Information Disclosure and Locations Choices: a Study of the Marcellus Shale*

Wenshu Guo
Lingnan Univeristy

July 14, 2022

Abstract

Information disclosure usually affects the decision-making process of both buyers and sellers when the product quality is uncertain, with its ultimate consequence being hard to predict. The phenomenon is particularly evident in the natural resource exploration industry, where the true mineral recovery potential is seldomly known. This research studies the effect of production information disclosure on shale gas operators' lease location choices. Using a novel dataset from the Marcellus Shale, this article finds that the disclosure of production information makes shale gas operators lease more productive parcels despite higher cost, leading to a more efficient resource allocation. A theoretical model is then proposed to explain the learning process of the firms. While the finding of the article justifies the cost of collecting and distributing the production information by the regulating agency, it also shows that centralized information disclosure facilitates learning among firms, creating room for potential strategic behaviors in the long run.

Key Words: *Information Disclosure, 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, and Lingnan University for their comments. Contact information: wenshuguo@ln.edu.hk

I Introduction

Information disclosure usually affects the decision-making process of both buyers and sellers when the product quality is uncertain, with its ultimate consequence being hard to predict. In the natural resource exploration industry, for example, production information provides valuable knowledge for firms and mineral rights owners to infer the distribution of high-quality reserves. The resulting impact of the disclosure of such information on the firms' decisions as to where to explore and on the costs of obtaining relevant mineral rights is subject to the bargaining process between the firms and the mineral rights owners, which is not clear before it is empirically investigated. This research seeks to understand the effect of production information disclosure on shale gas operators' lease location choices, using the case of the Marcellus Shale. Since November 2010, the Pennsylvania government has been publishing the Marcellus Shale's production data on its website, providing information about reserve quality distribution to shale gas firms and landowners. Using a novel dataset of lease records and production quantities that has not been used in previous studies, this article finds that the disclosure of production information makes shale gas operators lease more productive parcels of land despite higher costs, leading to a more efficient resource allocation. A theoretical model is then proposed to reflect the tradeoff faced by the shale gas firms between greater output potentials and higher leasing costs. While the finding of the article justifies the cost of collecting and distributing the production information by the regulating agency, it also shows that centralized information disclosure facilitates learning among firms, creating room for potential strategic behaviors in the long run.

The prosperity of the shale gas industry has transformed the United States into a natural gas net exporter in the past decade. Among the major shale gas deposits in the United States, the Marcellus Shale is the largest, accounting for more than one-third of the country's shale gas production. Although the geological property of the Marcellus Shale is well understood today, estimates of its recoverable natural gas reserve quantity varied significantly over the past 20 years. In 2002, well before the Marcellus Shale was developed, the United States Geological Survey estimated that the Marcellus Shale contained 1.9 trillion cubic feet technically recoverable natural gas[1]. In the 2011's Annual Energy Outlook published by the United States Energy Information Administration, the estimate jumped remarkably to 410 trillion cubic feet[2]. One year later, the number was cut by more than one half to 141 trillion cubic feet in the 2012's Annual Energy Outlook[3]. The continuous revision reveals that there was considerable uncertainty regarding the output potential of the Marcellus Shale, which is due to several factors. First, the Marcellus Shale had a relatively short production history. Its first producing well was drilled in 2004, later than the development of other major shale gas formations, and only after 2010 had its output begun to proliferate. Second,

[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`

most of Marcellus Shale's early shale gas wells were drilled at the two productive "sweet spots" in Northeast and Southwest Pennsylvania, leaving the potentials of other regions hard to detect. Third, until 2010, there had not been public well-level production data. These facts underline the necessity of detailed production information for shale gas companies to learn about the geological features of the Marcellus Shale Formation. On November 1, 2010, the Pennsylvania Department of Environmental Protection (PADEP) began posting well-specific production data of the Marcellus Shale on its website. The data was updated semi-annually before 2015 and monthly afterward. Because the law of Pennsylvania forbids any form of seismic surveys before the shale gas firms sign lease contracts with the landowners, the production information presumably reshapes people's assessment of the shale gas reserve distribution. Regions with higher productivity are likely to be more attractive to firms. On the other hand, their leasing costs may also be higher as landowners are now more aware of the value their lands can generate. To what extent the firms' lease location choices are affected is an empirical question. The information disclosure hence provides a valuable opportunity to examine its impact on the mineral rights market.

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. To the extent that innovation and discovery are concerned, the shale gas exploration industry is a good example to study such impacts as the input and output can be easily described and measured. The current research addresses this question by following the dual approach of first conducting an empirical investigation into the role of information in shale gas drilling using the lease and production data from the Marcellus Shale and then building a theoretical model to explain the findings. The empirical strategies include a difference-in-difference study of the correlation between the cost and productivity of leases and an auto-regressive model embedded in a regression discontinuity design that aims to quantify the impact of the production information disclosure on shale gas operators' land leasing decisions. The results show that, following the disclosure of the production information, a 1% increase in productivity at locations that focus on shale gas production leads to an additional 0.0014% increase in royalty rates compared to the locations that focus on traditional natural gas production, suggesting that the disclosure of the production information makes shale gas operators target more productive leases. Prior to the disclosure, a 1 acre increase of the leases obtained by an operator in the past three months leads to a 0.593 acres increase in its current month's leases at the same location. Once the production information is available, the number goes up by 0.007 acres to 0.6 acres for each 1% increase in the location's productivity. The estimation also finds that every 1% increase in productivity is associated with a 23.591 acres direct increase in an operator's the current month's leases obtained.

The theoretical part of the paper considers a model in which investors sequentially choose between two locations to make their investments. The two locations differ in their quality, and higher quality leads to a higher returns. The investors don't know which location is of higher quality and, without other information, conduct Bayesian learning based on their private signals and the investment history. Each investor chooses the location that gives it the higher expected return after deducting the investment cost. As the number of investors choosing a particular location grows, the cost of investing in that location rises. The model does not dig into the mechanism that shapes the cost but assumes a reduced-form upward-sloping supply curve to reflect the scarcity of resources at each location. When additional information partially discloses the quality of the two locations, the investors' prior beliefs are modified. Analysis of the two-firm example shows that when the new information correctly adjusts the prior beliefs, the expected number of firms choosing the location with higher quality goes up. When there are more than two investors, simulations show that the expected number of investors choosing the location with higher quality increases in the investors' prior beliefs that the location is of higher quality. Thus when the disclosed information exposes the locations' true potentials, the expected number of investors choosing the higher quality location increases even though doing so becomes more costly, which induces more efficient investments. Further inspection of the simulation results suggests that the probability that an investor chooses a location is positively correlated with the fact that it observes a positive signal of that location, and the correlation first declines and then goes up as the investor's prior belief that the location is of higher quality gets larger. These patterns prove to be valid in the empirical analysis using actual lease and production data.

The finding of the paper highlights the value of information in improving resources allocation efficiency and has broader policy implications for industries featuring exploration and experimentation. The existing literature on information disclosure typically focuses the interaction between sellers and consumers. Dranove and Jin (2010) in their survey paper point out that voluntary disclosure is incomplete in many industries and provide evidence that disclosure affects both consumer choices and product quality. Bessen and Maskin (2009) show that when innovation is sequential, society may be better off if imitation is costless. This article provides empirical evidence showing that free imitation can improve investment efficiency in certain industries using shale gas exploration as an example. It is worth pointing out that mandatory disclosure is necessary because no firm has the incentive to release its production data unilaterally. This observation justifies the cost of collecting and verifying production information by the regulating agency and provides support to promoting similar centralized information disclosure policies to other industries.

Information externalities have long been documented in studies of oil and gas extraction (Hendricks and Robert H. Porter (1988); Hendricks and Kovenock (1989); Hendricks and Robert H. Porter (1996)). While this article shows that the disclosure of the shale gas production information leads to more efficient lease investments, in general, information externalities can have negative impacts on social welfare due to strategic disclosure. The example of shale gas exploration in Pennsylvania has a number of features that mitigates this concern. First, the law of Pennsylvania requires the shale gas firms to report their production data truthfully, eliminating the possibility of strategic disclosure. Second, the timing of implementing the disclosure policy is exogenous and unpredictable. The shale gas firms could not foresee the availability of additional information and plan accordingly. This is not to say that strategic concerns do not exist in this industry, and free-riding behaviors can naturally arise in a repeated disclosure setting. An early disclosure would allow the firms to take more advantage of the information externalities, while on the other hand exacerbate strategic delays. The study of strategic concerns and optimal timing of disclosure is left for future research.

The rest of the article is organized as follows: section II discusses the related literature. Section III reviews the background of the industry. Section IV describes the data. Section V introduces the empirical model. Section VI presents the estimation results. Section VII provides a theoretical framework. Section VIII concludes the article.

II Related Literature

This research is mainly related to three strands of literature. The first is the strand of literature that studies quality disclosure. Early theory papers typically focus on sellers' incentives to disclose their product quality (see Akerlof (1970), S. J. Grossman and Hart (1980), Sanford J. Grossman (1981), Milgrom (1981), Jovanovic (1982)). One condition that is frequently assumed is costless unravelling. When quality disclosure is costly, Sanford J. Grossman (1981) and Jovanovic (1982) show that mandatory disclosure may reduce social welfare, while Matthews and Postlewaite (1985) and Shavell (1994) show that compulsory disclosure policies may induce sellers to collect less information. In the shale gas industry, the cost for regulators to collect and verify production information is not negligible. It is therefore worth asking what would otherwise happen to the shale gas operators had the production information not been disclosed. Dranove and Jin (2010) provide a comprehensive review of the theoretical and empirical development of the literature on quality disclosure. The more recent theory papers cited in the survey relevant to this research include Fishman and Hagerty (2003) and Board (2009). Fishman and Hagerty (2003) point out that mandatory disclosure may prove attractive to consumers when product information is difficult to understand. Board (2009) shows that under competition between sellers, voluntary disclosure may not be realistic, and mandatory disclosure policies are called for to raise consumer surplus. In the shale gas industry, mineral rights owners usually do not have good knowledge of the exploration and production plans of the shale gas operators, and it is hard to find incentives for the shale gas operators to disclose their production information voluntarily. Should the information be beneficial to mineral rights owners, mandatory disclosure may be the optimal option. Still, this article differs from the theoretical quality disclosure literature because it is the shale gas operators, which is the demand side of the mineral rights market, that are required to disclose their production information. How does buyer's disclosure affect the market participants? incentives and decisions will be one of the contributions made by this article to the existing literature.

A large number of empirical studies of information disclosure are related to this article, among which a majority of earlier papers are surveyed by Dranove and Jin (2010). Mathios (2000) demonstrates that many producers of high-fat salad dressing withheld information before the Nutrition

Labeling and Education Act came into effect. Jin (2005) finds that only half of all health maintenance organizations (HMOs) disclosed their quality due to production differentiation concerns. Their findings show that voluntary quality disclosure is often hampered in practice, and mandatory disclosure policy is needed when necessary. All available empirical evidence supports vertical sorting resulting from quality disclosure (Ippolito and Mathios (1990), Hastings and Weinstein (2008), Jin and Sorensen (2006), and Dafny and Dranove (2008)). However, research by Marshall et al. (2000) and Dranove and Sfekas (2008) show that quality disclosure may fail to affect demand if ratings are difficult to understand or only confirm what consumers already know. Thus whether the disclosed production information affects the decisions of shale gas operators is subject to empirical examination. Regarding information disclosure's impact on product quality, different papers give different answers. Jin and Leslie (2003) find that the requirement for restaurants to post hygiene grade cards reduced hospitalizations. By contrast, Dranove, Kessler, et al. (2003) find that cardiac surgery report cards led to selection behaviours by providers. It is, therefore, worth studying the impact of the production information disclosure on the shale gas industry-wide production outcomes. Some of the more recent empirical papers related to this research include Luco (2019)'s study of price disclosure in the retail gasoline market and Guo, Sriram, and Manchanda (2021) that studies the impact of the transparency law on industry payments to physicians.

The second strand of related literature concerns learning and information externality, especially in the oil and gas industry. In the seminal papers by Banerjee (1992)and Bikhchandani, Hirshleifer, and Welch (1992), the authors propose sequential Bayesian learning models to study herding behaviors and informational cascades. The models can be well applied to the current setting to demonstrate how shale gas operators learn about reserve distributions. In the series of studies of offshore oil and gas extraction, Hendricks, Robert H. Porter, and Boudreau (1987), Hendricks and Robert H. Porter (1988), Hendricks and Kovenock (1989), Hendricks, Robert H. Porter, and Wilson (1994), Robert H. Porter (1995), Hendricks and Robert H. Porter (1996) and Hendricks, Pinkse, and Robert H Porter (2003) provide thorough investigations into the role of information in the auctions of offshore oil and gas leases and the exploration of offshore wildcat tracts. The above studies of information externality can be broadly classified as examples of social learning, which has been documented in a variety of other industries and is a crucial driving force for technology adoption and information spillover (Irwin and Klenow (1994), Foster and Rosenzweig (1995), Munshi (2004), Conley and Udry (2010)). Bolton and Harris (1999), in their study of strategic experiments in a multi-player setting, show that when information is a public good, both a free-rider problem and an incentive to experiment in order to induce future experimentation by other players can arise. Bessen and Maskin (2009) further point out that when innovation is sequential, society may be better off if imitation is costless. This article follows the literature and studies if social learning facilitates the shale gas firms to understand the Marcellus Shale's reserve distribution, which leads to more efficient investments.

The third strand of literature this article contributes to is the recent progress in industrial organization about oil and gas extraction. Some of the most representative articles include 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). This article will provide a new angle to the above literature by evaluating the production information disclosure policy, which has not been studied before.

III Industry Background

This section provides the institutional backgrounds of shale gas exploration in the United States and the Marcellus Shale in particular. Shale gas is a kind of "unconventional natural gas", a term used by the industry to denote natural gas trapped in clay shales (shale gas), in coal beds (coal bed methane), and in dense sandstones with low permeability (tight gas, see Zhiltsov and Semenov (2017)). The first shale gas well (also the first natural gas well) in the United States was drilled by William Hart in the shallow shales of Fredonia, New York, in 1821. Due to the limited extraction technology and the lack of adequate geological knowledge about shale formations, commercialized shale gas production was negligible before 1980. The natural gas shortage in the 1970s urged the government to launch a series of policies to stimulate the discovery of unconventional natural gas resources. Among the most prominent ones were incentive pricing, tax credits, and the Eastern Gas Shales Project, an R&D program targeting the Devonian-age shales in the Eastern United States. According to Wang and Krupnick (2015), at the start of the gas shales program, "the industry had a poor understanding of the physical and chemical characteristics of Devonian shale, and the estimates of recoverable reserves were highly uncertain." The goal of the program was to assess the shale gas resource base and to introduce more efficient production technology. Starting from the 1980s, Mitchell Energy began to develop the Barnett shale in northern Texas by applying technologies that were still in the early stages, such as massive hydraulic fracturing, 3-D seismic imaging, and horizontal drilling. The promising findings of Mitchell Energy, along with high natural gas prices in the early 2000s, encouraged the exploration of other shale gas "plays" - shale formations containing a significant amount of natural gas - including the Fayetteville Shale in Arkansas, the Haynesville Shale in Louisiana, the Woodford Shale in Oklahoma, and the Marcellus Shale in Pennsylvania. Figure 1 displays locations of the major shale gas plays in the lower 48 states. The exploitation of the shale plays drove the shale gas output to rapidly expand. The share of natural gas from shales and tight oil plays as a percentage of the United States' dry natural gas production jumped from 10.9% in 2005 to 50.2% in 2015, as shown by Figure 2. The fast growing shale gas output was accompanied by a drop of natural gas price, from 8.69 dollars per Million Btu in 2005 to 2.62 dollars per Million Btu in 2015. Cheap natural gas shadowed the development of the shale gas industry. As firms struggled to make a balance, a series of consolidations followed. This industry-wide restructuring goes beyond the scope of this article and is not discussed here.

Among the major shale gas plays, the Marcellus Shale has the largest output. The Devonian Marcellus shale extends throughout most of the Appalachia Basin in the Northeastern United States. Figure 4 shows a structural map of the Marcellus Formation, which stretches from southern New

York, across northern and western Pennsylvania, down to West Virginia and eastern Ohio. Its depth ranges from 5000 feet to 9000 feet. The Eastern Gas Shales Project in the 1970s identified the Marcellus Shale as a potential source of natural gas supply. Yet early exploration yielded no significant outcomes due to technology limitations (see Zagorski, Wrightstone, and Bowman (2012)). Following Mitchell Energy's success in the Barnett Shale, Range Resources completed its first Marcellus well in October 2004 in Washington County, Pennsylvania, establishing the production rates high enough to attract wide industry interest. As more operators joined the exploration process, the Marcellus Shale quickly became the major driving force of shale gas development, and in September 2012, it surpassed the Haynesville Shale to become the leading producer of both shale gas and overall natural gas in the United States. As Figure 3 shows, in January 2016, more than one-third of the shale gas in the United State was produced from the Marcellus Shale, making up 17.3% of the United States' natural gas gross withdraws[4].

