y	Y
Time FE	Y	Y	Y	N	Y
Instrument for Royalty Rate	NA	Y	Y	Y	N
No. of Observations	1,056,319	1,056,319	1,056,319	1,056,319	1,056,319
R^2	0.0677	0.3597	0.3586	0.3659	0.3752
Prob>F	0.0000	0.0000	0.0000	0.0000	0.0000

Significance codes: 0.05 '**', 0.1 '*'.

Note: Current Period's Acreage is the log of the lease acreage operator i obtains in municipality j in the current month. Own Recent Acreage is the log of the lease acreage operator i obtained in municipality j in the past three months. Peer's Recent Acreage is the log of the lease acreage operators other than i obtained in municipality j in the past three months. Productivity is the average log daily production of the horizontal wells in the neighboring municipalities, calculated using the latest production report available in month t. The value of Productivity is zero when $t < 47$. Population and Land Area are in log forms.

Column 1 is the first stage result of the 2SLS estimation. The dependent variable is Royalty Rate. Columns 2-5 are the 2SLS estimation results.

Table 9: Estimation Results of Model (5) under Various Constructions of The Signal Vector

Components of Signal Vector	Past 3 Months	Past 6-4 Months	Cumulative
Dep. Var.: Current Period's Acreage	(1)	(2)	(3)
Own Recent Acreage	0.4295**	0.3325**	
	(0.0091)	(0.0102)	
Peer's Recent Acreage	0.0052**	0.0067**	
	(0.0023)	(0.0030)	
Own Cumulative Acreage			0.1102**
			(0.0117)
Peer's Cumulative Acreage			0.0202
			(0.0156)
Productivity	0.0035**	0.0042**	0.0188**
	(0.0012)	(0.0015)	(0.0079)
Productivity × Own Recent Acreage	0.0011	0.0075**	
	(0.0023)	(0.0025)	
Productivity × Peer's Recent Acreage	−0.0005**	−0.0006**	
	(0.0002)	(0.0003)	
Productivity × Own Cumulative Acreage			−0.0026**
			(0.0009)
Productivity × Peer's Cumulative Acreage			-0.0007
			(0.0011)
Royalty Rate	−2.2304*	−2.7289*	-12.3858
	(1.2891)	(1.6330)	(11.0149)
Natural Gas Price	0.0033	0.0026	0.0257**
	(0.0054)	(0.0067)	(0.0051)
Public	0.0043*	0.0069**	0.0390**
	(0.0022)	(0.0030)	(0.0051)
Population	-0.0018	-0.0026	-0.0087
	(0.0025)	(0.0031)	(0.0148)
Land Area	0.0026**	0.0046**	-0.0102
	(0.0011)	(0.0014)	(0.0138)
Operator FE	Y	Y	Y
Time FE	Y	Y	Y
Instrument for Royalty Rate	Y	Y	Y
No. of Observations	1,056,319	1,056,358	1,057,749

Significance codes: 0.05 '**', 0.1 '*'.

Note: Current Period's Acreage is the log of the lease acreage operator i obtains in municipality j in the current month. Own Recent Acreage is the log of the lease acreage operator i obtained in municipality j in the past three months or the three-month period before the past three months. Peer's Recent Acreage is the log of the lease acreage operators other than i obtained in municipality j in the past three months or the three-month period before the past three months. Own Cumulative Acreage and Peer's Cumulative Acreage are the logs of the lease acreage operator i and operators other than i obtained in municipality j up to the past three months. Productivity is the average log daily production of the horizontal wells in the neighboring municipalities, calculated using the latest production report available in month t. The value of Productivity is zero when $t < 47$. Population and Land Area are in log forms.

Column 1 is the benchmark result. Column 2 is the estimation result using lease acreages in the three-month period before the past three months as the signals. Column 3 is the estimation result using cumulative lease acreages up to the past three months as the signals

Table 10: Estimation Results of Model (5) for Varying Trimmed Samples

Sample Period	2007-2014	2008-2013	2009-2012	2010-2011
Dep. Var.: Current Period's Acreage	(1)	(2)	(3)	(4)
Own Recent Acreage	0.4295**	0.4286**	0.4327**	0.4261**
	(0.0091)	(0.0089)	(0.0113)	(0.0132)
Peer's Recent Acreage	0.0052**	0.0050**	0.0041**	0.0055**
	(0.0023)	(0.0019)	(0.0015)	(0.0016)
Productivity	0.0035**	0.0033**	0.0027**	0.0020**
	(0.0012)	(0.0009)	(0.0006)	(0.0006)
Productivity × Own Recent Acreage	0.0011	-0.0012	-0.0052	-0.0031
	(0.0023)	(0.0025)	(0.0029)	(0.0036)
Productivity × Peer's Recent Acreage	−0.0005**	−0.0004*	-0.0003	-0.0003
	(0.0002)	(0.0002)	(0.0002)	(0.0003)
Royalty Rate	−2.2304*	−1.9414**	−1.5205**	-1.7300
	(1.2891)	(0.9409)	(0.5310)	(0.4702)
Natural Gas Price	0.0033	0.0076**	−0.0118**	0.0068**
	(0.0054)	(0.0027)	(0.0037)	(0.0019)
Public	0.0043*	−0.0085**	0.0005	-0.0071
	(0.0022)	(0.0027)	(0.0026)	(0.0064)
Population	-0.0018	-0.0016	-0.0017	−0.0022
	(0.0025)	(0.0020)	(0.0015)	(0.0015)
Land Area	0.0026**	0.0025**	0.0027**	0.0026**
	(0.0011)	(0.0010)	(0.0008)	(0.0010)
Operator FE	Y	Y	Y	Y
Time FE	Y	Y	Y	Y
Instrument for Royalty Rate	Y	Y	Y	Y
No. of Observations	1,056,319	845,870	566,542	299,222

Significance codes: 0.05 '**', 0.1 '*'.

Note: Current Period's Acreage is the log of the lease acreage operator i obtains in municipality j in the current month. Own Recent Acreage is the log of the lease acreage operator i obtained in municipality j in the past three months. Peer's Recent Acreage is the log of the lease acreage operators other than i obtained in municipality j in the past three months. Productivity is the average log daily production of the horizontal wells in the neighboring municipalities, calculated using the latest production report available in month t. The value of Productivity is zero when $t < 47$. Population and Land Area are in log forms.

Column 1 is the benchmark result using the full sample. Columns 2-4 are the estimation results using trimmed subsamples where the first and last years of observations are removed.

Table 11: Estimation Results of Model (5) For Pre- and Post-Disclosure Subsamples

Sample Period	2007-2014	2011-2014	2007-2010	2007-2010
Report to Calculate Productivity	Recent Report	Recent Report	Report 1	Report 9
Dep. Var.: Current Period's Acreage	(1)	(2)	(3)	(4)
Own Recent Acreage	0.4295**	0.4262**	0.4032**	0.3042**
	(0.0091)	(0.0185)	(0.0110)	(0.0173)
Peer's Recent Acreage	0.0052**	0.0096*	−0.0009	−0.0005
	(0.0023)	(0.0052)	(0.0019)	(0.0010)
Productivity	0.0035**	0.0051**	0.0005	0.0008
	(0.0012)	(0.0021)	(0.0006)	(0.0006)
Productivity × Own Recent Acreage	0.0011	0.0017	0.0096**	0.0210**
	(0.0023)	(0.0030)	(0.0026)	(0.0026)
Productivity × Peer's Recent Acreage	−0.0005**	−0.0010*	-0.0002	-0.0004
	(0.0002)	(0.0006)	(0.0002)	(0.0002)
Royalty Rate	−2.2304*	−3.5801*	1.5468	1.8594*
	(1.2891)	(2.1465)	(1.1159)	(1.0777)
Natural Gas Price	0.0033	0.0013	0.0168**	0.0185**
	(0.0054)	(0.0012)	(0.0064)	(0.0062)
Public	0.0043*	0.0064**	−0.0087**	-0.0058
	(0.0022)	(0.0032)	(0.0040)	(0.0043)
Population	-0.0018	-0.0035	0.0008	0.0016
	(0.0025)	(0.0056)	(0.0014)	(0.0013)
Land Area	0.0026**	0.0015**	0.0066**	0.0072**
	(0.0011)	(0.0018)	(0.0010)	(0.0010)
Operator FE	Y	Y	Y	Y
Time FE	Y	N	Y	Y
Instrument for Royalty Rate	Y	Y	Y	Y
No. of Observations	1,056,319	614,907	441,412	441,412

Significance codes: 0.05 '**', 0.1 '*'.

Note: Current Period's Acreage is the log of the lease acreage operator i obtains in municipality j in the current month. Own Recent Acreage is the log of the lease acreage operator i obtained in municipality j in the past three months. Peer's Recent Acreage is the log of the lease acreage operators other than i obtained in municipality j in the past three months. Productivity is the average log daily production of the horizontal wells in the neighboring municipalities, calculated using the latest production report available in month t. The value of Productivity is zero when $t < 47$. Population and Land Area are in log forms.

Column 1 is the benchmark result using the full sample. Column 2 is the estimation result using the post-disclosure subsample. Both Column 1 and Column 2 use productivity calculated from the most recent production report. Columns 3-4 are the estimation results using the pre-disclosure subsample. Column 3 uses productivity calculated from the 1st production report, and Column 4 uses productivity calculated from the 9th production report.

Appendix

This appendix provides further details on the simulation analysis of the multi-firm case in the theoretical model and offers a deeper inspection of different components of the dataset used for the empirical analysis.

The Multiple-Firm Case

When there are multiple firms, analytical deduction becomes less tractable. Instead, in this appendix a simulation-based approach is adopted to explore the quantitative relationship between the additional information and firms' location choices. A summary of the simulation's settings is given in Table 12.

The simulation is run 10,000 times under each of the three assumptions of the royalty rate functional forms. In each round, 50 firms sequentially choose between two locations L_1 and L_0 with $Q_1 = 1$ and $Q_0 = 0$. The signal s_i for each firm i is an independent binomial draw following the distribution $\Pr(s_i = 1) = 1 - \Pr(s_i = 0) = p = 3/5$, and all firms have the common belief that $\Pr(s_i = j|Q_j = 1) = p$, $j = 0, 1$. The firms conduct Bayesian updating based on the location choice histories and private signals they observe and choose either L_1 or L_0 to solve the profit-maximizing problem (1). The number of firms choosing each location is recorded.

Such an exercise is done for a range of common prior beliefs of $\Pr(Q_1 = 1)$. The average number of firms ended at each location and the resulting average royalty rates are calculated for each prior belief. The relationship between the quantities and the prior beliefs is shown in Figure 4. Figure 4a and Figure 4b plot the average numbers of firms and royalty rates at L_1 and L_0 against different values of prior beliefs under the setting of a linear royalty rate function $r(x) = x/N$. Figure 4c and Figure 4d are the same plots for a concave royalty function $r(x) = 1 - 1/(x + 1)$, and Figure 4e and Figure 4f are based on a convex cost function $r(x) = x^2/N^2$. In all three settings, the average numbers of firms choosing L_1 increase in the prior beliefs, and so do the average royalty rates. The results of the simulation show that when the disclosed productivity information correctly adjusts the operators' prior beliefs, the investment efficiency of the whole industry improves.