Although the Marcellus Shale spreads across a number of states, over 90% of its output is produced in Pennsylvania. Two "sweet spots" in Pennsylvania have been identified, which are Susquehanna County and Bradford County in the northeast, and Washington County and Greene County in the southwest. This article focuses on the period between 2007 and 2014, a time in which shale gas production in Pennsylvania boomed and the industry consolidation resulting from cheap natural gas prices had not yet commenced. Figure 5 plots the daily gross natural gas withdraws and the daily shale gas production in Pennsylvania from 2007 to 2014. The gross natural gas productivity increased over twentyfold, from less than 0.5 billion cubic feet per day in 2007 to more than 11 billion cubic feet per day in 2014, mostly driven by the expansion of the shale gas output. One contributor to the growth is the prevalence of horizontal drilling. A horizontal well is first drilled vertically until the drill bit reaches the target shale formation. The drill bit then turns and drills horizontally through the shale formation for about one mile. Figure 6 illustrates the configurations of the two types of wells. Compared with vertical wells, horizontal wells are more costly to drill, but are much more productive and less space consuming. Figure **??** plots the numbers of newly drilled vertical and horizontal wells in Pennsylvania from 2007 to 2014. Among the wells drilled in 2007, less than one percent were horizontal. That share jumped to 60% in 2010, and continued growing to over 90% in 2014. All horizontal wells are devoted to shale gas production, whereas a significant share of vertical wells still target conventional natural gas. For these reasons, this article focuses on the natural gas production from horizontal wells.

The law of Pennsylvania requires firms to obtain mineral rights by signing lease contracts with the landowners, usually local residents, before conducting seismic surveys or drilling natural gas wells. The lease contracts are publicly viewable in county courthouses. The term of a contract is typically three to five years. If production meets the minimum requirement, the lease extends to as long as the production continues. Otherwise the lease expires at the end of the contractual term. The cost of the lease consists of two components: an up-front payment (bonus) and a share of the future revenue (royalty). The common range of the royalty rate is 12.5% to 17.5%. If a firm wants

[4]See the trend and data provided by EIA: `https://www.eia.gov/dnav/ng/hist/n9010us2m.htm`

to drill a well, it first needs to apply for a drilling permit from the PADEP. There are two types of permits. A conventional permit is issued for a well targeting traditional natural gas, and an unconventional permit is issued for a well targeting shale gas. On November 1, 2010, the PADEP began posting well-specific production data of the Marcellus shale gas wells on its website. The first set of data covers the production from July 2009 to June 2010. It documents the owner, location, production quantity, and the number of producing days of every shale gas well in Pennsylvania. The production data is updated semi-annually before 2015 and monthly thereafter. Prior to the disclosure, a shale gas firm could not observe the productivity of the wells drilled by other companies. The following sections analyze the impact of the production information disclosure on firms' land leasing decisions.

IV Data

This research relies on a novel dataset that has not been used in previous studies, which documents the lease and production information of the Marcellus Shale during 2007-2014. This section describes the data used in this article along with the necessary steps through which the data is processed for empirical analysis. The dataset is made up of several components. The first set of data contains the drilling permits issued by the PADEP from January 2007 to December 2014. The sample is available on the PADEP's website. Each observation includes the operator, issuing date, location (in terms of county and municipality), and type (conventional or unconventional) of a drilling permit. The data is used to select operators focusing on shale gas exploration. Table **??** shows the number of unconventional permits received by different operators. There are 390 operators receiving at least one drilling permit, among which 107 received at least one unconventional drilling permit. Chesapeake tops the list with 3062 unconventional permits. Four other operators, Range Resources, SWEPI, EQT Production, and Talisman, acquired more than 1000 unconventional permits each. Apart from the five operators, there are 22 operators receiving at least 100 unconventional permits, and another 31 operators receiving at least 10 unconventional permits. The 58 operators combined received 18157, of 99.13% of the unconventional permits issued by the Pennsylvania DEP during the sample period. These operators are treated as having a significant presence in the shale gas exploration field.

The second set of data consists of the lease records. The data is acquired from Enverus (previously Drillinginfo), an energy consulting firm. It includes all the lease contracts in Pennsylvania signed between January 2007 and December 2014, which shows the dynamic locations choices of the shale gas operators and the corresponding costs. Each contract documents the acreage, operator, location (county and municipality), starting date, royalty rate, and term of a lease. The bonus variable is not observable as it is kept secret by the operators. To focus on the leases taken for shale gas extraction, the leases granted to the operators with fewer than 10 unconventional permits received between January 2007 and December 2014 are dropped out of the sample.

There are two types of lease records. Apart from the original lease contracts, there are lease memos, which are simplified versions of contracts in which some information, including the royalty rate, is concealed. There are duplicated observations of lease contracts and lease memos for a single plot of land, possibly due to different grantors or subplots. For example, a 20-acre lease may contain three subplots, A, B, and C. Subplot A is leased on April 1, 2009 from household h1, and its original contract is observable. Subplot B is leased on April 1, 2009 from household h1 and h2, each with a lease memo. Subplot C is 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 the 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 April 5's leases are for subplot C, not for a distinct parcel, is to refer to the plot map included in the contract, a method impractical to follow 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 considered 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 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 dates, and the 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 is 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 possibly different from the reality. Nevertheless, it correctly reflects the timing and structure of the lease. The selected sample contains 90241 lease records.

Table **??** shows the numbers, average acreages, average contractual term, and acreage weighted average royalty rates of the leases granted in each year from 2007 to 2014. The number of granted leases peaked in 2008 at over 20000, then dropped to fewer than 5000 in 2012, after which it rose to more than 10000 in 2014. The average area of a lease was around 70 acres in 2007 and fell below 30 acres starting from 2009. The average royalty rate in the last column is weighted by acreages and kept fluctuating around 15%.

There are 67 counties and 2464 municipalities in Pennsylvania. The selected lease sample involves 28 counties and 675 municipalities, and spans across 96 months from January 2007 to December 2014. For each combination of municipality and month, the acreage-weighted average royalty rate is used as the expected cost of leases. To accommodate for the pairs of municipality and month in which there are no lease observations, this article explores the spatial correlation of the royalty rates. Figure **??** displays the Moran's I for the municipality level average royalty rate based on different neighborhood defining distances, using the subsample of the original contracts. The Moran's I measures the spatial autocorrelation of a variable. It is defined as

$$I = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij}(X_i - \overline{X})(X_j - \overline{X})}{\sum_i (X_i - \overline{X})^2} \tag{1}$$

where 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 otherwise. 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 municipality i and municipality j is less than or equal to d, and $w_{ij} = 0$ otherwise. The figure shows that the Moran's I for the municipality level average royalty rate is significantly positive when the neighborhood defining distance is between 0 and over 175 miles, suggesting that the royalty rate is highly spatially correlated. Let L denote the set of the 675 municipalities in which there are lease records. To get an approximation of the expected royalty rate $R_{l,m}$ for each pair of municipality and month (l, m), $l \in L$ and $m = 1, 2, \ldots, 96$, the following loop is run over a list of neighborhood defining distances and averaging periods. The algorithm takes a few steps:

- 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 the unit of miles. Create the list of averaging periods $T = \{1, 3, 6\}$ with the unit of 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 at municipality l within the 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 municipality i and municipality 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 with 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 again used to filter the lease records that are made for the purpose of conventional 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 chosen 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 third set of data keeps track of the shale gas production history in Pennsylvania from July 2009 to June 2014. The data is acquired from the PADEP's website. It consists of nine production reports that document the locations (longitude and latitude), configuration types (vertical or horizontal), operators, production quantities, and number of producing days of the shale gas wells in Pennsylvania. Table **??** shows the time at which the reports are released and the production period each report covers. The first report came out in November 2010. It covers the shale gas production from July 2009 to June 2010. The other reports were released in Februaries and Augusts of the years from 2011 to 2014, each covering six months of production. Table **??** describes the summary statistics of the production data. The first two columns are the numbers of active wells in each production report, both horizontal and vertical. The number of vertical wells remains relatively stable, whereas the number of horizontal wells steadily increases, from fewer than 500 in the first report to more than 4800 in the ninth report. The third and the fourth columns are the mean producing days of the horizontal and vertical wells in each production report. The horizontal wells are active for 133 days on average in the second half of 2010. The number grows to 159 days in the second half of 2012 and remains at about 160 days thereafter. The fifth and the sixth columns are the average daily production quantities of the two types of wells. The unit is million cubic feet (MMcf.). The average productivity of the horizontal wells fluctuates between 2.81 MMcf./d in the second half of 2010 and 2.56 MMcf./d in the second half of 2012. In comparison, the highest average productivity of the vertical wells is 0.28 MMcf./d in the first half of 2014. The last two columns are the standard deviations of the daily production quantities. The productivity variations of the horizontal wells are significantly and consistently higher than those of the vertical wells. These statistics show that there are obvious distinctions in both scale and productivity between horizontal and vertical wells, leading to potentially different inferences of the shale gas reserve distribution. To avoid possible confusions, only production of horizontal wells is used for empirical analysis. Figure **??** plots the Moran's I for the municipality level horizontal well productivity against different values of neighborhood defining distances, using the ninth production report. The average daily production of the horizontal wells in each municipality is used as the variable X in Equation 1. The value of the Moran's I is significantly positive for a wide range of distances from 0 to near 200 miles, which implies that shale gas productivity, like royalty rate, is also positively spatially correlated. The fact highlights the possibility to speculate about the shale gas reserve quality of an unexplored region based on the productivity of the adjacent areas.

One feature of the 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 **??** plots the frequency histograms of the horizontal wells' daily production using the first, the fifth, and the ninth production report. The daily production is in logarithmic form, measured by thousand cubic feet per day (Mcf./d). The upper diagram of Figure **??** is the productivity histogram of the horizontal wells in report 1. The vertical axis measures the number of wells. The lower diagram of Figure **??** 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 **??** is the analogous comparison between the productivity histogram of the horizontal wells in report 5 and the productivity histogram of the horizontal wells in report 9, where the wells in report 5 are excluded. Table **??** reports the means and medians of the histograms in Figure ??. 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 the shale gas operators benefit from the production data by choosing the locations with higher potentials.

The gazetteer of Pennsylvania is downloaded from the United States Census Bureau. The file provides tractable indices to the counties and municipalities in Pennsylvania, along with the population, land area, the longitude and latitude of each municipality surveyed by the 2010 Census. A municipality is a subdivision of a county. It is the smallest geographic unit used in the empirical analyisis. 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 are treated as a single unit with their population sizes and land areas added up. The middle points of the longitudes and latitudes of the two municipalities are used as the coordinates of the combined unit.

The final piece of data is the natural gas price. The monthly natural gas spot prices at the Henry Hub are used as as approximations to the wellhead prices in Pennsylvania. Located in Erath, Louisiana, the Henry Hub is a vital distribution hub of the natural gas pipeline system, and has been for years a benchmark for the natural gas futures trading in the United States. Its natural gas spot price is closely correlated with North American wellhead prices. The data spans across 96 months, from January 2007 to December 2014. The highest price occurs in June 2008, at 12.69 dollars per Mcf. The price falls below 10 dollars per Mcf. after July 2008, and declines to its lowest level in April 2012, at 1.95 dollars per Mcf., after which it slightly bounces to around 4 dollars per Mcf. Figure ?? plots the monthly natural gas spot prices along with the logarithm of the lease acreages granted in each month in Pennsylvania from January 2007 to December 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.

V Empirical Model

This section employs reduced form models to analyze the impact of the production information on the land leasing decisions of shale gas operators, using the data presented in the last section. It begins with a series of difference-in-difference type of models to study the correlation between productivity and the cost of acquiring leases. An auto-regression model embedded in a regression discontinuity design is then introduced to quantify the impact of the production information disclosure on the shale gas operators' land leasing decisions. To motivate the empirical analysis, Figure ?? plots the acreage-weighted average royalty rate against the average log daily production of the horizontal wells for each municipality in Pennsylvania where the values of both variables are observed. The unit of the daily production is thousand cubic feet per day, or Mcf./d. The productivity is calculated using the ninth report, which covers the shale gas production from January 2014 to June 2014. Thus it can be understood as the "true" productivity of a municipality. The royalties rates of the circular dots are calculated using the leases granted between January 2007 and October 2010, prior to the information disclosure. The royalties rates of the crosses are calculated using the post-disclosure leases granted between November 2010 and December 2014. The dashed and the solid curves are the fitted linear regression lines, with the 95% confidence intervals given by the shaded regions. The figure implies that royalty rates are more responsive to productivity when the shale gas production data is publicly available. The intuition behind the comparison is that if the production data generates additional information about the quality of the shale gas reserves, locations with higher productivity will attract greater lease investment, leading to higher royalty rates.

More formally, consider the following regression model:

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

where t stands for period, which is either "before disclosure" or "post disclosure", and j stands for municipality. $Royalty_{j,t}$ is the average royalty rate of municipality j at period t, weighted by the lease acreages. $Prod_j^9$ is the average log daily production (Mcf./d) of municipality j, calculated using the ninth production report. It is fixed for each municipality. D_t is the information disclosure indicator variable that is equal to 0 if period t is "before disclosure" and 1 if period t is "post disclosure". $\varepsilon_{j,t}$ is an unobserved error term that is 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 will represent an informative disclosure.

To see the relationship between productivity and royalty rate on a yearly basis from 2007 to 2014, the following regression is run:

$$Royalty_{j,t} = \beta_0 + \beta_{1,t} Prod_j^1 + \beta_t + \varepsilon_{j,t} \tag{3}$$

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

Finally, the set of municipalities in the sample is divided into sub-groups that are "shale gas oriented" and "non-shale gas oriented", based on the number of active horizontal wells. Because the disclosed production 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 lease cost and productivity across the two sub-samples, the impact of information disclosure can be identified. Specifically, consider the following regression model:

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

where t stands for period, which is either "before disclosure" or "post disclosure", and j stands for municipality. $Royalty_{j,t}$, $Prod_j^9$, and D_t are defined as before. S_j is the municipality type indicator variable that takes value 1 if municipality j is classified as shale gas oriented and 0 otherwise. There are 32 counties in which at least one horizontal well is active from January 2014 to June 2014. The municipalities in the 12 counties each with more than 50 active horizontal wells are classified as shale gas oriented. $\varepsilon_{j,t}$ is assumed to be i.i.d. normal. The coefficient of productivity is allowed to be different both for different periods and for different groups of municipalities. The critical parameter in this model is β_4, which captures the distinction in the impacts of the production information disclosure on the locations of leases taken for share gas production and other purposes.

The above models all seek to provide evidence on the production information disclosure's impacts indirectly by examining the dynamics of the correlation between productivity and the cost of acquiring leases. The remaining part of the section looks into an auto-regressive model that directly measures the effects of the disclosure. Let $\mathcal{I}$ denote the set of shale gas operators, $\mathcal{J}$ denote the set of municipalities. Let t denote the number of months past January 2007, where $t = 1, 2, \ldots, 96$. The production information is disclosed at $t = 47$. For each operator $i \in \mathcal{I}$ and each municipality $j \in \mathcal{J}$, let $a_{i,j,t}$ be the acreages of the leases granted to operator i in municipality j and month t, measured by acres. Define

$$\mathcal{O}_t = \left\{ i \in \mathcal{I} \middle| \sum_{j \in \mathcal{J}} a_{i,j,t} > 0 \right\} \tag{5}$$

to be the set of operators that are actively taking out leases in month t. Let $d_{j,k}$ denote the distance in miles between municipalities j and k. For each municipality $j \in \mathcal{J}$, define $\mathcal{N}_j = \{k \in \mathcal{J} | d_{j,k} \leq 5\}$ to be the set of its neighboring municipalities. Note that with this definition, $j \in \mathcal{N}_j$, i.e., a municipality is itself one of its neighboring municipalities. The data generating process is assumed to be the following: in each month t, $1 \leq t \leq 96$, each operator $i \in \mathcal{O}_t$ decides $a_{i,j,t}$, i.e., the acreages of the leases it shall take in municipality j, for each $j \in \mathcal{J}$. For an operator i that is not actively taking leases, i.e. $i \notin \mathcal{O}_t$, define $a_{i,j,t} = 0$ for every municipality $j \in \mathcal{J}$. The operators observe the history of lease taking of all operators in all municipalities $\{a_{i,j,\tau}\}_{i \in \mathcal{I}, j \in \mathcal{J}, 1 \leq \tau < t}$, but do not observe the productivity of the shale gas wells until $t = 47$. The value of $a_{i,j,t}$ is assumed to be determined by the following latent variable model:

$$a_{i,j,t} = \begin{cases} a^*_{i,j,t}, & \text{if } a^*_{i,j,t} > 0 \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

and the latent acreage $a^*_{i,j,t}$ is given by

$$a^*_{i,j,t} = \alpha_i + \boldsymbol{S}_{i,j,t}\boldsymbol{\delta} + \theta\overline{\log q}_{j,t} + \left(\overline{\log q}_{j,t} \times \boldsymbol{S}_{i,j,t}\right)\boldsymbol{\gamma} + \boldsymbol{X}_{i,j,t}\boldsymbol{\beta} + \varepsilon_{i,j,t} \tag{7}$$

where $\boldsymbol{S}_{i,j,t}$ is a vector of signals regarding the quality of the leases at location j, observed by operator i in month t. $\overline{\log q}_{j,t}$ is the average log daily production of the horizontal wells in the municipalities in $\mathcal{N}_j$, calculated using the most recent production report available in month t[5]. The unit of the horizontal wells' daily production is thousand cubic feet per day (Mcf./d). $\boldsymbol{X}_{i,j,t}$ is a vector of control variables, including natural gas price p_t, royalty rate $r_{j,t}$, population size Pop_j and land area $Aland_j$ of municipality j, an indicator variable Pub_i that takes value one if operator i is public, and a trend variable t. α_i is the operator fixed effect. $\varepsilon_{i,j,t}$ is an unobserved random term that is assumed to be i.i.d. normal across operators, municipalities, and months. $\boldsymbol{\delta}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\beta}$ are coefficient vectors. For notational simplicity, define $a_{i,j,t} = 0$, $\forall i \in \mathcal{I}, j \in \mathcal{J}$, and $t = -3, -2, -1, 0$. The vector of signals $\boldsymbol{S}_{i,j,t}$ include the following variables:

1. Own recent lease acreages $A^{Recent}_{i,j,t}$: the lease acreages operator i obtains in municipality j in the past three months, defined by

$$A^{Recent}_{i,j,t} = \sum_{\tau=t-3}^{t-1} a_{i,j,\tau} \tag{8}$$

2. Peer's recent lease acreages $A^{Recent}_{-i,j,t}$: the lease acreages operators other than i obtain in municipality j in the past three months, defined by

$$A^{Recent}_{-i,j,t} = \sum_{\tau=t-3}^{t-1} \sum_{k\in\mathcal{I}\setminus\{i\}} a_{k,j,\tau} \tag{9}$$

3. Accumulated lease acreages $A^{Accumlated}_{j,t}$: the accumulated acreages that have been leased in municipality j up to month $t-4$, defined by

$$A^{Accumlated}_{j,t} = \sum_{\tau=-3}^{t-4} \sum_{i\in\mathcal{I}} a_{i,j,\tau} \tag{10}$$

The above model is built upon the assumption that leasing decisions reflect operators' beliefs of a location's output potential and should be correlated over time. The own recent lease acreage and the peer's recent lease acreage stand for 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 lease acreages are separated to distinguish between the effects of signals from different sources. The accumulated lease acreage represents the lease investment history in the municipality. A larger accumulated lease

[5] To accommodate zero production, this article uses $\overline{\log(q_{j,t} + 1)}$

acreage implies that more operators think positively of the municipality in the past. The average productivity of the horizontal wells provides a direct measurement of the municipality's shale gas reserve quality. The neighboring municipalities are included because there exists significant positive spatial correlation in shale gas productivity. When $t < 47$, the productivity signal is not available and is assumed to be zero for all the municipalities[6]. For each month $t \geq 47$, the latest production report available in month t is used to calculate the average productivity. The coefficient θ is expected to be positive if the production information generates new knowledge of the reserve quality. The coefficients of the interactive terms between the past lease acreages and the average productivity will also be positive if higher productivity signals leads to stronger autocorrelations between leasing decisions.

VI Estimation Results

This section presents the estimation results of the empirical models discussed in the last section. Table **??** reports the coefficient estimates of the model given by (2). The coefficient of productivity is 0.0037 with a standard error of 0.0011, indicating that before the disclosure of the production information, a 1% increase in a municipality's daily production is associated with 0.0037% increase in absolute value of the municipality's average royalty rate. The coefficient of the interactive variable between productivity and disclosure dummy is estimated to be 0.0058 with a standard error of 0.0009. Thus once the production data is publicly available, a 1% increase in a municipality's daily production leads to an additional 0.0058% increase of the the municipality's royalty rate in absolute value, or 1.57 times higher compared to the prior-disclosure situation. The finding is consistent with the pattern in Figure **??** and the hypothesis that the production information makes locations with higher productivity attract significantly more lease investments, leading to more efficient resource allocation.

For the year-by-year regression model given by (3), the coefficient of productivity in 2007 is estimated to be -0.0012 with a standard error of 0.0027, which indicates that there is no statistically significant correlation between royalty rate and productivity in 2007. Figure **??** displays the coefficient estimates of the interaction between productivity and the year fixed effects. The round dots connected by the solid line represent the change of productivity's coefficient estimates in the corresponding years relative to the 2007's value. The dashed lines connects the upper and lower bounds of the 95% confidence intervals for the estimated coefficients. The vertical line stands for the time of the initial production data disclosure (November 2010). The coefficients of productivity in 2008 and 2009 are estimated to be higher than the value in 2007, though the increments are not statistically significant under the 5% significance level. The coefficient of productivity in 2010 is significantly higher than the 2007 estimate by 0.0088, and the increase jumps to 0.0137 in 2011 with a standard error of 0.0041. The estimates in 2012, 2013, and 2014 are not significantly different from the estimate in 2007. The result suggests that in the year following the disclosure of the

[6] An operator can observe its own production even when $t < 47$. This information, however, is not available in the data, and hence is dismissed in the model.

production information, royalty rates became positively correlated with productivity. The change in the correlation is significant both statistically and in magnitude. The positive change, however, does not continue into the years after 2011. One possible explanation to this finding is that the disclosed production information introduces a shock to the firms' understanding about the shale gas reserve distribution, which is completely digested in the following years.

Table **??** reports the coefficient estimates of the model given by (4). The "benchmark" coefficient of productivity is estimated to be 0.0047 with a standard error of 0.0011. It implies that prior to the production information disclosure, for municipalities that are not shale gas oriented, royalty rate is slightly positively correlated with the productivity of horizontal wells, with 1% increase in daily production leading to 0.0047% absolute increase in royalty rate. The coefficient of municipality type indicator variable is estimated to be -1.37×20^{-5} with a standard error of 0.0004, which is not significant either statistically or economically. Thus prior to the information disclosure, there is no testable difference in terms of the responsiveness of royalty rate to productivity between the two types of municipalities. The coefficient period indicator variable is estimated to be 0.0009 with a standard error of 0.0006, which is not significant either. It shows that for the group of municipalities that are not shale gas oriented, the disclosure of the production information has no significant impact on the choices of lease locations. The coefficient of the interactive variable between the type indicator and the period indicator is estimated to be 0.0014, with a standard error of 0.0006. The estimate is positive and statistically significant, suggesting that the disclosure of the production data generates an additional 0.0014% absolute increase in royalty rate with 1% increase in productivity for the shale gas oriented municipalities. Although the value is economically small, it nevertheless accounts for around one third of the marginal contribution of productivity to royalty rate for municipalities that are not shale gas oriented. This finding is consistent with the intuition that the production information disclosure only affects land leasing decisions targeting shale gas production.

The model defined by equations (6) and (7) is estimated using maximum likelihood. Table **??** reports the estimation results. The second and third columns are the coefficient estimates and standard errors of the baseline model. The coefficient of the operator's own lease acreage in the past three months is estimated to be 0.593 with a standard error of 0.007. It implies that, other things equal, for a one acre increase in the leases granted to an operator in a given municipality in the past three months, there corresponds to a 0.593 acres increase in the latent lease acreage taken by the operator in the same municipality in the current month, indicating a strong correlation of the leasing decisions over time. The coefficient of the peer's lease acreage in the past three months is estimated to be -0.003 and is insignificant. The result is somewhat surprising in that it suggests that operators don't rely so much on their competitors' recent land leasing decisions as signals of the shale gas reserve quality, and, in fact, tend to make opposite choices. The coefficient of the accumulated lease acreages in the given municipality is estimated to be 0.029 with a standard error of 0.001. The sign of the coefficient is as expected, though its value is much smaller than that of the coefficient of the own recent lease acreage, which makes sense as recent leasing decisions carry a lot more information concerning an operator's interest at a certain location. The coefficient of productivity is estimated to be 23.591 with a standard error of 0.924. It suggests that, conditioning on the past leasing decisions, a 1% increase in the daily production of the horizontal wells in the neighboring municipalities leads to a 23.591 acres increase in an operator's latent lease acreage in the given municipality. To get a sense of the magnitude of the estimate, the mean value of the monthly acreage an operator leases in a municipality is 2.15 acres, and the conditional mean given that the acreage is positive is 144.41 acres. The impact of productivity information on an operator's land leasing decision is thus huge. The coefficient of the interaction between productivity and the own recent lease acreage is estimated to be 0.007 with a standard error of 0.002. It implies that once the production information is available, a 1% increase in the average productivity generates an additional 0.007 acre to the latent lease acreages for every one acre increase in the operator's own recent lease. An explanation to the finding is that higher productivity leads to higher prior beliefs of the municipality's shale gas output potential, which makes the correlation of an operator's leasing decision over time also stronger. The coefficient estimate of the interaction between productivity and the peer's recent lease acreage is not significant. It reinforces the earlier finding that operators do not rely on their competitors' leasing decisions to make their choices. The coefficient of the interaction between productivity and the accumulated lease acreages is estimated to be -0.001 with a standard error of 0.0001, suggesting that higher productivity decreases the impact of the accumulated lease acreage on an operator's land leasing decision, though the magnitude is small. The coefficient of the royalty rate is estimated to be 631.163 with a standard error of 49.983. It implies that, other things equal, a 1% absolute increase in the average royalty rate of a municipality leads to a 6.31 acres increase in an operator's latent lease acreage. The positive sign of the estimate clearly reflects that the royalty rate variable is endogenously determined, which will lead to biases in the estimates of other coefficients. To address this issue, an alternative model is proposed in the later part of the section. The coefficient of the natural gas price is estimated to be 25.163 with a standard error of 1.116, consistent with the strong correlation between natural gas price and leased acreages displayed by Figure **??**. The public firm indicator variable's coefficient is estimated to be 173.033 with a standard error of 16.054, which implies that, compared to private operators, the latent lease acreage of a public operator is 173.033 acres larger on average. Finally, the coefficient of the time trend variable is estimated to be -3.904 with a standard error of 0.128, echoing a downward trend of lease acreages observed in Table **??**.

The fourth and fifth columns of the Table **??** present the estimation results of the same model with the operator fixed effects removed. The coefficient estimates and the standard errors are not significantly different from those of the baseline model estimation. The last two columns are the estimation results of a linear model, where the dependent variable is the observed lease acreage $a_{i,j,t}$ and the independent variables are the same as the baseline model. Compared to the first two models, the coefficient estimates of the linear model become mush smaller in absolute values, although the signs of the estimates remain mostly unchanged. One reason that the linear model may not be a good choice in the current setting is that the dependent variable is by nature non-negative, and a large share of the observed dependent variables are zeros. The nonlinear latent variable model proposed in the last section captures this non-negative feature.

The remaining part of the section discusses a few robustness tests of the estimation results of the last model. The first test re-performs the estimation based on different window widths. The purpose is to examine the possibility that the estimates are significant due to noises at the sample's tails. Table **??** presents the outcomes of the test. The operator fixed effects are dropped due to the shrinkage of the sample size. The removal of the operator fixed effects is not of a major concern given the similarity of the estimation results with and without the fixed effects in the baseline model. The second and the third columns are the coefficient estimates and the standard errors based on the full sample from 2007 to 2014, which are the same as the fourth and the fifth column in Table **??**. The next two columns in Table **??** give the estimates using the sample with a narrower window, with acreage observations from 2007 and 2014 removed. The sixth and the seventh columns are the estimates based on an even narrower window from 2009 to 2012, and the last two columns are estimated from the sample consisting only of the observations in 2010 and 2011. The four sets of estimation generate comparable outcomes, both in terms of the signs and the significance levels of the coefficient estimates. One exception is the coefficient of the interaction between average productivity and own recent lease acreage, estimated using the sample from 2008 to 2013, which is negative and less significant. However, with samples of narrower windows, the coefficient estimates are positive and significant again, and are greater than the original estimate. Another exception is the last estimate of the natural gas price's coefficient, which becomes negative and insignificant. As the sample's window turns narrower, the magnitudes of some of the coefficient estimates change monotonically. For example, the coefficient estimate of own recent lease acreage drops from 0.660 in the first estimation to 0.553 in the last estimation, whereas the coefficient of peer's recent lease acreage goes down from -0.015 in the first estimation to -0.139 in the last estimation. The most notable changes include the coefficient estimate of productivity, from 24.899 in the first estimation to 12.587 in the last estimation, and the coefficient estimate of royalty rate, from 651.460 in the first estimation to 3308.702 in the last estimation. As the changes are monotonic and do not flip the signs of the original estimates, the estimation results are robust.

The second robustness test conducts the same estimation using an artificially created sample. The sample is constructed by retaining only the acreage observations from 2011 to 2014 in the original sample, and changing the productivity observations in 2011 and 2012 to zeros. It intends to examine the virtual effect of the production information on operators' land leasing decisions had it been disclosed at the end of 2012. Table **??** presents the estimation results. The operator fixed effects are not included. The second and third columns are the coefficient estimates and the standard errors based on the original sample. The fourth and fifth columns are the estimation results based on the artificial sample. The first major change is that the coefficient estimate of peer's recent lease acreage becomes insignificant. The second major change is that the coefficient estimate of the interaction between average productivity and own recent lease acreage becomes negative and insignificant. In addition, the coefficient estimate of the natural gas price drops from 24.988 to

0.737 and becomes insignificant. The estimation results based on the artificial sample contradicts to the intuition that a higher productivity increases the correlation between an operator's leasing decisions overtime. Thus there is not sufficient support that the production information, had it been disclosed at the end of 2012, introduces shocks to the shale gas operator's land leasing decisions. The finding is consistent with the fact that most of the changes in the land leasing patterns ought to be realized in the year following the real disclosure of the production information.

Finally, to address the endogeneity problem of the royalty rate variable, the following log-linear model is proposed, where the acreage variables are all transformed to their logarithm forms:[7]

$$\log a_{i,j,t} = \alpha_i + \log \boldsymbol{S}_{i,j,t}\boldsymbol{\delta} + \theta\overline{\log q}_{j,t} + \left(\overline{\log q}_{j,t} \times \log \boldsymbol{S}_{i,j,t}\right)\boldsymbol{\gamma} + \boldsymbol{X}_{i,j,t}\boldsymbol{\beta} + \varepsilon_{i,j,t} \tag{11}$$

where $a_{i,j,t}$ is the lease acreages taken by operator i in municipality j and month t, $\log \boldsymbol{S}_{i,j,t} = \left(\log A^{Recent}_{i,j,t}, \log A^{Recent}_{-i,j,t}, \log A^{Accumlated}_{j,t}\right)$ stands for the vector of past lease acreages in log forms. The per capita income of the county where the municipality locates in the same year the lease is granted is used as an instrument for the royalty rate. The logic is that the per capita income represents the opportunity cost of the landowner's bargaining time and is hence correlated with the cost of acquiring leases. Table **??** reports the estimation results. First note that the coefficient of royalty rate turns negative when IV is used, which makes sense as the variable stands for the "price" of leases. The coefficients of own recent acreage, accumulated acreage and productivity remain positive and significant, whereas the coefficient of peer's recent acreage now becomes positive and significant. The coefficients of the interactive variables are all small, and the interaction between productivity and own recent acreage becomes insignificant. A somewhat surprising result is that natural gas price now is also insignificant. Compared with the estimation results without IV, apart from the sign change of the coefficient of royalty rate, another finding is that the coefficient of productivity is highly underestimated when IV is not used, suggesting an even stronger impact of the production information on leasing decisions than what is earlier estimated.

VII Theoretical Analysis

This section introduces a simple Bayesian learning model that seeks to explain how additional production information helps improve the efficiency of lease investment and justify the empirical findings presented in the last section. The framework follows the herding model described in Bikhchandani, Hirshleifer, and Welch (1992) and abstracts from the bargaining problem between operators and landowners by assuming a reduced-form upward sloping supply function. In the setting elaborated below, operators perform Bayesian updating about the potential outputs of two locations before making their lease investment decisions. The additional production information is assumed to affect the firms' prior beliefs. The model aims to explore the relationship between the production information and the expected numbers of firms taking leases at each of the two locations.

Assume N operators are sequentially choosing to allocate their lease investment between two

[7] $\log(a_{i,j,t} + 1)$ is used in practice to accommodate the zero acreage cases.

locations, L_1 and L_2. Let Q_1 and Q_2 denote the potential output of the of two locations. Assume $Q_1 = Q^H$, $Q_2 = Q^L$, and $Q^H > Q^L$. The true potentials are unknown to the operators. Instead, they have identical prior beliefs $\Pr(Q_1 = Q^H) = \Pr(Q_2 = Q^H) = 1/2$.

Each operator i, $i = 1, 2, \ldots, N$, chooses one location to invest, possibly due to budget concerns. Let $d_{i,j} \in \{0, 1\}$ denote operator i's investment decision at L_j, $j = 1, 2$, where 1 stands for investing and 0 stands for not investing. Then $d_{i,1} + d_{i,2} = 1$. For a given order of investment $1, 2, \ldots, N$, let $d_{0,j} = 0$ for both $j = 1, 2$, and $\mathbf{h}_{i,j} = (d_{0,j}, d_{1,j}, d_{2,j}, \ldots, d_{i-1,j})$ be the history of investment decisions at location L_j prior to operator i. Before investor i makes its investment decision, it observes $\mathbf{h}_i = (\mathbf{h}_{i,1}, \mathbf{h}_{i,2})$ and a signal $s_i \in \{1, 2\}$. It is known to all the operators that $\Pr(s_i = j)$ $= p > 1/2$ if $Q_j = Q^H$, $j = 1, 2$. Define $r_j(\cdot)$ to be the inverse supply function of lease resources at L_j that relates the royalty rate at L_j with the investment history at L_j. Assume $r_j(\cdot)$ is increasing, $r_j(0) = 0$, and $r_j(N) \leq 1$, $j = 1, 2$. Define

$$|\mathbf{h}_{i,j}| \equiv \sum_{k=0}^{i-1} d_{k,j}, j = 1, 2, i = 1, 2, \ldots, N \tag{12}$$

to be the number of operators that have invested at L_j before operator i. Operator i's return from L_j is given by

$$R_{i,j} = (1 - r_j(|\mathbf{h}_{i,j}|))Q_j \tag{13}$$

The return is increasing in the potential output and decreasing in the royalty rate. The natural gas price is normalized to one, and there is assumed to be no fixed investment cost.

Let $\mathbf{d}_i = (d_{i,1}, d_{i,2})$ denote operator i's lease investment decision. Operator i conducts Bayesian updating whenever possible, and chooses $\mathbf{d}_i$ to maximize its posterior expected return. Let $J_i(\mathbf{h}_i, \mathbf{d}_i)$ be the set of signal s_i so that operator i chooses $\mathbf{d}_i$ given that it observes investment history $\mathbf{h}_i$. Operator i solves the following objective function

$$\max_{\mathbf{d}_i} E\left[R_{i,1} + R_{i,2} | s_i, J_k(\mathbf{h}_k, \mathbf{d}_k), k < i\right] \tag{14}$$

Let N_1 and N_2 denote the expected numbers of firms choosing L_1 and L_2.