The simulation results also generate reduced-form implications that can be taken on to data. One implication is that the correlation between the private signal and the operator's location choice changes along with the common prior belief. Let Y_i be firm i's location choice and X_i be firm i's signal. The following probit model depicts the firm's location choice as a function of its signal:

$$Y_i = \begin{cases} 1, & Y_i^* > 0 \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

and

$$Y_i^* = \beta_0 + \beta_1 X_i + \varepsilon_i, \tag{7}$$

where ε_i is an independent standard normal unobserved term. The resulting sample from the simulation exercise can be used to estimate the above model.

Figure 5 plots $\hat{\beta}_1$, the estimated coefficient of X_i in Equation (7), against prior beliefs of $\Pr(Q_1 = 1)$ ranging from 0.1 to 0.9[24], using simulated samples with a linear royalty cost function $r(x) = x/N$. The coefficient estimates are positive for all prior beliefs, suggesting that the signal observed by an operator is predictive of the operator's location choice. In addition, the coefficient of the signal variable first decreases along with the prior belief when its value is in the lower range between 0.1 and 0.2 and then increases fast when the value of the prior belief gets larger. Finally, when the prior belief's value is greater than 0.5, the coefficient's value does not change much. The finding suggests that, when the quality of the prior beliefs is very low, the disclosure of additional positive information may diminish the correlation between an operator's location choice and the signal it observes. When prior beliefs are slightly more accurate, the disclosure of additional positive information tends to amplify this correlation. Lastly, when firms are relatively well-informed, the correlation between their location choices and their private signals stays flat.

The Lease Sample

Two types of lease records show up in the sample of leases. Apart from the original lease contracts, there are also lease memos, which are simplified versions of contracts in which some information, including the royalty rate, is concealed. There could be duplicated observations of lease contracts and lease memos for a single plot of land, possibly due to different grantors or because they are generated for different subplots. For example, a 20-acre lease may contain three subplots, A, B, and C. Suppose subplot A was leased on April 1, 2009, from household h1, and its original contract is observable. Subplot B was leased on April 1, 2009, from households h1 and h2, each with a lease memo. Subplot C was leased on April 5, 2009, from household h3. Both its original contract and its lease memo are available. The three subplots generate five observations of lease records. The acreage of each individual subplot is not known. Instead, all five observations have the same acreage value, which is 20 acres.

One fact that complicates the matter is that there is not a variable to identify observations of the same plot of land. The only way to precisely determine that the lease on April 5 is for subplot C, not for a distinct parcel, is to refer to the plot map included in the contract, a method impractical to implement when dealing with a large number of observations. As a compromise, the following procedures are taken to adjust for the possible duplication.

First, observations with the same operator, location (county and municipality), date, acreage, and type (lease contract or lease memo) are treated to be of the same plot, and duplicated observations are dropped. This step leaves one original contract for subplot A, one lease memo for subplot B, and both the original contract and the lease memo for subplot C for the hypothetical example.

In the next step, for each combination of operator, location (county and municipality), date, and acreage, if there are both the original contract and the lease memo, the lease memo gets dropped. This step leaves one original contract for subplot A, one original contract for subplot C, and no

[24]For extremely small prior probabilities, possibility exists that no operator chooses location L_1, which generates a problem for the regression analysis.

lease records for subplot B.

In the third step, observations with the same operator, county, and acreage are grouped and sorted in the ascending order of their leasing dates, and the time lags between consecutive observations are calculated. For each ordered group, an observation is indicated to be of a distinct lease if one of the following requirements is satisfied: a). its area is smaller than 10 acres; b). it is the leading observation of the group; c). its area is smaller than 50 acres, and the lag between its date and the date of the previous observation in the same group is greater than 30 days; d). its area is larger than 50 acres, and the lag between its date and the date of the previous observation in the same group is greater than 60 days.

The assumptions behind the criteria are the following: if a lease is smaller than 10 acres, it is obtained with a single contract on a single day; if a lease is larger than 10 acres and smaller than 50 acres, it could be obtained with multiple contracts signed at different dates, and the lag between any two consecutive dates is less than 30 days; if a lease is larger than 50 acres, the lag between any of its two consecutive contracts is less than 60 days. Each ordered group contains the leases in the same county but not necessarily in the same municipality because the subplots of a lease may be in different municipalities. According to this rule, the contracts for subplots A and C are treated to be of the same plot of land.

Finally, the number of contracts for each lease is counted, and the average acreage of each contract is calculated. For the hypothetical example, the 20-acre plot is leased in two days, once on April 1, 2009, and the other time on April 5, 2009, each with 10 acres. The result is likely to be different from reality. Still, it correctly reflects the timing and structure of the lease. The refined sample contains 90,241 lease records.

There are 2464 municipalities in Pennsylvania, and the selected lease sample involves 675 of them. For each combination of municipality and month, denoted as a *location-time pair*, the acreage-weighted average royalty rate of the leases granted in that pair is used as the expected cost of obtaining leases. To accommodate for the location-time pairs in which there are no lease observations, the spatial correlation of royalty rates is exploited. A common measure for the spatial autocorrelation of a variable is its Moran's I, which is defined as

$$I = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij}(X_i - \overline{X})(X_j - \overline{X})}{\sum_i (X_i - \overline{X})^2}. \tag{8}$$

In the above definition, N is the number of spatial units, X is the variable of interest, and w_{ij} is the element in the ith row and jth column of the spatial weighting matrix. w_{ij} takes value 1 if unit i and unit j are "neighbors", and 0 if they are not. Given the spatial weighting matrix, a positive Moran's I implies that high values of X tend to cluster with other high values, and low values of X tend to cluster with other low values. A negative Moran's I implies that high values of X tend to cluster with low values of X. If the values of X are randomly distributed, the Moran's I will be close to 0.

In the royalty rate example, N is the number of municipalities in which royalty rates are observed.

X is the acreage-weighted average royalty rate of a municipality. For each given distance d, $w_{ij} = 1$ if $i \neq j$ and the distance between municipalities i and j is less than d, and $w_{ij} = 0$ otherwise. Figure 6 displays the Moran's I for the municipality-level average royalty rates based on different neighborhood-defining distances, using the subsample of the original contracts. The figure shows that Moran's I for the municipality-level average royalty rate is significantly positive when the neighborhood defining distance is within the range of 175 miles, suggesting that royalty rates are highly spatially correlated. The finding inspires the following approximation algorithm.

Let L denote the set of the 675 municipalities in which lease records are observed. To get an approximation of the expected royalty rate $R_{l,m}$ for each location-time pair (l, m), the following looping algorithm is run over a list of neighborhood-defining distances and averaging periods:

- Initialize $R_{l,m}$ as a missing value, $\forall l \in L$ and $m = 1, 2, \ldots, 96$.
- Create the list of neighborhood defining distances $D = \{0, 5, 10, 15, 20, 25, 30, 40, 50, 60, 70\}$ with their units in miles. Create the list of averaging periods $T = \{1, 3, 6\}$ with their units in months.
- Outer loop: loop through D in ascending order. For each $d \in D$:
  - Inner loop: loop through T in ascending order. For each $t \in T$:
    - For each municipality l and month m, calculate the acreage-weighted average royalty rate $r_l^{m,t}$ and the sum of acreages $s_l^{m,t}$ using the leases made in municipality l within the time interval $[m - t, m + t]$. Denote $r_l^{m,t}$ as a missing value if $s_l^{m,t} = 0$.
    - Define d_{ij} to be the distance between municipalities i and j. For each municipality l, define $N_{l,d}$ to be the set of neighboring municipalities of l such that $n \in N_{l,d}$ if and only if $d_{nl} \leq d$.
    - Calculate the acreage-weighted average royalty rate $R(l, d, m, t)$ for each municipality l and month m, using $\{(r_n^{m,t}, s_n^{m,t}) | n \in N_{l,d}\}$. Denote $R(l, d, m, t)$ as a missing value if $s_n^{m,t} = 0$, $\forall n \in N_{l,d}$.
    - Set $R_{l,m} = R(l, d, m, t)$ if $R_{l,m}$ is a missing value and $R(l, d, m, t)$ is not.

Another concern about the lease data is that it is not clear whether a lease is granted for conventional wells or unconventional wells. Given that the geological features of conventional natural gas are well understood, additional information on shale gas productivity is unlikely to shift the locations where firms drill conventional wells. To analyze the impact of the shale gas production information disclosure on firms' land leasing decisions, it is essential to distinguish between the two types of leases.

The permit data is used to filter the lease records that are made for the purpose of unconventional well drilling. The filtering process takes the following steps: first, for each pair of operator i and county c, the numbers of conventional and unconventional permits issued to operator i in county c are counted, which are denoted as $n_{i,c}^{c}$ and $n_{i,c}^{uc}$. Their ratio $n_{i,c}^{c}/n_{i,c}^{uc}$ is calculated. In the second step, for each operator i, its numbers of conventional and unconventional permits are counted, which are denoted as n_i^c and n_i^{uc}, and their ratio n_i^c/n_i^{uc} is calculated. In the third step, for a given threshold h, which is set to be 1/5, the leases granted to operator i in county c are dropped if $n_{i,c}^c/n_{i,c}^{uc} > h$. In the fourth step, if operator i acquired no drilling permits in county c, its leases in county c are dropped if $n_i^c/n_i^{uc} > h$.

The reasoning behind the filtering process is that if $n_{i,c}^c/n_{i,c}^{uc}$ gets too large, the leases granted to operator i in county c will be contaminated by those intended for conventional well drilling, which can be misleading if used for empirical analysis. In the cases where neither $n_{i,c}^c$ nor $n_{i,c}^{uc}$ is available, the ratio n_i^c/n_i^{uc} is used as a general gauge of the weights operator i attaches to conventional and unconventional drilling. The method is conservative in the sense that it is possible for the conventional permits to be issued on the existing leases, whereas the newly acquired leases are used for shale gas production, in which case the above procedure unnecessarily drops usable observations. The defect, if exists, has at most limited impacts, given that the filtered sample contains 81207, or 89.99% of the initial observations.

The Production Reports

An essential precondition that productivity information disclosure induces positive spatial autocorrelation of royalty rates is that productivity is itself positively spatially correlated. Figure 7 plots Moran's I for the municipality-level horizontal well productivity against different values of neighborhood-defining distances, using the last production report. The average daily production of the horizontal wells in each municipality is used as the variable X in Equation 8. The value of Moran's I is significantly positive for a wide range of distances up to 200 miles, which implies that shale gas productivity, like royalty rate, also exhibits strong positive spatial correlations.

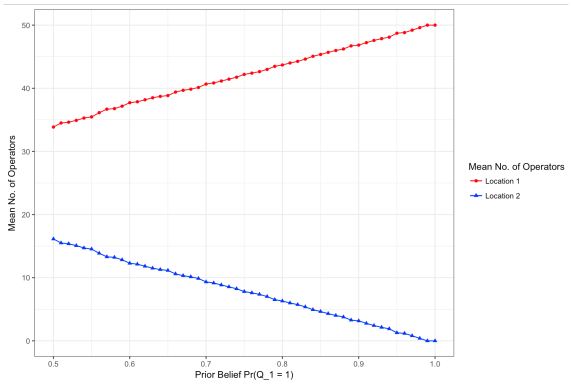
One feature of horizontal wells is that their productivity declines over time as the inner pressure of the shale gas formation gets weaker. This is the major reason that the mean productivity of the horizontal wells fluctuates over time. To get a more precise sense of how the production information disclosure may affect shale gas productivity, figure 8 plots the frequency histograms of the horizontal wells' daily production using the first, the fifth, and the ninth production reports. The daily production, measured by thousand cubic feet per day (Mcf./d), is in logarithmic form. The upper diagram of Figure 8a is the productivity histogram of the horizontal wells in report 1. The vertical axis measures the number of wells. The lower diagram of Figure 8a is the productivity histogram of the horizontal wells in report 5, where the wells in the first report are excluded. Therefore, the lower left diagram solely reflects the productivity of the horizontal wells drilled between July 2010 and December 2011. Figure 8b is the analogous comparison between the productivity histogram of the horizontal wells in report 5 and report 9, where the wells in report 5 are excluded. Table 13 reports the means and medians of the histograms in Figure 8. The mean and median log daily production of the horizontal wells in report 1 are 7.54 and 7.62. By contrast, the mean and median log daily production of the horizontal wells drilled between July 2010 and December 2011 are 7.67 and 7.70. The numbers grow further to 7.93 and 7.96 for the horizontal wells drilled between January 2012 and December 2013. The finding is consistent with the hypothesis that shale gas operators benefit from the production information disclosure by choosing locations with higher potential.