Now assume there exists an additional set of productivity signals q_1 and q_2, $q_1 > q_2$, so that all the operators' prior beliefs becomes

$$\Pr(Q_1 = Q^H) = q_1/(q_1 + q_2) \tag{15}$$

$$\Pr(Q_2 = Q^H) = q_2/(q_1 + q_2) \tag{16}$$

For the same given order of investment and inverse supply functions $r_1(\cdot)$, $r_2(\cdot)$, let $M_1(q_1/q_2)$ and $M_2(q_1/q_2)$ denote the expected numbers of firms choosing L_1 and L_2 with signals q_1 and q_2 available. The question of interest is how $M_1(q_1/q_2)$ and $M_2(q_1/q_2)$ compare with N_1 and N_2. If $M_1(q_1/q_2) > N_1$ and $M_2(q_1/q_2) < N_2$ when $q_1 > q_2$, the introduction of the productivity signal improves the efficiency of resource allocation. To begin with, the following example seeks to illustrate the intuition of the model by focusing on the special case when $N = 2$. For simplicity, assume $Q^H = 1$, $Q^L = 0$, $r_1(\cdot) = r_2(\cdot) = r(\cdot)$.

When $N = 2$, given that the operators have even priors, the first operator always chooses L_{s_1} because the posterior probability $\Pr(Q_{s_1} = 1) = p > 1/2$. The second operator, observing the first operator's choice, is able to infer s_1. If $s_2 \neq s_1$, the second operator again has identical posterior beliefs $\Pr(Q_{s_1} = 1)$ and $\Pr(Q_{s_2} = 1)$. Because $r(\cdot)$ is increasing, the second operator will choose L_{s_2}. If $s_2 = s_1$, the second operator's posterior belief that $Q_{s_1} = 1$ becomes

$$\Pr(Q_{s_1} = 1 | s_1 = s_2) = \frac{p^2}{p^2 + (1-p)^2} \tag{17}$$

Let L_{3-s_1} denote the location other than L_{s_1}. The second operator's expected return from choosing each of the two locations, given its posterior beliefs, are

$$E[R_{2,s_1} | s_2 = s_1] = (1 - r(1)) \frac{p^2}{p^2 + (1-p)^2} \tag{18}$$

$$E[R_{2,(3-s_1)} | s_2 = s_1] = \frac{(1-p)^2}{p^2 + (1-p)^2} \tag{19}$$

When $r(1) > 1 - (1-p)^2/p^2$, $E[R_{2,(3-s_1)} | s_2 = s_1] > E[R_{2,s_1} | s_2 = s_1]$. Hence even if $s_2 = s_1$, the second operator still chooses L_{3-s_1} due to the high royalty cost at L_{s_1}. It can be seen that given $r(1) > 1 - (1-p)^2/p^2$, the two operators always choose different locations. Therefore, $N_1 = N_2 = 1$.

Now assume that the two operators observe the productivity signals q_1 and q_2 with $q_1 > q_2$, so that their priors become $\Pr(Q_1 = 1) = q_1/(q_1 + q_2)$. If the first operator observes $s_1 = 1$, it will choose L_1 for sure. If it observes $s_1 = 2$, its posterior belief that $Q_1 = 1$ becomes

$$\Pr(Q_1 = 1 | s_1 = 2) = \frac{(1-p)q_1}{(1-p)q_1 + pq_2} \tag{20}$$

Hence when $q_1/q_2 > p/(1-p)$, the first operator will choose L_1 regardless of its signal. Next, if $s_2 = 1$, the second operator's posterior belief that $Q_1 = 1$ becomes

$$\Pr(Q_1 = 1 | s_2 = 1) = \frac{pq_1}{pq_1 + (1-p)q_2} \tag{21}$$

and its expected return from the two locations, given its posterior beliefs, are

$$E[R_{2,1} | s_2 = 1] = (1 - r(1)) \frac{pq_1}{pq_1 + (1-p)q_2} \tag{22}$$

$$E[R_{2,2} | s_2 = 1] = \frac{(1-p)q_2}{pq_1 + (1-p)q_2} \tag{23}$$

As long as

$$r(1) < 1 - \frac{(1-p)q_2}{pq_1} \tag{24}$$

the second operator will choose L_1, and there is a positive probability that both operators will end up at L_1. Since $q_1/q_2 > p/(1-p)$, $1 - \frac{(1-p)q_2}{pq_1} > 1 - (1-p)^2/p^2$. By setting

$$1 - (1-p)^2/p^2 < r(1) < 1 - \frac{(1-p)q_2}{pq_1} \tag{25}$$

it happens that $M_1 > N_1$ and $M_2 < N_2$, so that on average, there will more operators choosing the location will higher potential, and on average, the royalty rate at L_1 will also be higher. Another thing to note is that when $r(1) < 1-(1-p)^2/p^2$, without the productivity signal, there is a positive probability that both operators will choose L_2, which is the most inefficient outcome. On the other hand, with the productivity signal available, as long as $q_1/q_2 > p/(1-p)$, at least one operator will choose L_1, and the worst scenario is avoided.

An analytical investigation of the Bayesian updating process becomes less tractable when $N > 2$ given the existence of the royalty cost functions, and simulation is used to explore the relationship between the prior beliefs and the expected number of operators choosing each location. Figure **??** displays the simulation results based on different royalty cost functions. The potential output Q_1 for location L_1 is set to be 1, and the potential output Q_2 for location L_2 is set to be 0. The common belief that a signal points to location L_j given $Q_j = 1$ is taken to be 3/5 for both $j = 1, 2$, and the number of investing operators N is set to be 50. Given a royalty cost function, a sample of 10,000 experiments are simulated for each prior belief $\Pr(Q_1 = 1)$. In each experiment, a set of 50 signals s_i, $i = 1, 2, \ldots, 50$, are independently drawn. s_i takes value 1 with probability 3/5 and value 2 with probability 2/5. The operators, not knowing the true potentials of the two locations, conduct Bayesian learning based on the investment histories and the signals they observe, and choose either L_1 or L_2 by solving Equation (14). The number of operators choosing each location is recorded for every experiment. Figure **??** and Figure **??** plot the mean numbers of operators and the mean royalty rates at L_1 and L_2 against different values of the prior belief $\Pr(Q_1 = 1)$ for the linear royalty cost function $r(x) = x/N$, where $N = 50$. Figure **??** and Figure **??** are the same plots for the concave royalty cost function $r(x) = 1 - 1/(x + 1)$, while Figure **??** and Figure **??** are based on the convex royalty cost function $r(x) = x^2/N^2$, where $N = 50$. In all the three sets of plots, the mean number of operators choosing $L - 1$ is increasing in the prior belief $\Pr(Q_1 = 1)$, so is the mean royalty rate at L_1. The simulation results show that in a Bayesian learning setting, an informative productivity signal that correctly adjusts the operators' prior beliefs improves the efficiency of the operators' resource allocation.

The simulation results generate several reduced-form implications that provide justifications for the earlier empirical findings. 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 the indicator variable of operator i's location choice, where $Y_i = 1$ if operator i chooses L_1 and $Y_i = 0$ if operator i chooses L_2. Let X_i be the indicator variable of operator i's signal, where $X_i = 1$ if $s_i = 1$ and $X_i = 0$ if $s_i = 2$. Consider the following probit model of the operator's location choice as a function of the operator's signal:

$$Y_i^* = \beta_0 + \beta_1 X_i + \varepsilon_i \tag{26}$$

and

$$Y_i = \begin{cases} 1, & Y_i^* > 0 \\ 0, & \text{otherwise} \end{cases} \tag{27}$$

Figure **??** plots the estimation results of β_1 in Equation (26) against different values of the prior belief $\Pr(Q_1 = 1)$ ranging from 0.1 to 0.9[8]. The coefficient estimate for each prior belief is obtained using the simulated sample of size 10,000 based on the linear royalty cost function $r(x) = x/N$, where $N = 50$. 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 displays a generally increasing pattern when the prior belief becomes large. The finding suggests that, when prior beliefs are very pessimistic, the disclosure of additional positive information may diminish the correlation between an operator's location choice and the signal it observes. By contrast, when prior beliefs are more optimistic, the disclosure of additional positive information tends to amplify this correlation.

Another implication of the model is that the correlation between an operator's location choice and the investment history it observes also changes along with the prior belief. Let $|\mathbf{h}_{i,1}|$ be the number of operators that has invested at location L_1, Y_i be the same location choice indicator variable defined as above. Consider the following probit model:

$$Y_i^* = \beta_0 + \beta_1|\mathbf{h}_{i,1}| + \beta_2 i + \varepsilon_i \tag{28}$$

and

$$Y_i = \begin{cases} 1, & Y_i^* > 0 \\ 0, & \text{otherwise} \end{cases} \tag{29}$$

Figure **??** plots the coefficient estimates in Equations (28) and (29) against different values of the prior belief $\Pr(Q_1 = 1)$, using the same simulation results as before. The royalty cost function is still set to be linear with $r(x) = x/N$, where $N = 50$. Figure **??** plots the estimates of the coefficient of the number of investors choosing location L_1 against prior beliefs ranging from 0.1 to 0.9. The coefficient estimates are positive for all values of the prior belief, suggesting that an operator's location choice is positively correlated with the investment history at the location. An interesting feature is the U-shape of the curve: the coefficient first decreases when the prior belief is less than 0.4, and stays relatively flat in the range between 0.4 and 0.6, and then begins to climb up when the prior belief exceeds 0.6. The finding suggests that, when prior beliefs are pessimistic, additional positive information tends to lower the correlation between an operator's location choice and the location's accumulated investment history. When prior beliefs are in the intermediate range, the correlation does not change much following the disclosure, and when prior beliefs are optimistic, additional positive information turns to enhance the correlation. Finally, Figure **??** plots the estimates for the coefficient of an operator's position in the queue against its

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

prior beliefs. The coefficient estimate is generally decreasing in the prior belief, and changes from positive to negative when the prior belief increases. The interpretation is that latecomers tend to avoid promising locations due to higher costs, and this avoidance is magnified by higher prior beliefs of the location's output potential.

VIII Conclusion

This article looks into the the classic topic of information disclosure and information sharing in the context of shale gas exploration. It provides empirical evidence showing that in the early stages of an industry where exploration and experimentation is essential, mandatory information sharing is likely to improve the efficiency of resource allocation and industry wide production outcomes. The article also provides a theoretical framework seeking to explain the mechanism through which information disclosure affects firms' decisions while at the same time generating predictions that can be empirically verified. The findings of the article provide justifications for the regulatory agency to collect and disclose production information that otherwise cannot be shared spontaneously among firms.

Two questions that are fundamental to the shale gas exploration industry are not addressed in the current research. The first question is how royalty rates are determined. The theoretical model proposed by this article assumes a reduced form upward sloping supply curve. In practice, royalty rates and bonus payments are outcomes of the bargaining process between operators and landowners. An empirical bargaining model is needed to determine the equilibrium royalty rates before full-fledged structural estimation can be conducted. The second question is whether strategic concerns arise in a dynamic framework. Theoretically, when information sharing is not one shot but persistent or periodic, firms will have incentives to strategically pause their decision making in return for extra information. These two questions are important in that they are directly related to the measurement of the welfare consequence of the production information disclosure and how utility gains or losses are distributed between operators and landowners. Answer these questions are left for further research.

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.

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.

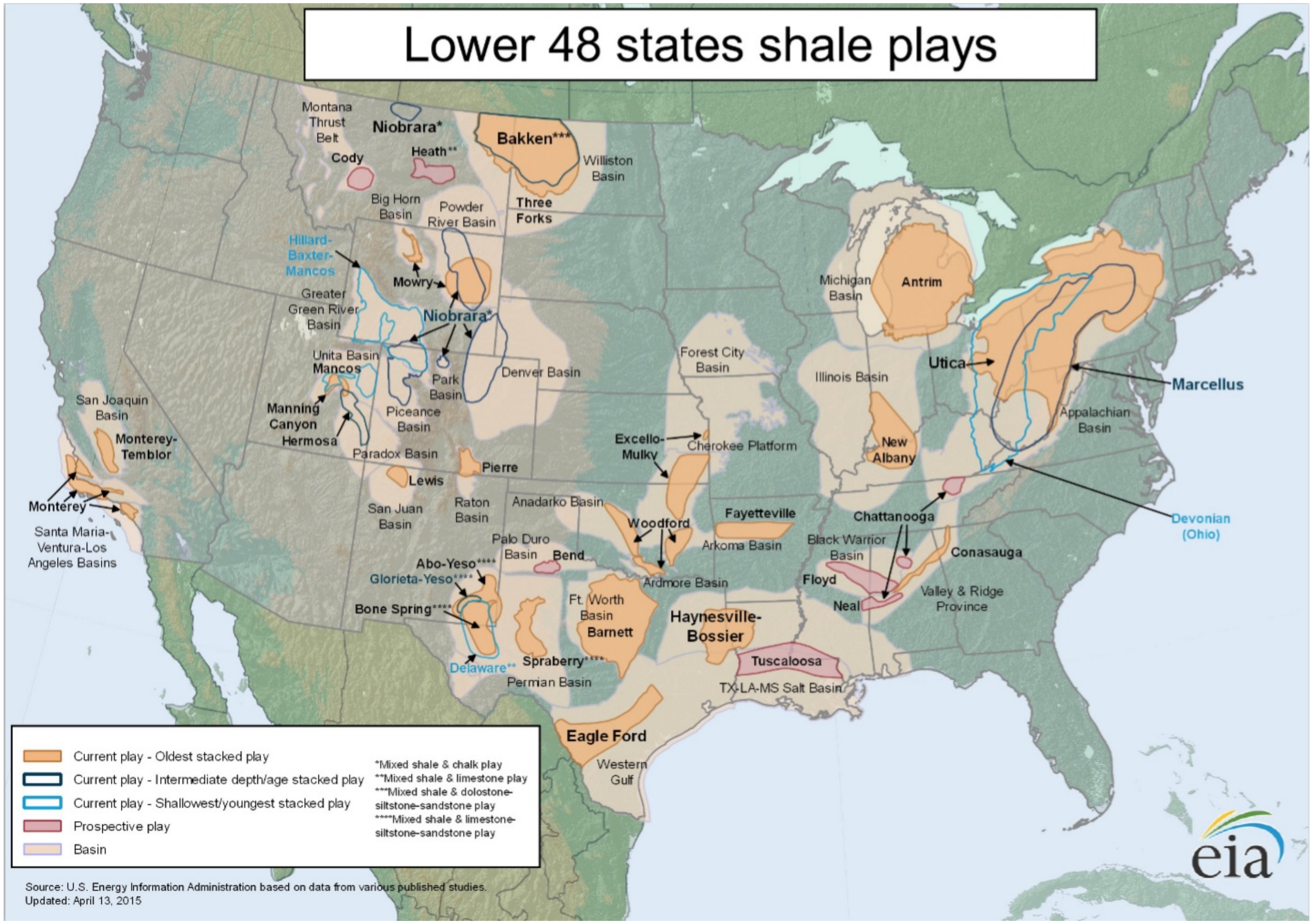
Steck, Andrew L (2018). “Industry dynamics with social learning: Evidence from hydraulic fracturing”. In: *Department of Economics*.

Wang, Zhongmin and Alan Krupnick (2015). “A Retrospective Review of Shale Gas Development in the United States: What Led to the Boom?” In: *Economics of Energy & Environmental Policy* 0(Number 1). URL: https://ideas.repec.org/a/aen/eeepjl/eeep4_1_krupnick.html.

Zagorski, William A, Gregory R Wrightstone, and Douglas C Bowman (2012). “The Appalachian Basin Marcellus gas play: Its history of development, geologic controls on production, and future potential as a world-class reservoir”. In.

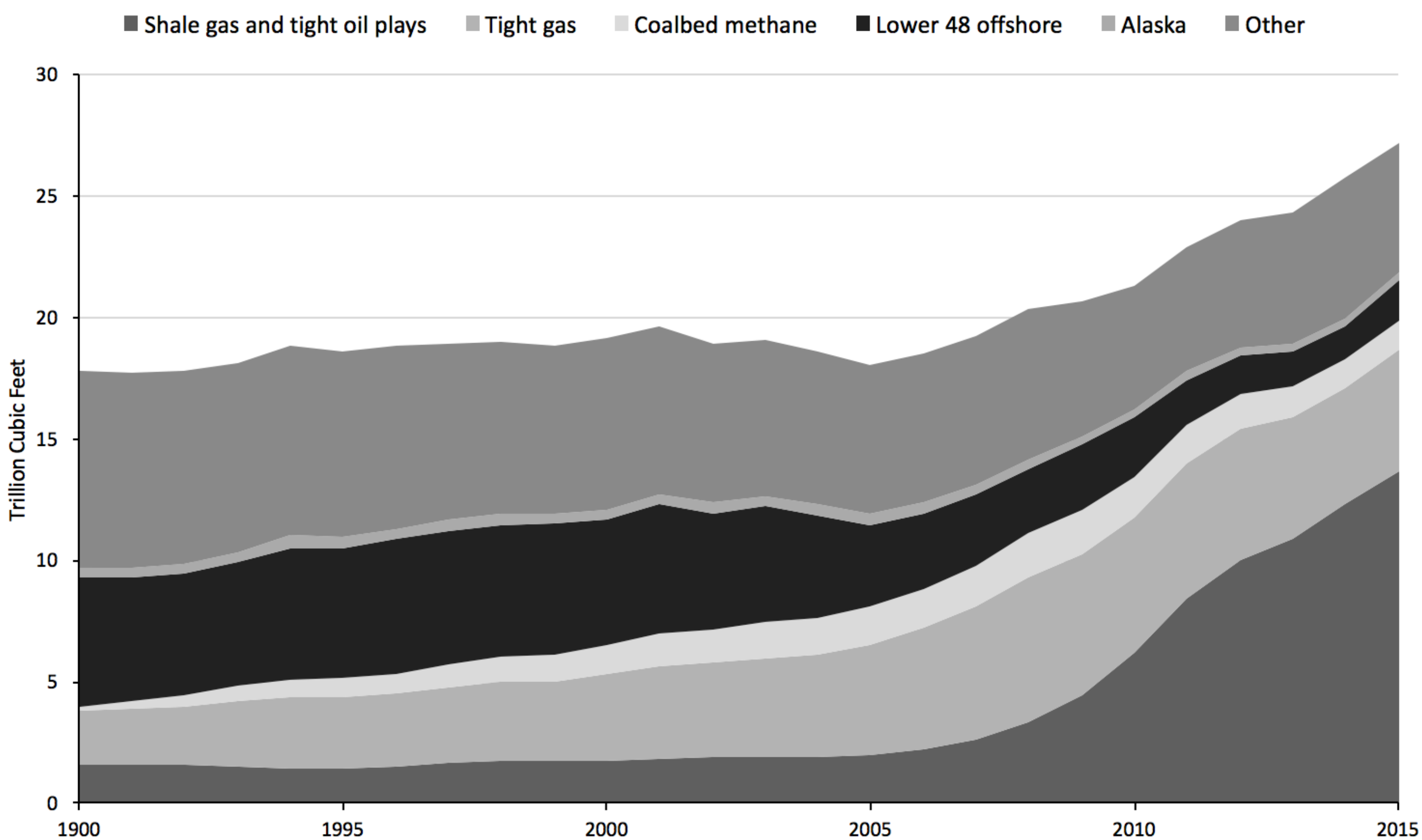
Zhiltsov, Sergey S. and A. V. Semenov (2017). “Shale Gas: History of Development”. In: *Shale Gas: Ecology, Politics, Economy*. Ed. by Sergey S. Zhiltsov. Springer International Publishing: Cham, pp. 9–16. ISBN: 978-3-319-50275-5. DOI: 10.1007/698_2016_49. URL: https://doi.org/10.1007/698_2016_49.

Figure 1: Major Shale Plays in the Lower 48 States



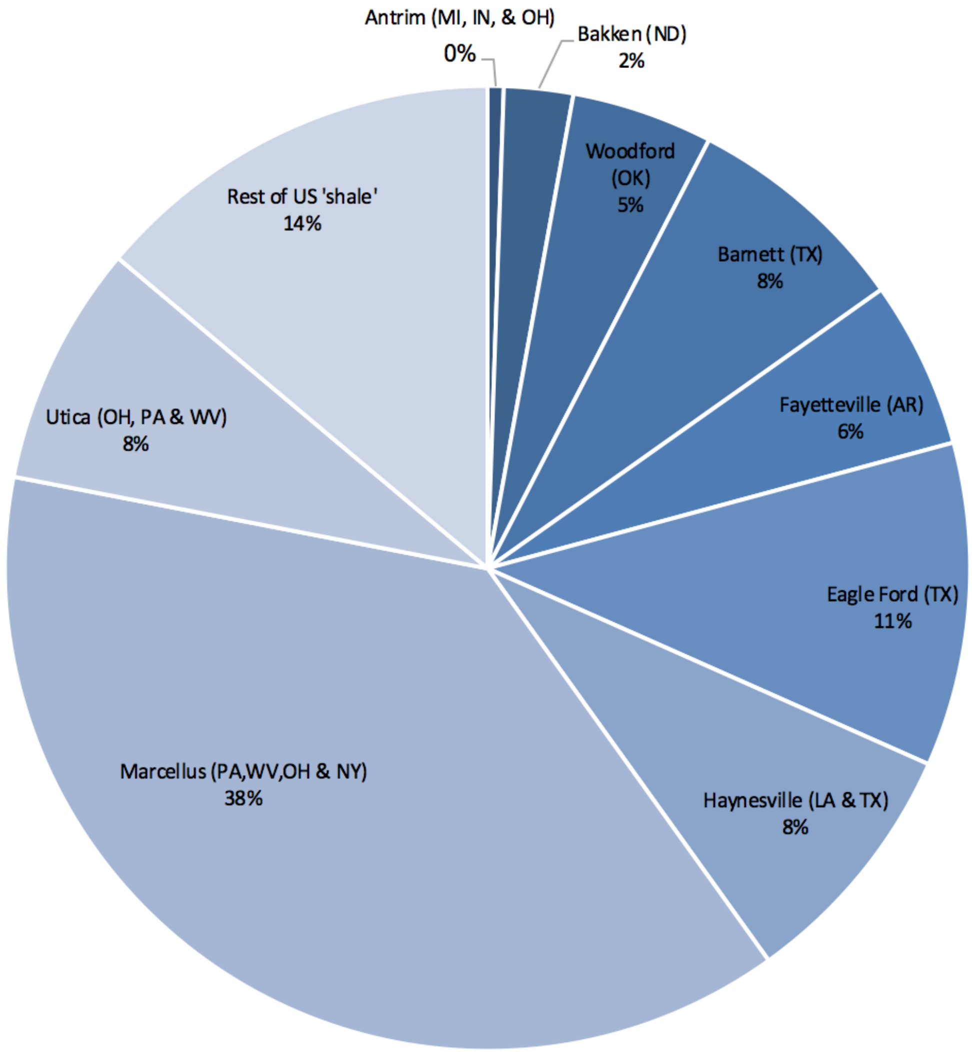
Source: EIA. Link: `http://www.eia.gov/oil_gas/rpd/shale_gas.pdf`

Figure 2: U.S. Dry Natural Gas Production by Sources



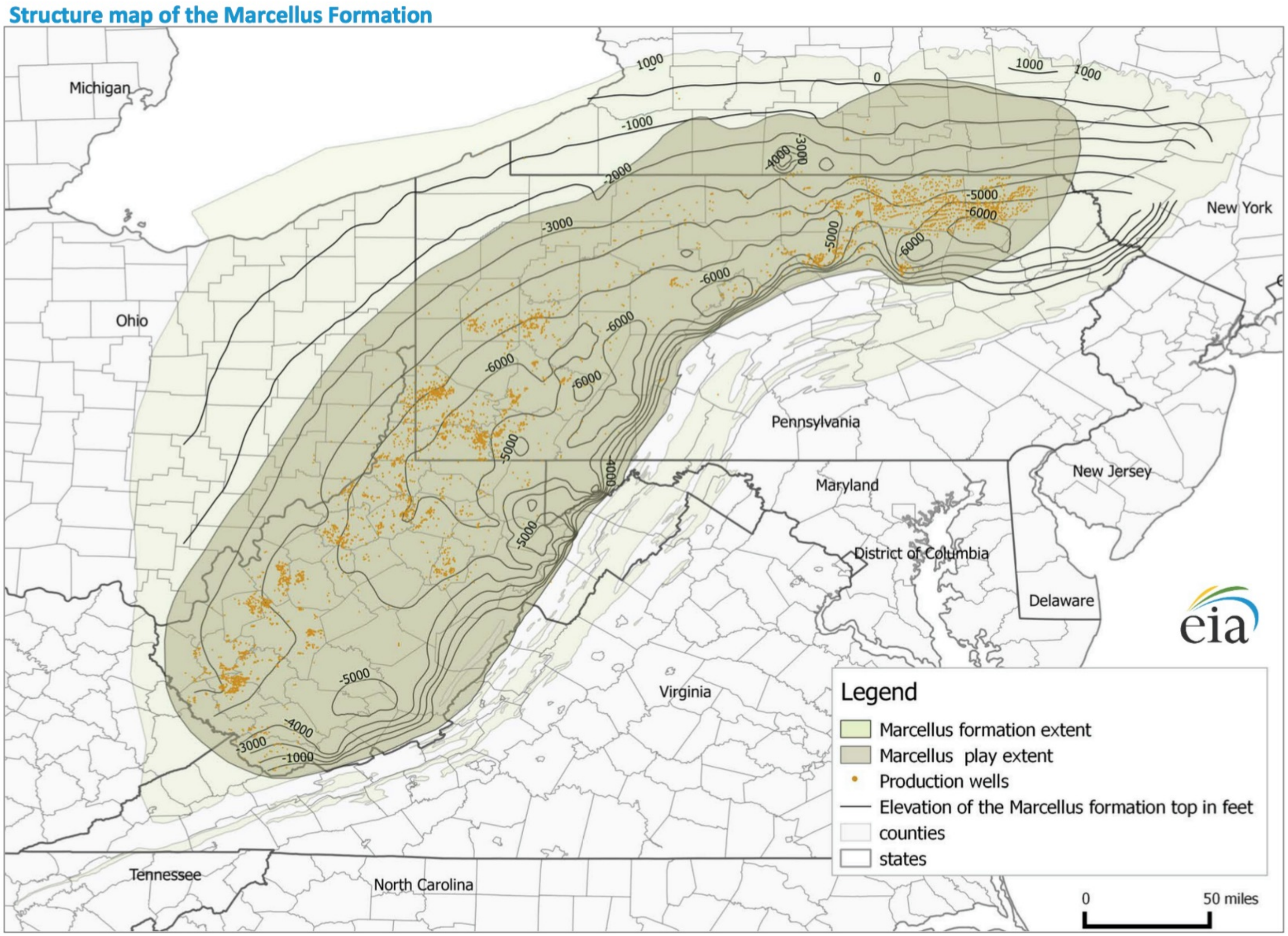
Note: the data is obtained from the *Annual Energy Outlook 2016* published by EIA, which can be viewed at `https://www.eia.gov/outlooks/archive/aeo16/MT_naturalgas.cfm#natgasprod_exp`. The regions, from below to above, are “Shale gas and tight oil plays”, “Tight gas”, “Coal bed methane”, “Lower 48 offshore”, “Alaska”, and “Other”, respectively. The horizontal axis measures the years. The vertical axis measure the production quantities in trillion cubic feet.

Figure 3: The Production of The Major Shale Gas Plays in January 2016



Note: the data is obtained from the *Natural Gas Weekly Update*, released by EIA on March 24, 2016, which can be viewed at `http://www.eia.gov/naturalgas/weekly/archive/2016/03_24/index.cfm`. The figure is based on the U.S. shale gas production in January 2016.

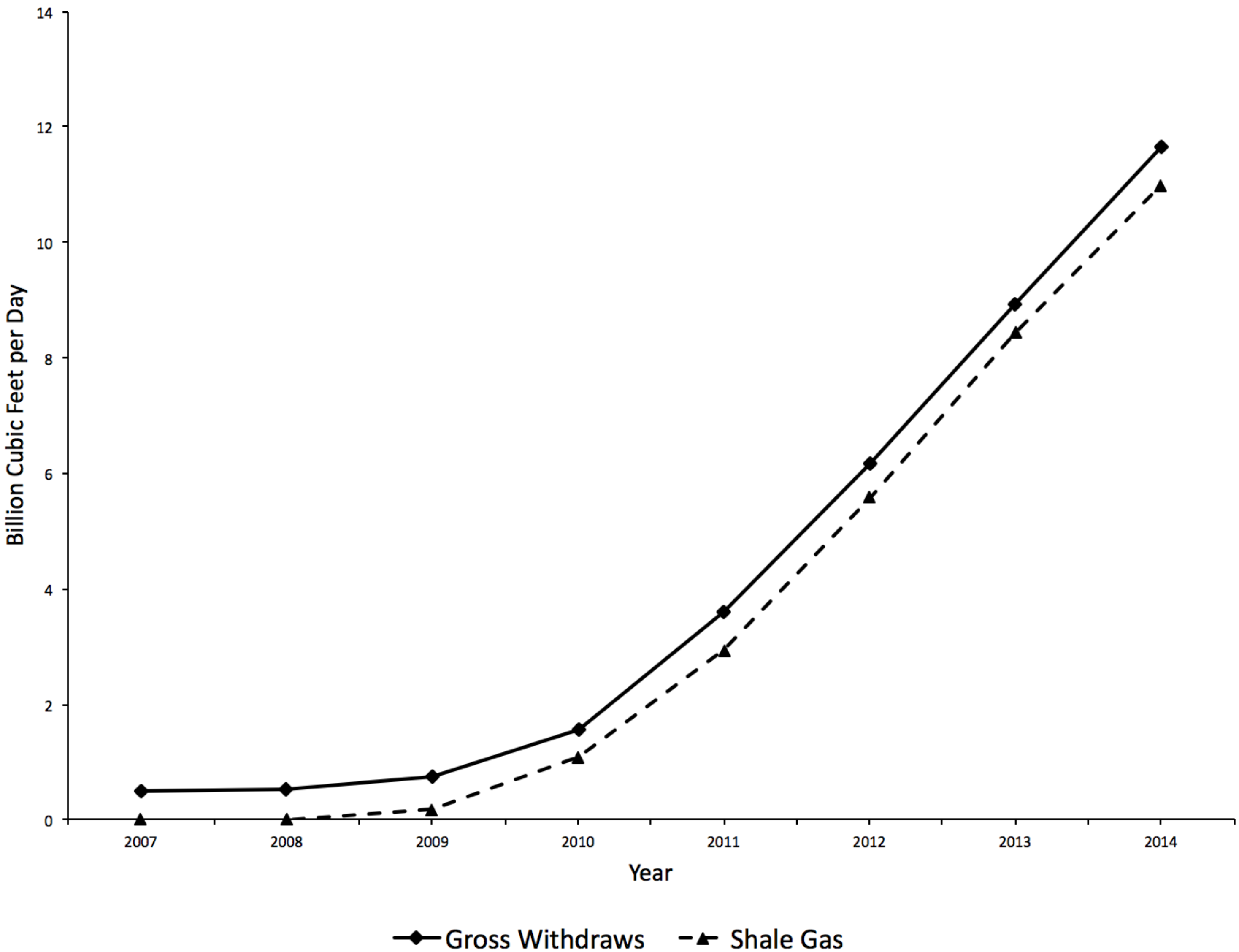
Figure 4: The Structure Map of the Marcellus Formation



Source: U.S. Energy Information Administration, based on DrillingInfo Inc., New York State Geological Survey, Ohio State Geological Survey, Pennsylvania Bureau of Topographic & Geologic Survey, West Virginia Geological & Economic Survey, and U.S. Geological Survey.
Note: Map includes production wells from January 2003 through December 2014.

Source: EIA. Link: `https://www.eia.gov/maps/pdf/marcellus-upd.pdf`

Figure 5: Daily Natural Gas Gross Withdraws and Shale Gas Production in Pennsylvania (2007 - 2014)



Note: the data is obtained from EIA. The links are `https://www.eia.gov/dnav/ng/hist/n9010pa2A.htm` and `http://www.eia.gov/dnav/ng/hist/res_epg0_r5302_spa_bcfa.htm`. The horizontal axis measures the years. The vertical axis measures the mean daily production quantities of each year. The unit is Billion cubic feet per day.