The Natural Gas Prices

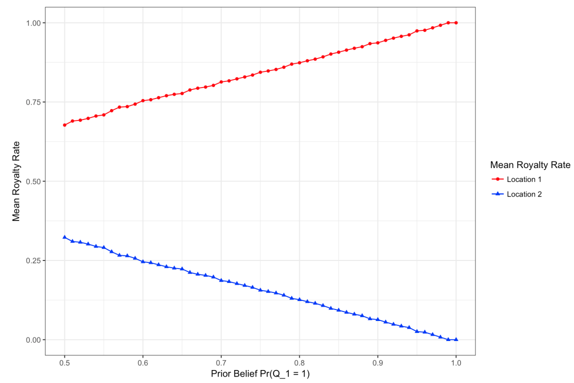
Located in Erath, Louisiana, the Henry Hub is a vital distribution hub of the natural gas pipeline system and has long been a benchmark for natural gas futures trading in the United States. Its natural gas spot price is closely correlated with North American wellhead prices. Figure 9 plots the monthly natural gas spot prices along with the logarithm of the lease acreages granted in each month in Pennsylvania from 2007 to 2014. The sample correlation between the two variables is 0.67, suggesting that natural gas price is likely to be one of the factors that affect firms' land leasing decisions.

Figure 4: Simulation Results for Different Royalty Cost Functions

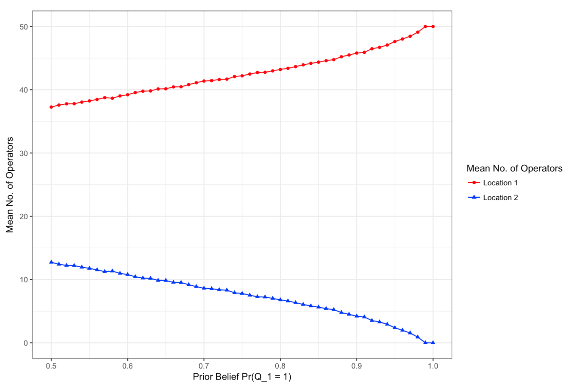
(a) Mean No. of Operators, $r(x) = x/N$



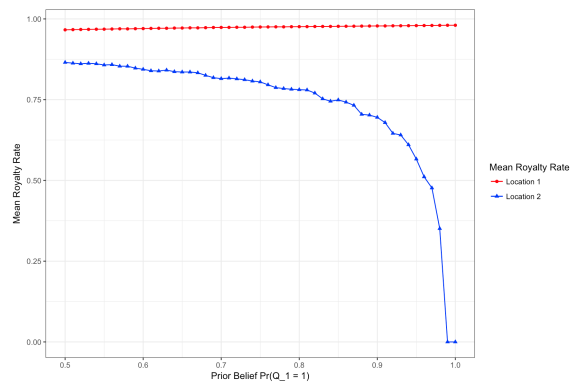
(b) Mean Royalty Rate, $r(x) = x/N$



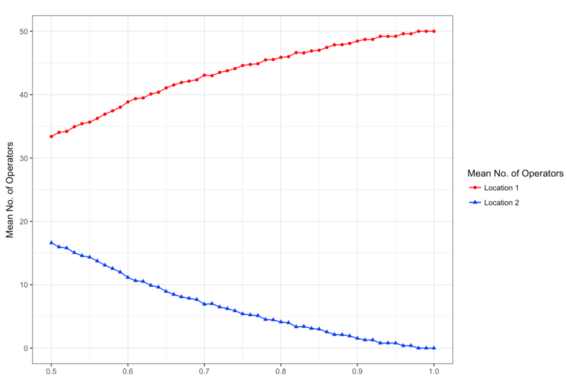
(c) Mean No. of Operators, $r(x) = 1 - 1/(x + 1)$



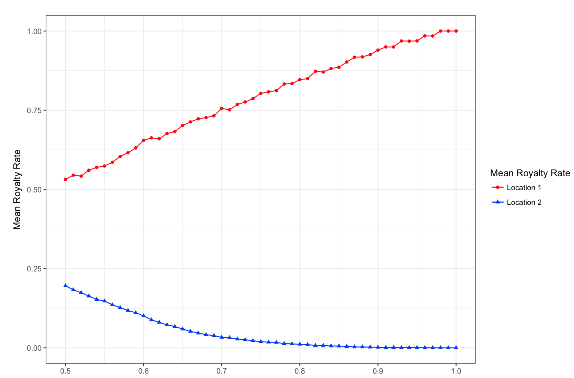
(d) Mean Royalty Rate, $r(x) = 1 - 1/(x + 1)$