Figure 6: Comparison between Vertical and Horizontal Well Pads

(a) Vertical Well Pads

(b) Horizontal Well Pads

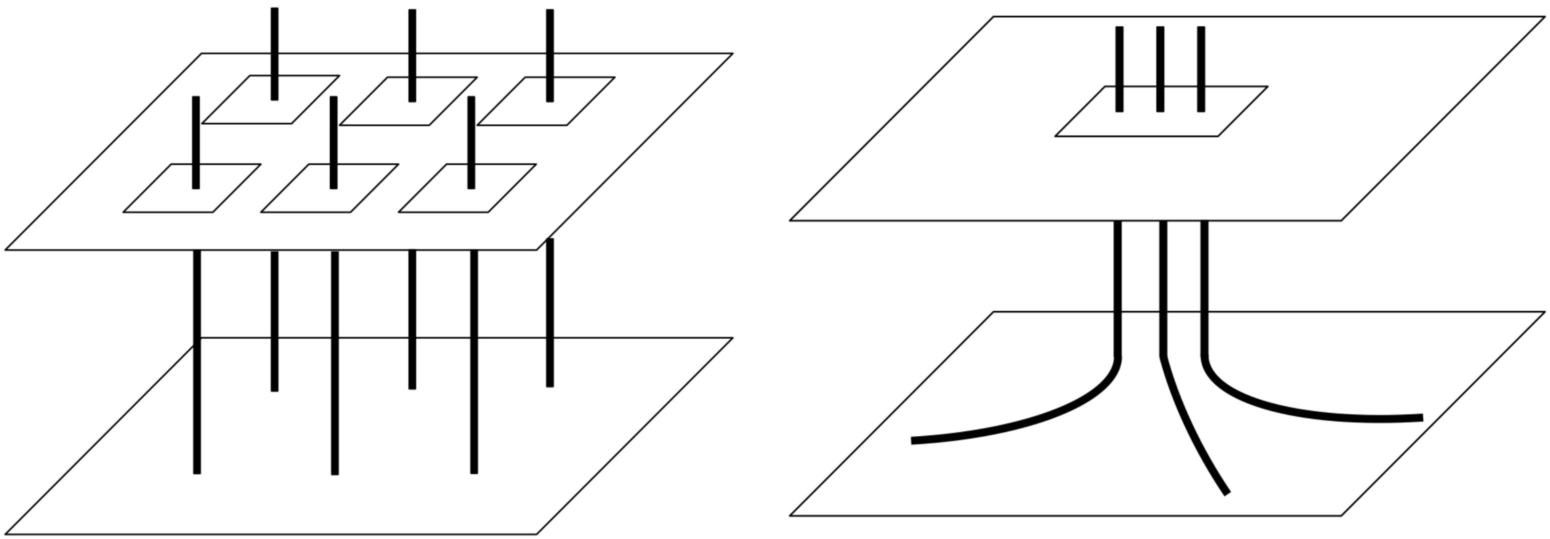
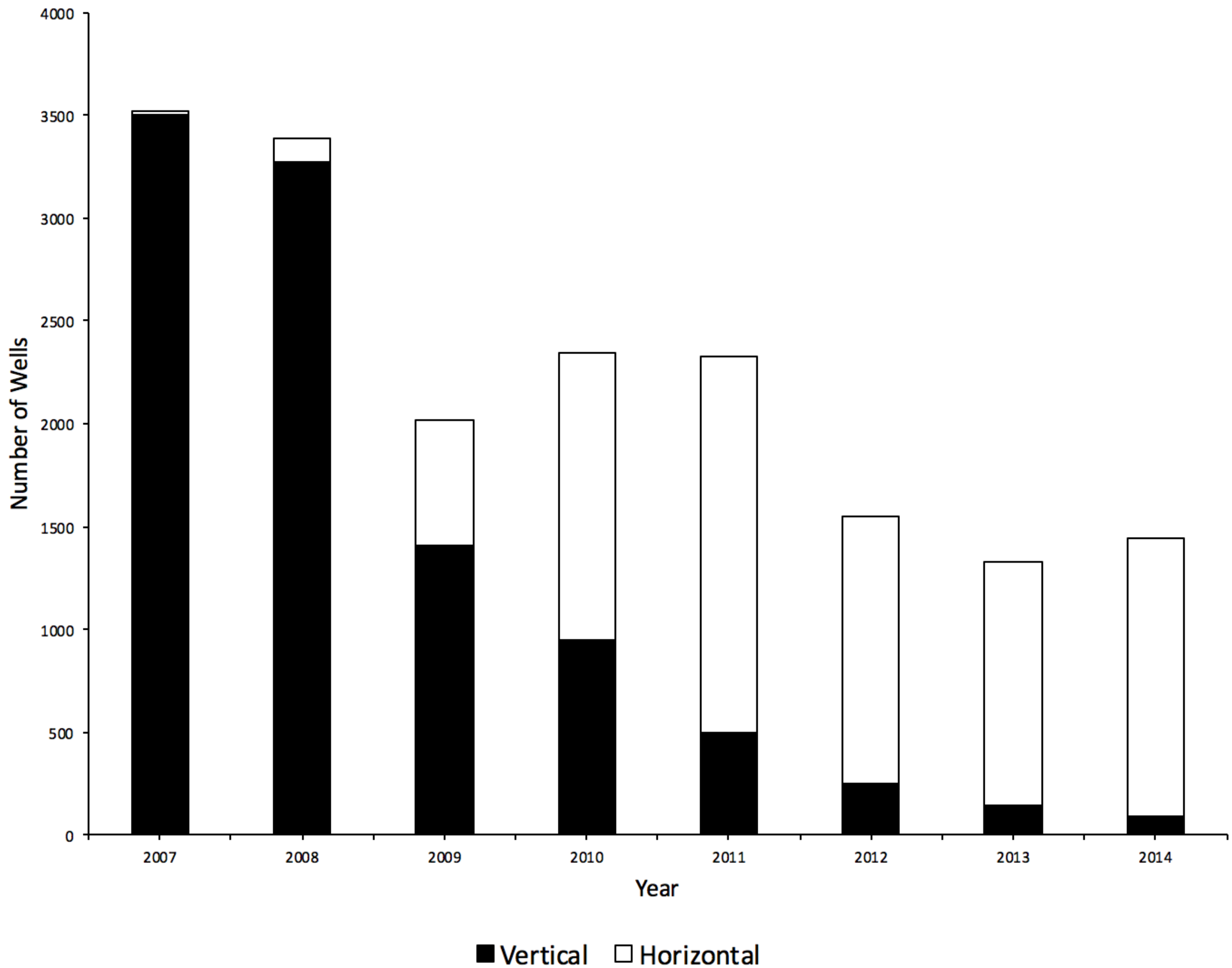
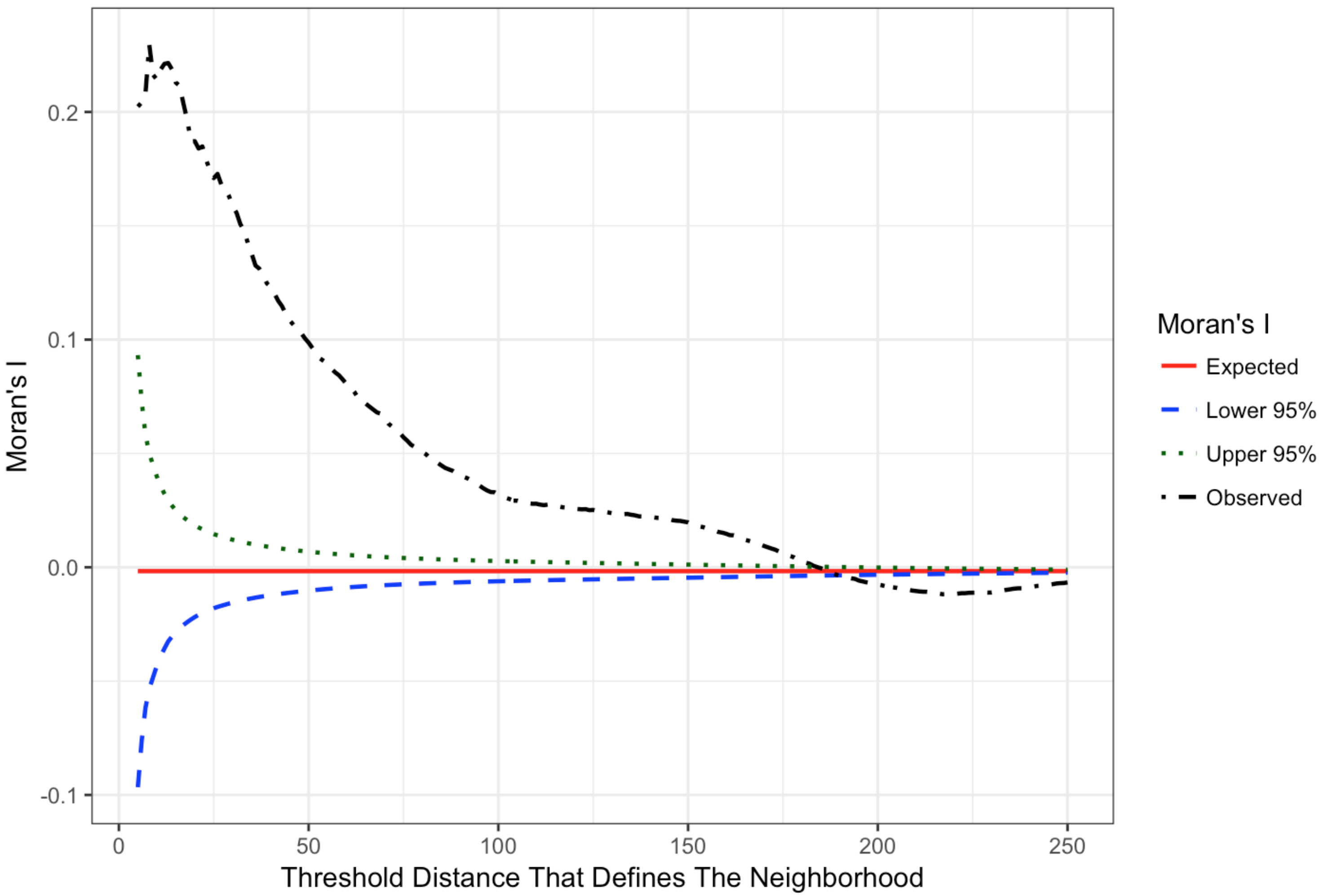


Figure 7: Annual Natural Gas Well Starts in Pennsylvania (2007 - 2014)



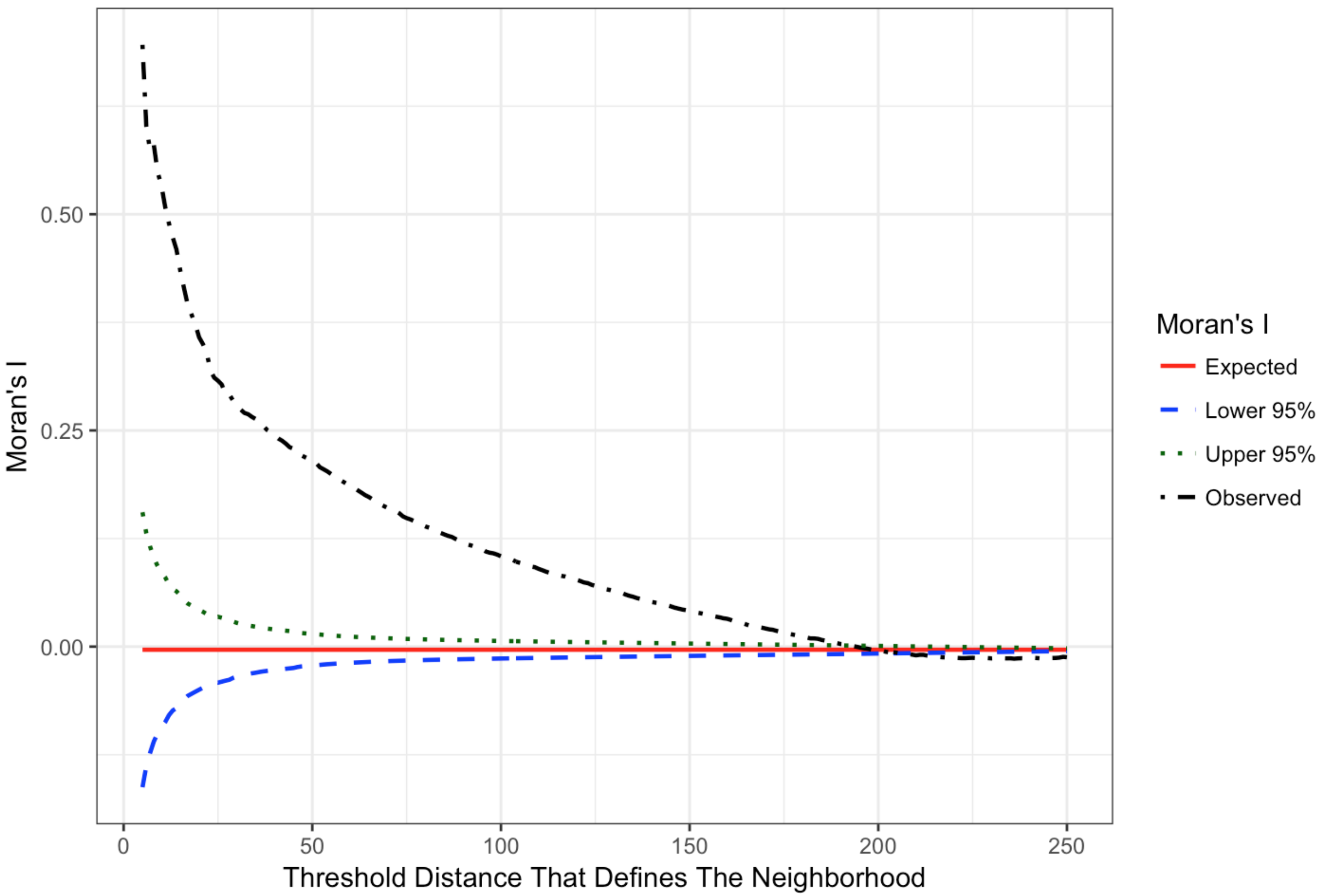
Note: the data is obtained from the PADEP. The link is `http://www.dep.pa.gov/DataandTools/Reports/Oil%20and%20Gas%20Reports/Pages/default.aspx#.Vye1useUpg0`. The figure measures the newly drilled wells in each year from 2007 to 2014.

Figure 8: 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.

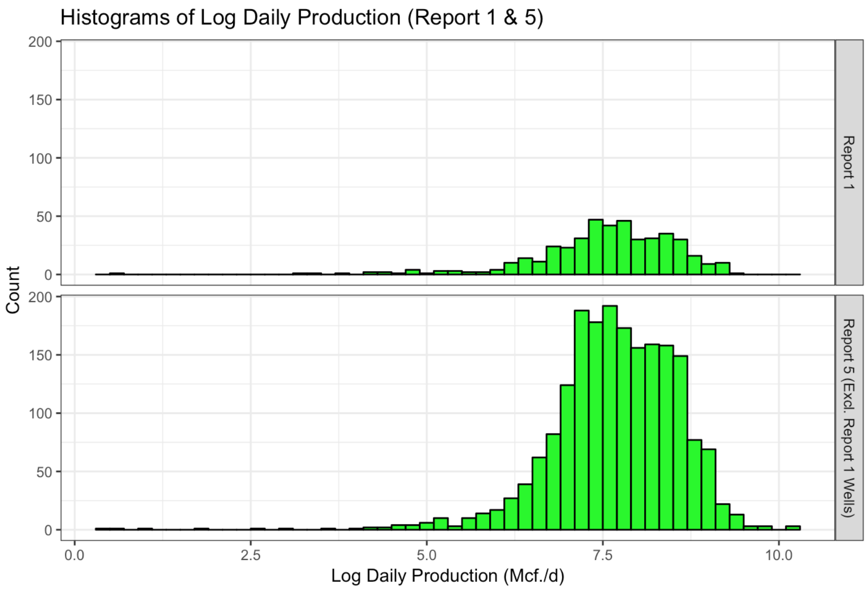
Figure 9: Moran's I for Municipality Level Productivity Based on Different Neighborhood Defining Distances



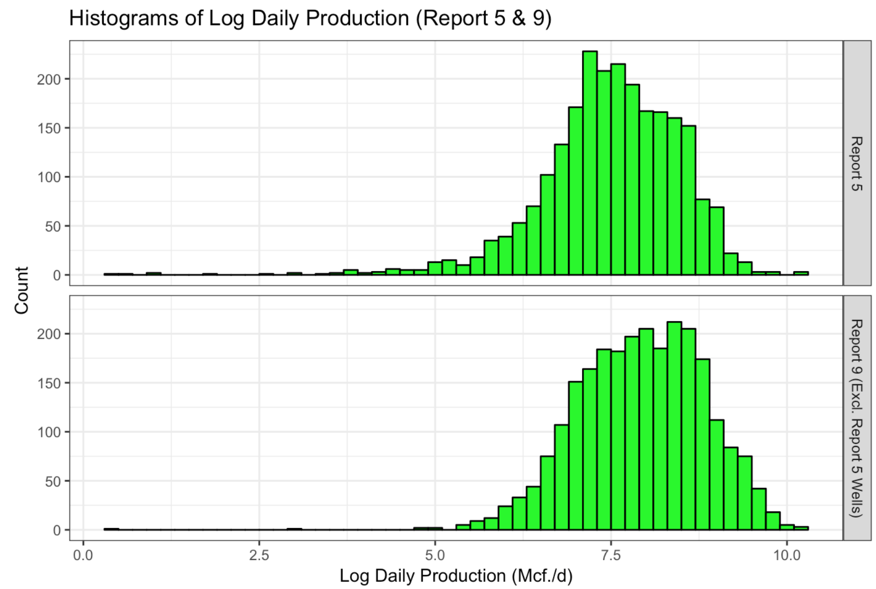
Note: the data is obtained from the PADEP. The solid curve stands for the expected Moran's I for the different values of neighborhood defining distance. The lower dashed curve and the upper dotted curve represent the 95% confidence intervals. The dotted-dashed curve at the top of the figure illustrates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

Figure 10: Horizontal Well Productivity Histogram of Production Report 1, Report 5, and Report 9

(a) Productivity Histogram of Report 1 and Report 5

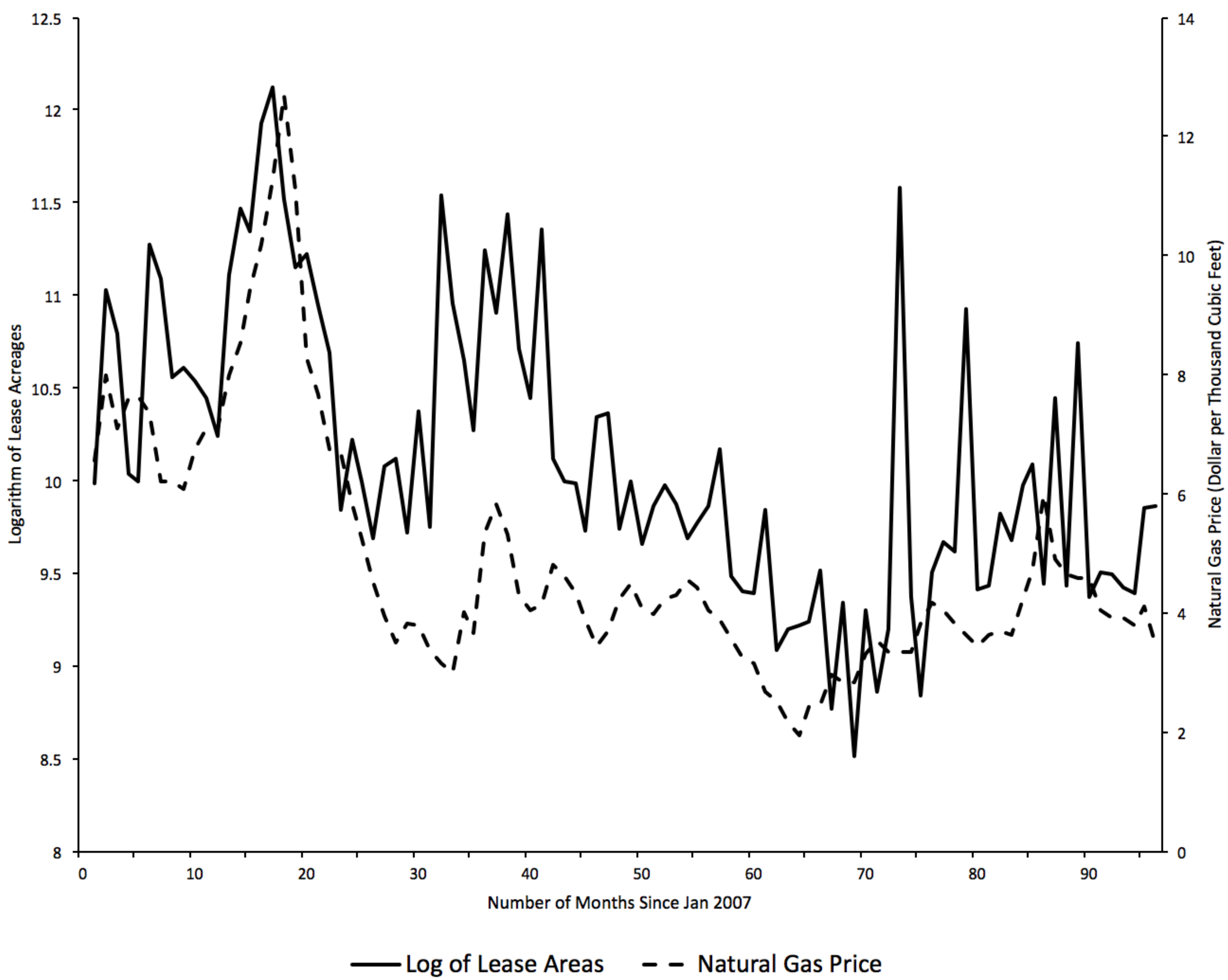


(b) Productivity Histogram of Report 5 and Report 9



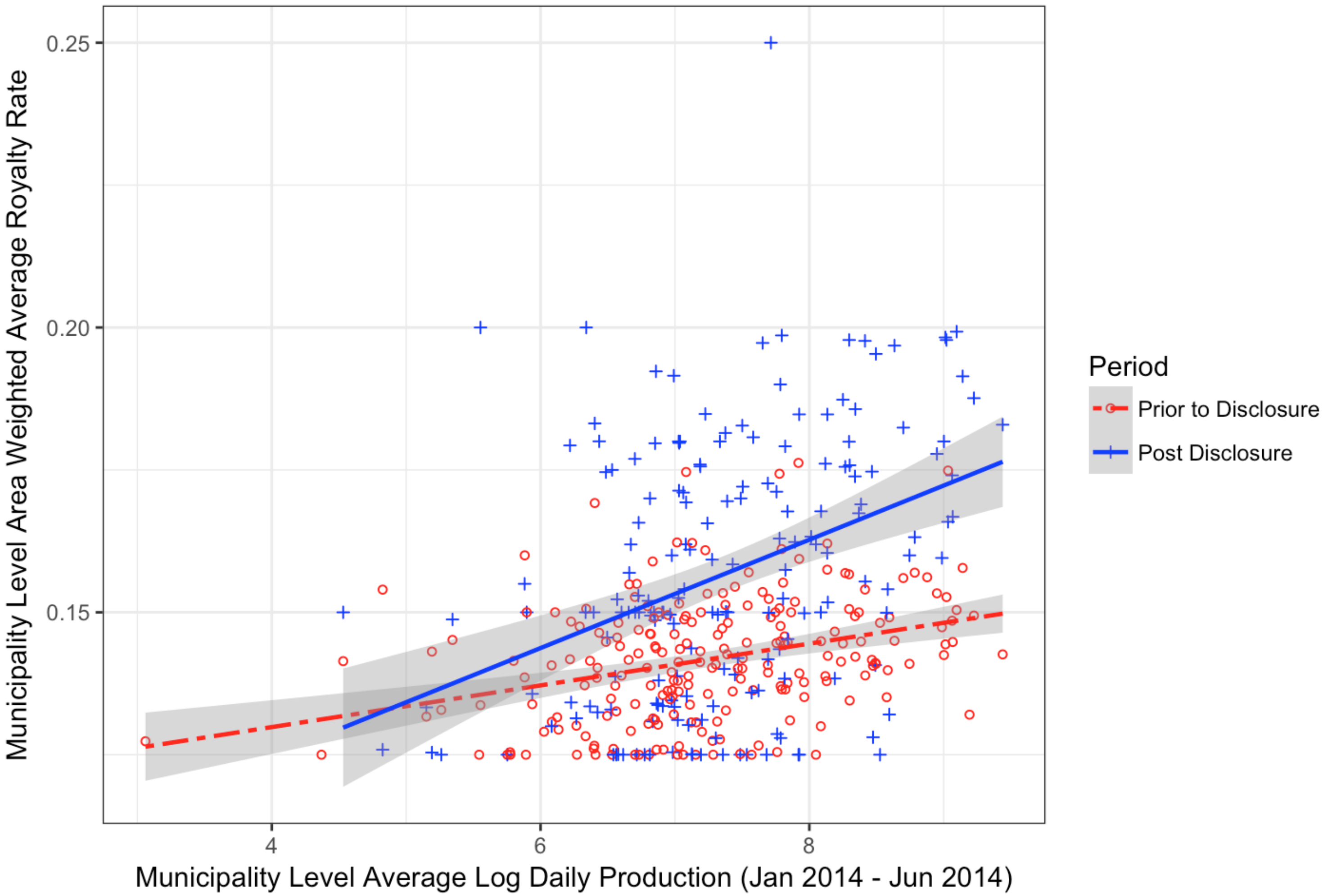
Note: the data is obtained from the PADEP. The lower histogram of Figure **??** is based on the horizontal wells in report 5 but not in report 1, whereas the upper diagram of Figure **??** is based on the entire set of horizontal wells in report 5. The lower histogram of Figure **??** is based on the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

Figure 11: Monthly Incremental Lease in PA (Log Form) and Natural Gas Price (January 2007 - December 2014)



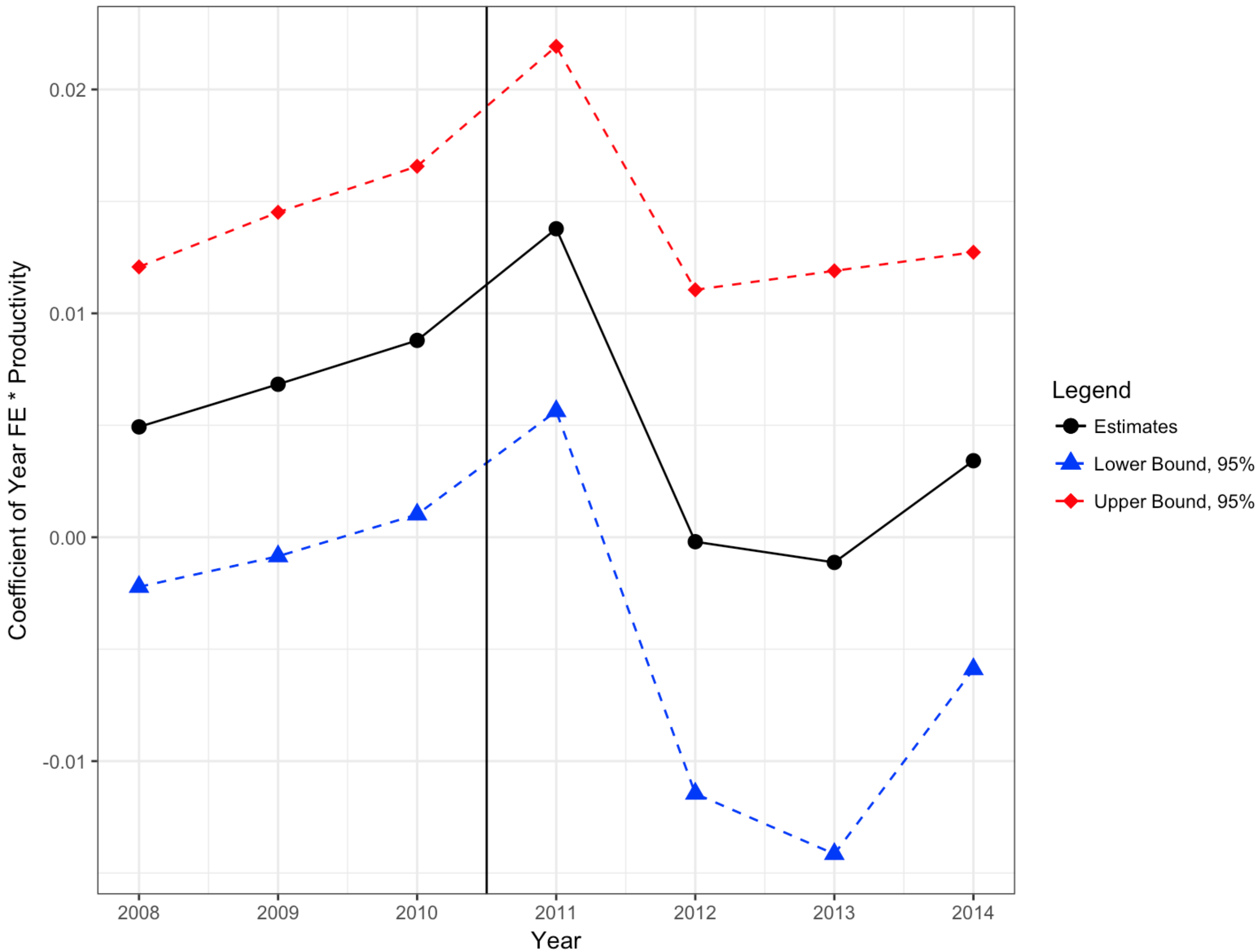
Note: the price data is obtained from EIA. The link is `https://www.eia.gov/dnav/ng/hist/rngwhhdm.htm`. The lease data is obtained from Enverus. The solid curve measures the logarithm of the lease acreages granted in each month in Pennsylvania. The dashed curve measures the natural gas spot price at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

Figure 12: 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 production of shale gas for each municipality. The royalty rate data is obtained from Enverus. The productivity data is obtained from the PADEP. The unit of productivity is thousand cubic feet per day. The average log daily production is calculated using the shale gas production from January 2014 to June 2014. The royalty rates of the red dots are averaged over the leases taken prior to the production information disclosure, and the royalty rates of the blue dots are averaged over the leases taken post to the production information disclosure.

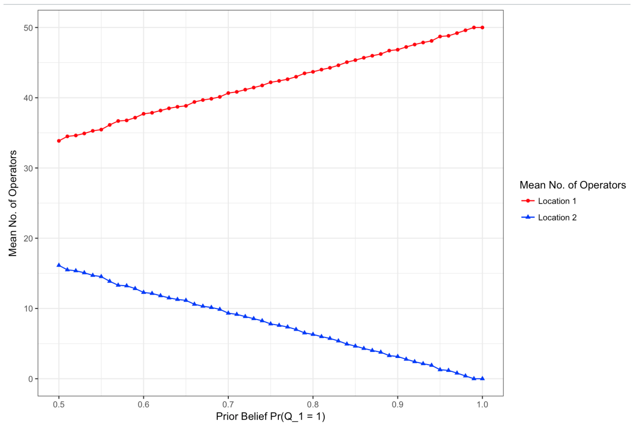
Figure 13: Plot of The Yearly Coefficient Estimates of Productivity in Equation (3)



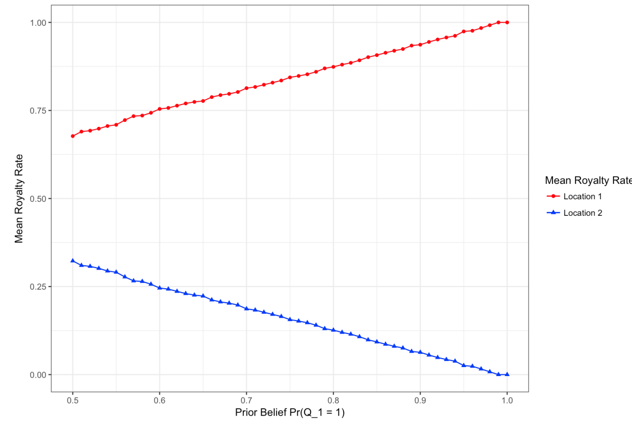
Note: the regression equation is $Royalty_{j,t} = \beta_0 + \beta_{1,t} Prod_j^1 + \beta_t + \varepsilon_{j,t}$. The coefficient of productivity in 2007 is estimated to be -0.0012 with a standard error of 0.0027. The solid line in the plot connects the coefficient estimates of the interacting variables between productivity and the year fixed effects. The values of the round dots, therefore, represent the change of the estimates of productivity's coefficient in the corresponding years relative to the estimate of productivity's coefficient in 2007. The dashed curves represent the upper and lower bounds of the 95% confidence intervals of the coefficient estimates.

Figure 14: Simulation Results for Different Royalty Cost Functions

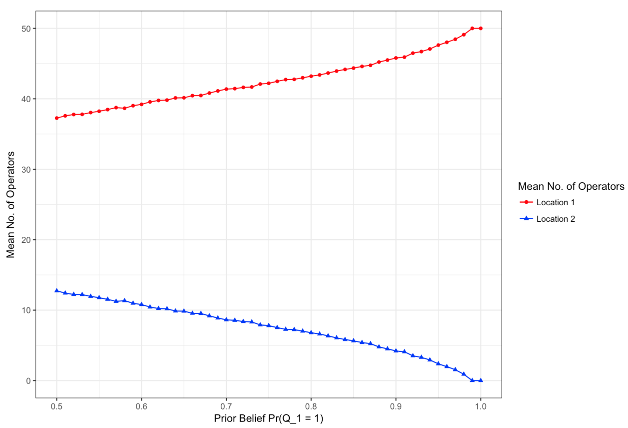
(a) Mean No. of Operators, $r(x) = x/N$



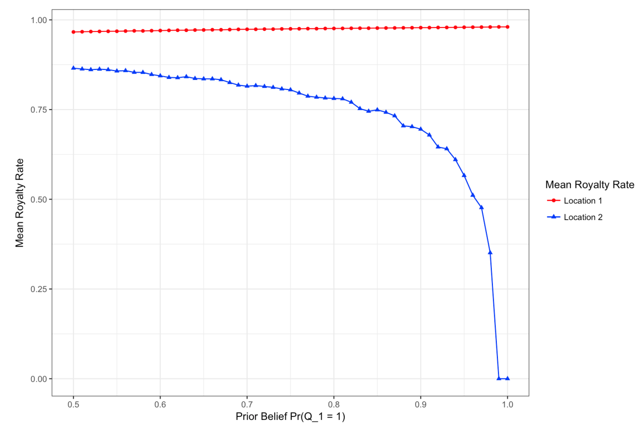
(b) Mean Royalty Rate, $r(x) = x/N$



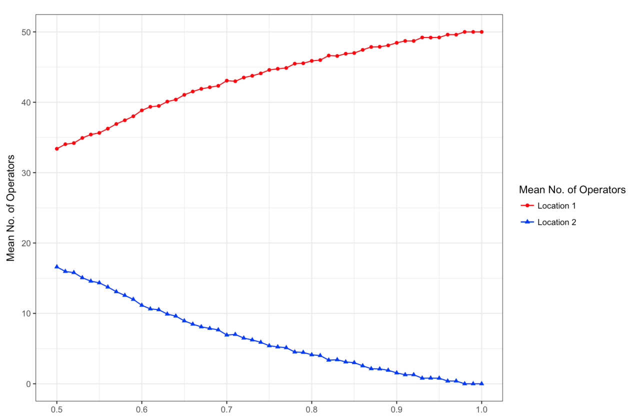
(c) Mean No. of Operators, $r(x) = 1 - 1/(x + 1)$



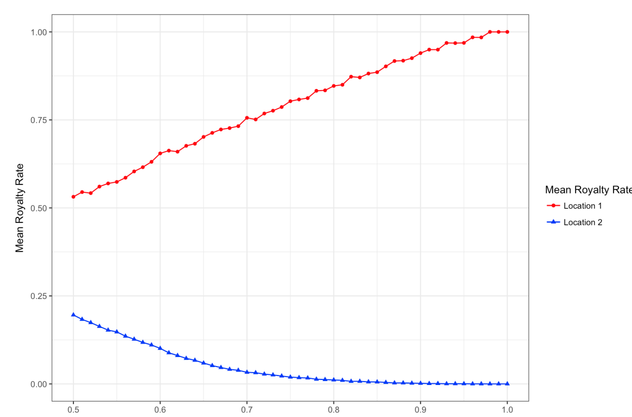
(d) Mean Royalty Rate, $r(x) = 1 - 1/(x + 1)$