(e) Mean No. of Operators, $r(x) = x^2/N^2$

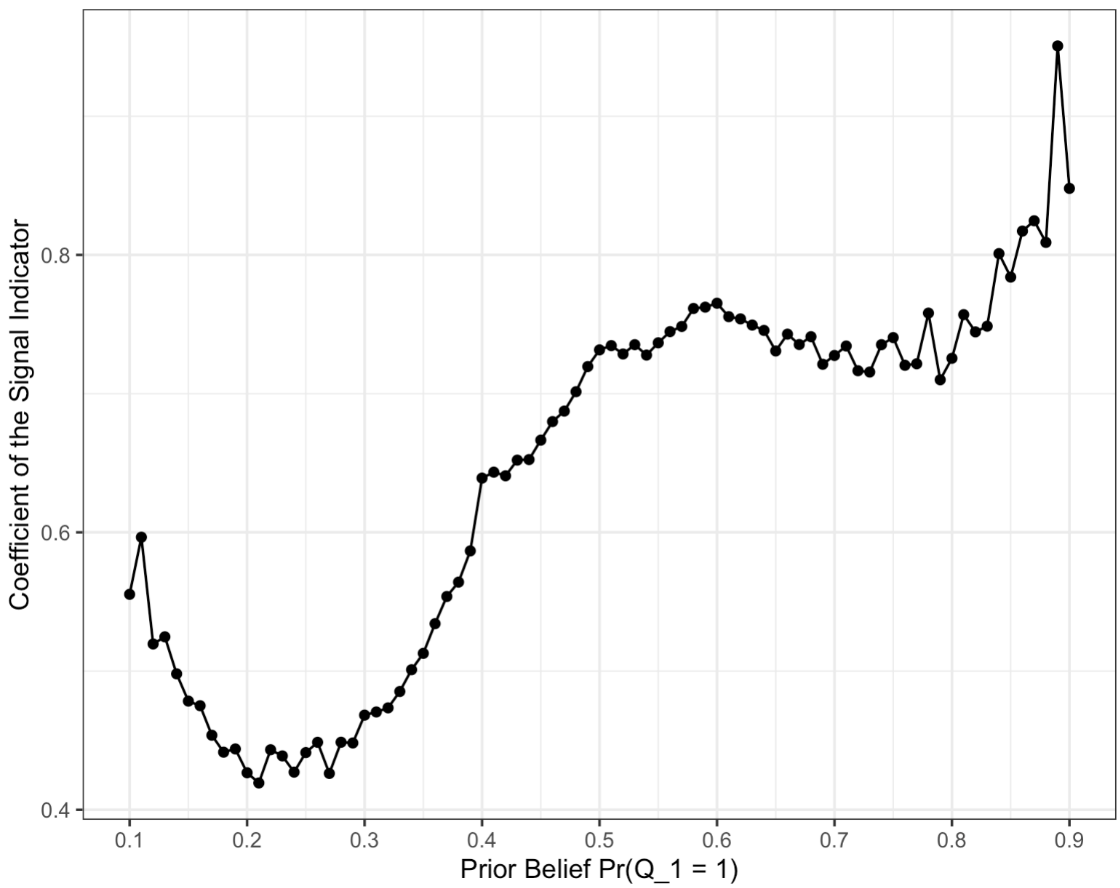


(f) Mean Royalty Rate, $r(x) = x^2/N^2$



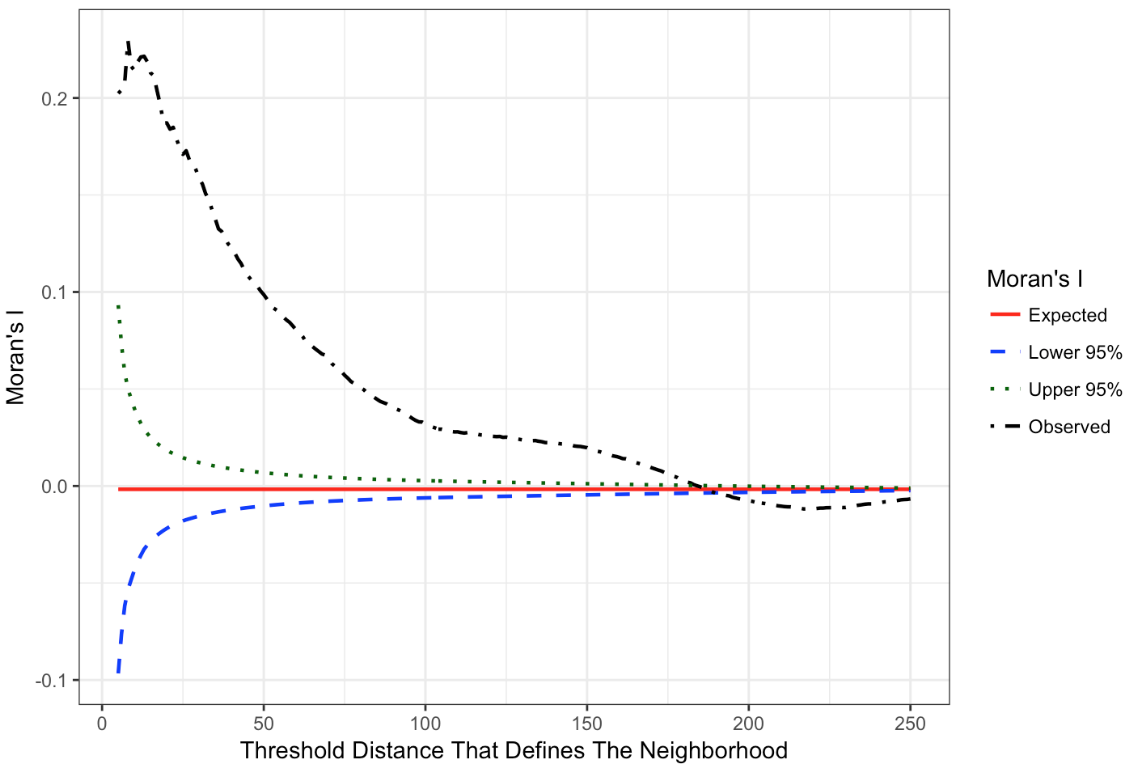
Note: this set of figures illustrates the simulation results of the Bayesian learning process in the theoretical model based on different royalty cost functions. The true potentials are $Q_1 = 1$, $Q_2 = 0$. The number of operators N is 50. The common belief $\Pr(s_i = j|Q_j = 1)$ is set to be 3/5 for each operator i and location j. The number of experiments is 10,000. The horizontal axes measure the prior beliefs that $Q_1 = 1$. Figures 4a and 4b are the mean numbers of operators and the mean royalty rates at the two locations for different prior beliefs based on the linear royalty cost function. Figures 4c and 4d are the same plots based on the reciprocal royalty rate function. Figures 4e and 4f are the plots based on the quadratic royalty rate function.

Figure 5: Coefficient Estimates of The Signal Indicator in Equation (7) and (6) Based on Different Prior Beliefs



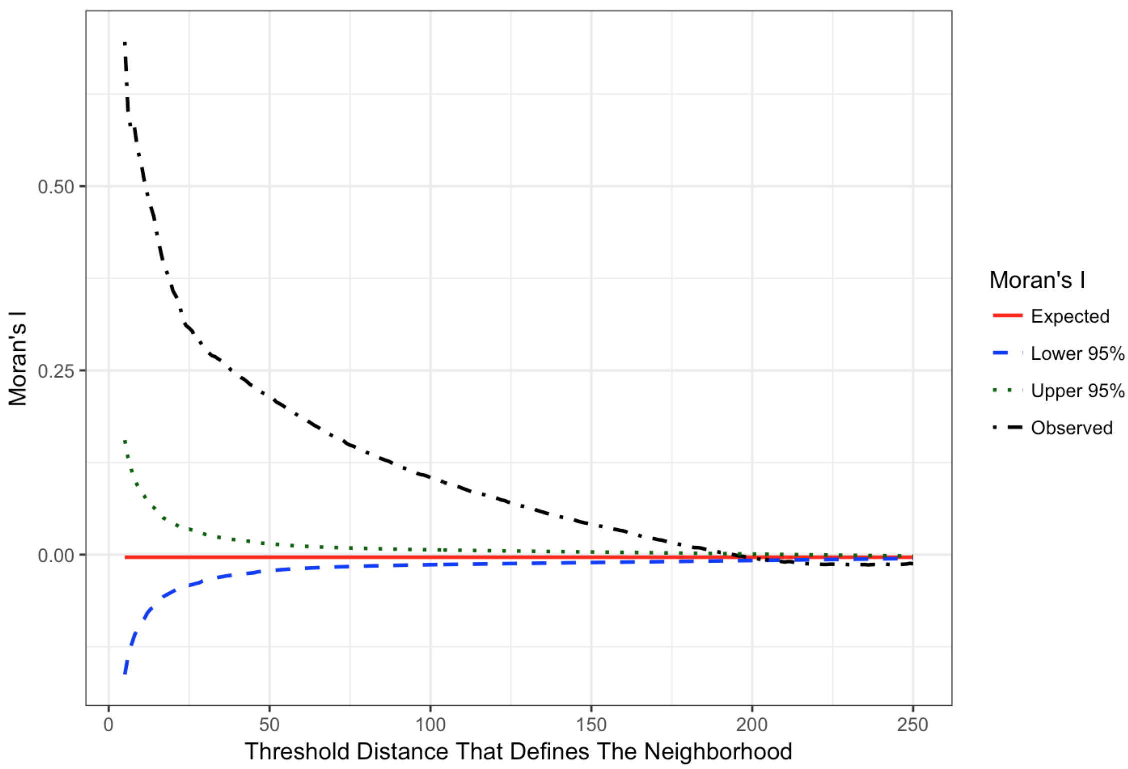
Note: the coefficient estimate for each prior belief is obtained using the simulated sample based on the linear royalty cost function $r(x) = x/N$, where N is set to be 50. Each sample consists of 10,000 experiments in which 50 operators sequentially choose their lease investment locations.

Figure 6: Moran's I for Municipality Level Royalty Rate Based on Different Neighborhood Defining Distances



Note: the data is obtained from Enverus. The average royalty rate of a municipality weighted by the acreages of the leases is used to calculate the Moran's I. The solid curve stands for the expected Moran's I for the different values of neighborhood defining distance. The lower dashed curve and the upper dotted curve represent the 95% confidence intervals. The dotted-dashed curve at the top of the figure illustrates the observed values of the Moran's I.

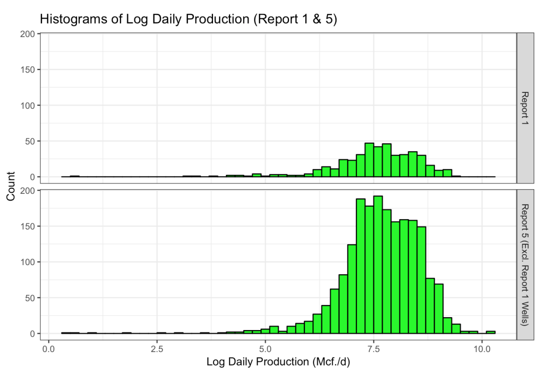
Figure 7: Moran's I for Municipality Level Productivity Based on Different Neighborhood Defining Distances



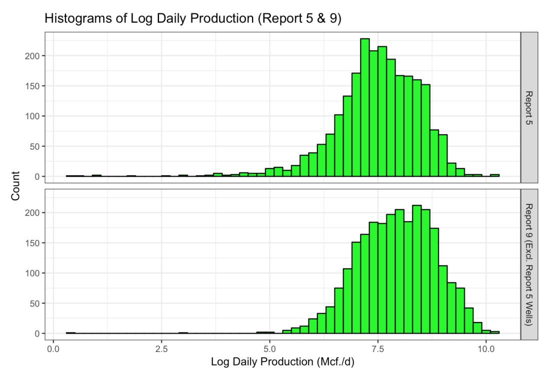
Note: the data is obtained from the PADEP. The solid curve stands for the expected Moran's I for the different values of neighborhood defining distance. The lower dashed curve and the upper dotted curve represent the 95% confidence intervals. The dotted-dashed curve at the top of the figure illustrates the observed values of the Moran's I.

Figure 8: Productivity Histogram of Production Report 1, Report 5, and Report 9

(a) Productivity Histogram of Report 1 and Report 5

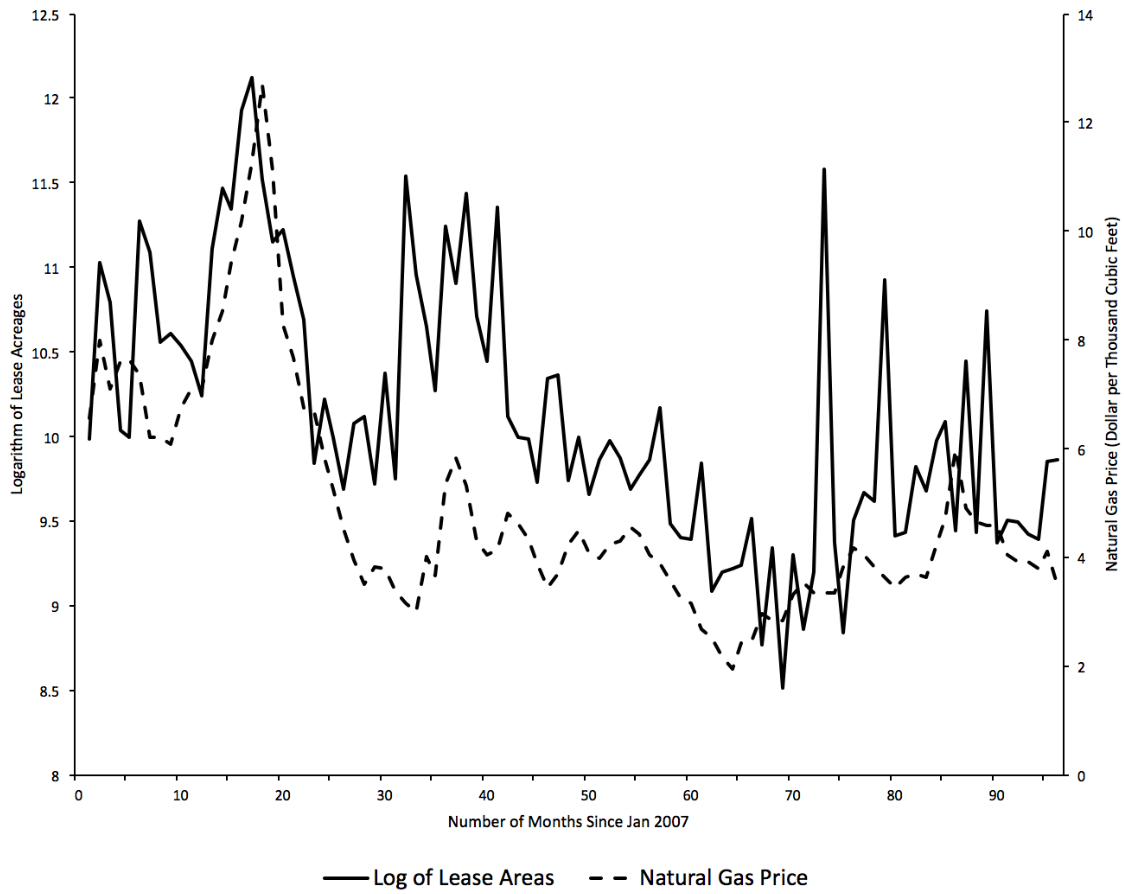


(b) Productivity Histogram of Report 5 and Report 9



Note: the data is obtained from the PADEP. The lower histogram of Figure 8a is based on the horizontal wells in report 5 but not in report 1, whereas the upper diagram of Figure 8b is based on the entire set of horizontal wells in report 5. The lower histogram of Figure 8b is based on the horizontal wells in report 9 but not in report 5

Figure 9: Monthly Incremental Lease in PA (Log Form) and Natural Gas Price (January 2007 - December 2014)



Note: the price data is obtained from EIA. The link is `https://www.eia.gov/dnav/ng/hist/rngwhhdm.htm`. The lease data is obtained from Enverus. The solid curve measures the logarithm of the lease acreages granted in each month in Pennsylvania. The dashed curve measures the natural gas spot price at the Henry Hub. The unit is dollar per thousand cubic feet.

Table 12: Parameters of The Simulation

Parameter	Setting 1	Setting 2	Setting 3
N	50	50	50
Q_1	1	1	1
Q_0	0	0	0
p	3/5	3/5	3/5
$r(x)$	x/N	$1 - \frac{1}{x+1}$	x^2/N^2
No. of rounds	10,000	10,000	10,000

Note: N is the number of firms in each round of the simulation. Q_1 and Q_2 are the potential outputs of L_1 and L_2. $p = \Pr(s_i = j | Q_j = Q^H), j = 1, 2$. $r(x)$ is the royalty rate at each location given x firms have chosen that location.

Table 13: Summary Statistics of The Productivity Histograms in Figure 8

	Figure 8a		Figure 8b	
	Upper	Lower	Upper	Lower
Mean	7.54	7.67	7.50	7.93
Median	7.62	7.70	7.55	7.96

Note: the lower diagram of Figure 8a is based on the horizontal wells in production report 5 but not in report 1, whereas the upper diagram of Figure 8b is based on the entire set of horizontal wells in report 5.