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

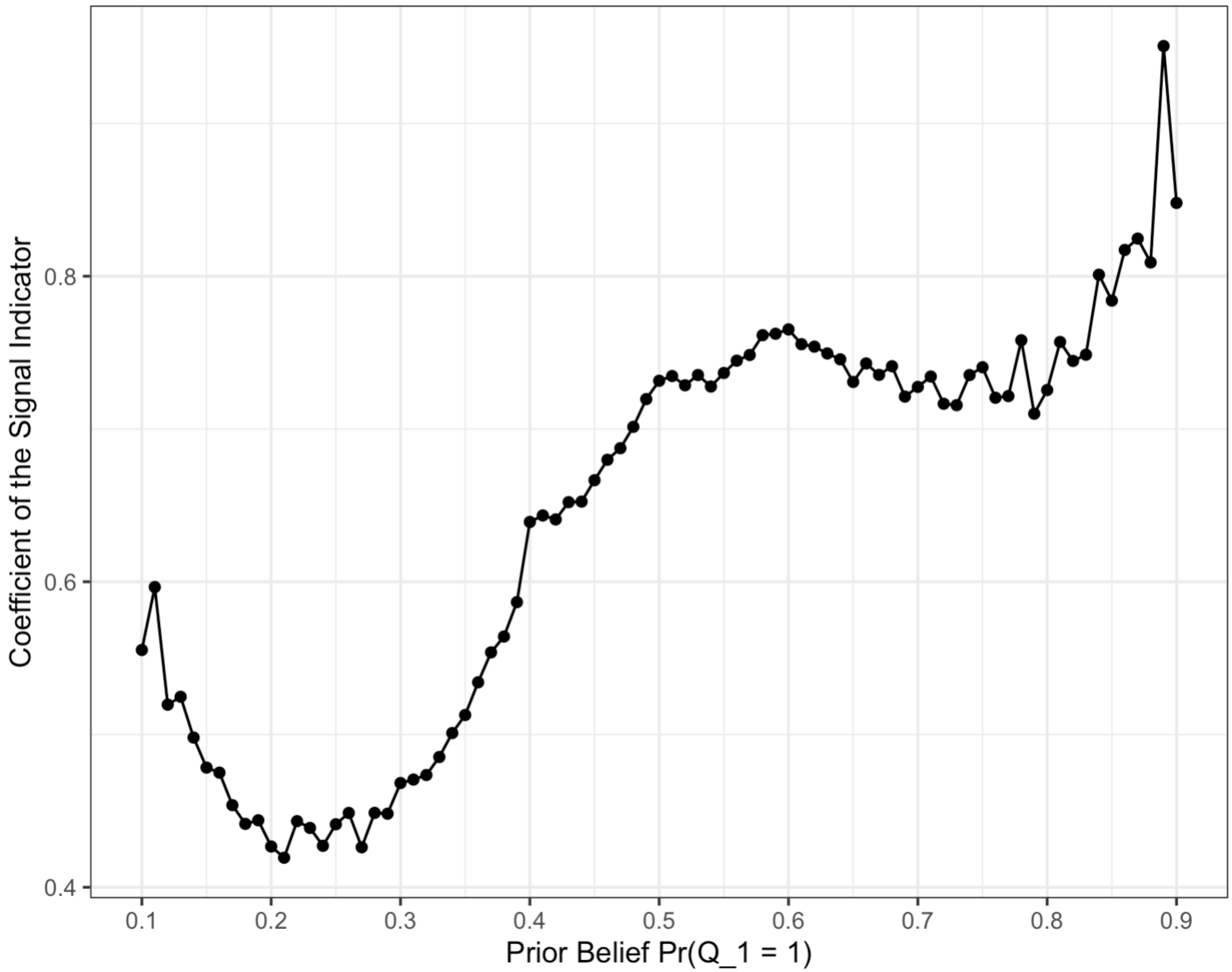


(f) Mean Royalty Rate, $r(x) = x^2/N^2$



Note: this set of figures illustrates the simulation results of the Bayesian learning process in the theoretical model based on different royalty cost function. 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 **??** and **??** 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 **??** and **??** are the same plots based on the reciprocal royalty rate function. Figures **??** and **??** are the plots based on the quadratic royalty rate function.

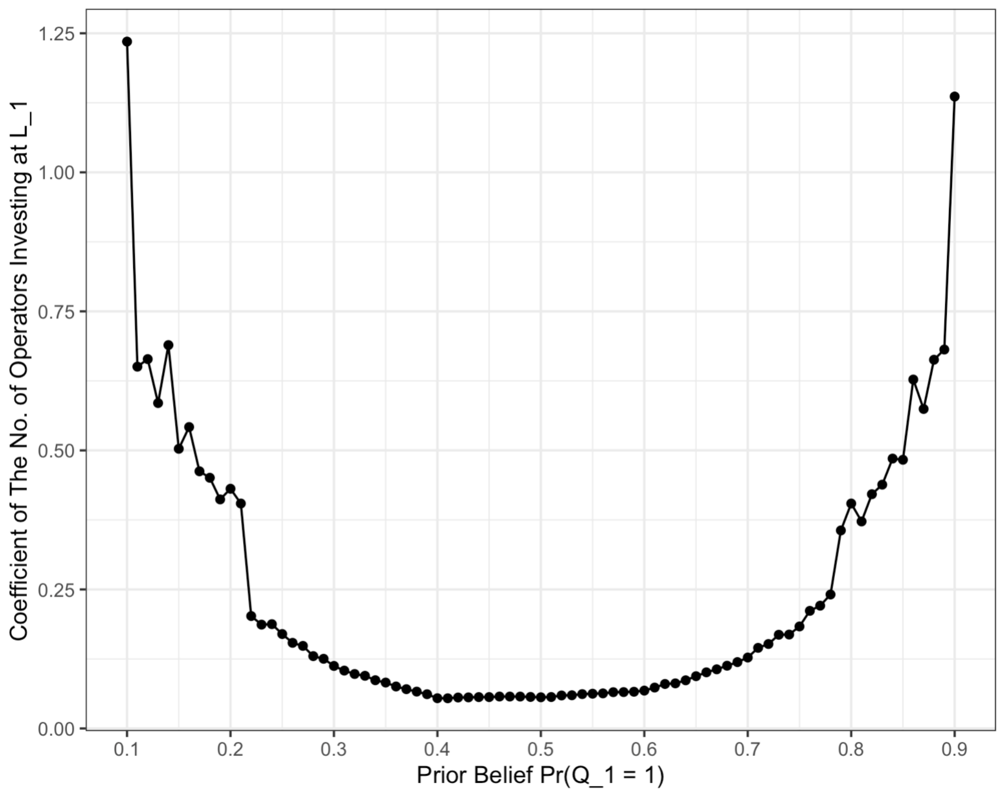
Figure 15: Coefficient Estimates of The Signal Indicator in Equation (26) and (27) Based on Different Prior Beliefs



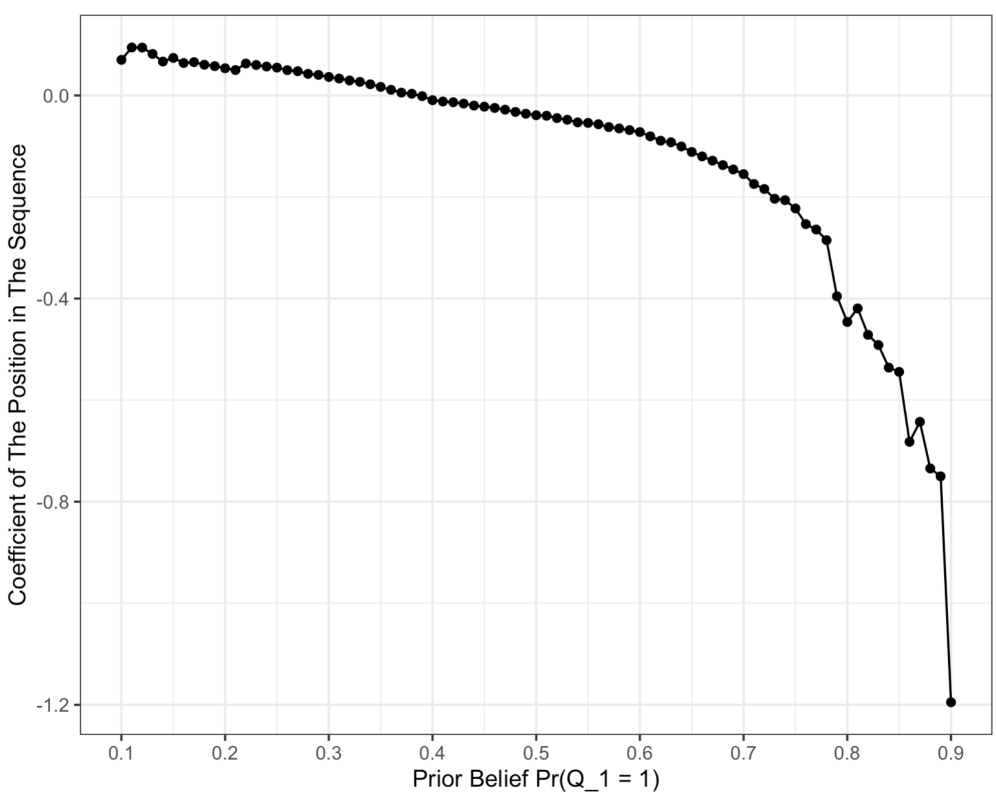
Note: the coefficient estimate for each prior belief is obtained using the simulated sample based on the linear royalty cost function $r(x) = x/N$, where N is set to be 50. Each sample consists of 10,000 experiments in which 50 operators sequentially choose their lease investment locations.

Figure 16: Coefficient Estimates of Equation (26) and (27) Based on Different Prior Beliefs

(a) Coefficient Estimates of The No. of Investors Based on Different Prior Beliefs



(b) Coefficient Estimates of The Position in Sequence 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.

Table 1: No. of Unconventional Permits Issued from Jan 2007 to 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 acquired from the 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 obtained 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 Data section. 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 acreage.

Table 3: The Release Time of and The Periods Covered by 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 is based on the production reports collected from the PADEP. The first report covers the production of twelve months from July 2009 to June 2010. Each other report covers the production of six months. The reports are assumed to be released on the first days of the months.

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 data is obtained from the 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: Summary Statistics of The Productivity Histograms in Figure **??**

	Figure **??**		Figure **??**	
	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 **??** is based on the horizontal wells in production report 5 but not in report 1, whereas the upper diagram of Figure **??** is based on the entire set of horizontal wells in report 5.

Table 6: Estimation Results of The Model Defined by Equation (2)

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

Significance codes: 0 '***', 0.001 '**', 0.01 '*', 0.05 '.'
Note: $Prod_j^9$ represents the average log daily production (Mcf./d) of municipality j calculated using the ninth production report. D_t is the information disclosure indicator variable that is equal to 0 if period t is "before disclosure" and 1 if period t is "post disclosure".

Table 7: Estimation Results of The Model Defined by Equation (4)

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

Significance codes: 0 '***', 0.001 '**', 0.01 '*', 0.05 '.'
Note: $Prod_j^9$ represents the average log daily production (Mcf./d) of municipality j calculated using the ninth production report. D_t is the information disclosure indicator variable that is equal to 0 if period t is "before disclosure" and 1 if period t is "post disclosure". S_j is the municipality type indicator variable that takes value 1 if municipality j is classified as shale gas oriented and 0 otherwise.

Table 8: Estimation Results of The Empirical Model Defined by Equations (6) and (7)

	Baseline		W/O FE		Linear Model	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
Own Recent Acreage	0.593	0.007***	0.660	0.007***	0.116	0.001***
Peer's Recent Acreage	-0.003	0.004	-0.015	0.004***	-0.000	0.000
Accumulated Acreage	0.029	0.001***	0.029	0.001***	0.000	0.000***
Productivity	23.591	0.924***	24.899	0.908***	0.079	0.022***
Productivity × Own Recent Acreage	0.007	0.002***	0.009	0.002***	-0.006	0.000***
Productivity × Peer's Recent Acreage	-0.000	0.001	0.000	0.001	0.000	0.000
Productivity × Accumulated Acreage	-0.001	0.000***	-0.001	0.000***	-0.000	0.000
Royalty Rate	631.163	49.983***	651.460	48.084***	1.054	1.920
Natural Gas Price	25.163	1.116***	24.988	1.055***	0.455	0.032***
Public	173.033	16.054***	150.768	4.208***	0.876	0.378*
Population	0.001	0.001**	0.002	0.001**	-0.000	0.000
Land Area	-0.000	0.000	-0.000	0.000*	0.000	0.000***
Month	-3.904	0.128***	-3.699	0.119***	-0.027	0.003***
Operator FE	Yes		No		Yes	
No. of Observations	1,057,741		1,057,741		1,057,741	

Significance codes: 0 '***', 0.001 '**', 0.01 '*', 0.05 '.'

Note: Own Recent Acreage is the lease acreages operator i obtains at municipality j in the past three months. Peer's Recent Acreage is the lease acreages operators other than i obtain at municipality j in the past three months. Accumulated Acreage is the accumulated lease acreages that have been taken at municipality j up to month $t-4$. 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 $t < 47$.

The second and the third columns show the coefficient estimates and the standard errors for the baseline latent variable model. The fourth and the fifth columns show the estimation results for the same model without the operator fixed effects. The last two columns show the estimation results for a linear model.

Table 9: Robustness Test Based on Different Window Widths

	2007 – 2014		2008 – 2013		2009 – 2012		2010 – 2011	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
Own Recent Acreage	0.660	0.007***	0.660	0.008***	0.631	0.010***	0.553	0.012***
Peer's Recent Acreage	-0.015	0.004***	-0.021	0.005***	-0.088	0.010***	-0.139	0.016***
Accumulated Acreage	0.029	0.001***	0.029	0.001***	0.026	0.001***	0.022	0.001***
Productivity	24.899	0.908***	21.045	1.031***	12.510	1.204***	12.587	1.611***
Productivity × Own Recent Acreage	0.009	0.002***	-0.006	0.002*	0.037	0.004***	0.028	0.005***
Productivity × Peer's Recent Acreage	0.000	0.001	-0.000	0.001	0.004	0.004	0.020	0.004***
Productivity × Accumulated Acreage	-0.001	0.000***	-0.001	0.000***	-0.001	0.000***	-0.001	0.000***
Royalty Rate	651.460	48.084***	915.313	49.872***	1256.381	56.864***	3308.702	160.854***
Natural Gas Price	24.988	1.055***	22.655	1.175***	10.407	3.352**	-2.667	6.910
Public	150.768	4.208***	148.485	4.515***	165.648	4.535***	135.818	7.325***
Population	0.002	0.001**	0.001	0.001*	0.001	0.001	0.001	0.001
Land Area	-0.000	0.000*	-0.000	0.000	-0.000	0.000***	0.000	0.000
Month	-3.699	0.119***	-4.151	0.170***	-5.204	0.248***	-9.636	0.688***
Operator FE	No		No		No		No	
No. of Observations	1,057,741		847,069		567,427		299,698	

Significance codes: 0 '***', 0.001 '**', 0.01 '*', 0.05 '.'

Note: the second and the third column show the estimation results using the full lease sample from 2007 to 2014. The fourth and the fifth column are the estimation results based on a narrower window width from 2008 to 2013. The sixth and the seventh column are the estimation results based on an even narrower sample from 2009 to 2012. The last two columns are the estimation results for the narrowest window width from 2009 to 2012. For all the samples, the production information is disclosed in the middle of the window width.The operator fixed effects are dropped for the estimation using all the four samples.

Table 10: Test of The Virtual Effect of The Production Information If Disclosed at The End of 2012

	Original Sample		Artificial Sample	
	Coef.	S.E.	Coef.	S.E.
Own Recent Acreage	0.660	0.007***	0.768	0.013***
Peer's Recent Acreage	-0.015	0.004***	-0.016	0.009
Accumulated Acreage	0.029	0.001***	0.020	0.001***
Productivity	24.899	0.908***	27.063	1.001***
Productivity × Own Recent Acreage	0.009	0.002***	-0.031	0.002
Productivity × Peer's Recent Acreage	0.000	0.001	0.001	0.001
Productivity × Accumulated Acreage	-0.001	0.000***	-0.001	0.000***
Royalty Rate	651.460	48.084***	211.932	51.690***
Natural Gas Price	24.988	1.055***	0.737	2.960
Public	150.768	4.208***	107.591	4.722***
Population	0.002	0.001**	0.007	0.001***
Land Area	-0.000	0.000***	-0.001	0.000***
Month	-3.699	0.119***	-3.957	0.211***
Operator FE	No		No	
No. of Observations	1,057,741		592,507	

Significance codes: 0 ‘***’, 0.001 ‘**’, 0.01 ‘*’, 0.05 ‘.’

Note: the artificial sample is created by removing the observations from 2007 to 2010 from the original sample, and changing the values of productivity in 2011 and 2012 to 0.

Table 11: Estimation Results of The Empirical Model Defined by Equation (11)

	Log-linear W/ IV		Log-linear W/ IV		Log-linear W/O IV	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
Own Recent Acreage	0.431	0.000***	0.434	0.000***	0.428	0.000***
Peer's Recent Acreage	0.005	0.000***	0.004	0.000***	-0.000	0.000***
Accumulated Acreage	0.007	0.000***	0.007	0.000***	0.003	0.000***
Productivity	0.003	0.000***	0.003	0.000***	0.001	0.000***
Productivity $\times$ Own Recent Acreage	-0.000	0.000	-0.000	0.000	0.000	0.000
Productivity $\times$ Peer's Recent Acreage	-0.000	0.000***	-0.000	0.000***	0.000	0.000
Productivity $\times$ Accumulated Acreage	0.000	0.000***	0.000	0.000***	-0.000	0.000
Royalty Rate	-3.951	0.322***	-4.012	0.324***	0.023	0.014
Natural Gas Price	0.000	0.000	-0.000	0.000	0.005	0.000***
Public	0.013	0.003***	0.011	0.000***	0.010	0.003*
Population	-0.000	0.000***	-0.000	0.000***	0.000	0.000***
Land Area	-0.000	0.000***	-0.000	0.000***	0.000	0.000
Month	-0.000	0.000***	-0.000	0.000***	-0.000	0.000***
Operator FE	Yes		No		Yes	
No. of Observations	1,057,741		1,057,741		1,057,741	
Adjusted R^2	0.3385		0.3362		0.3841	

Significance codes: 0 '***', 0.001 '**', 0.01 '*', 0.05 '.'

Note: Own Recent Acreage is the logarithm of the lease acreages operator i obtains at municipality j in the past three months. Peer's Recent Acreage is the logarithm of the lease acreages operators other than i obtain at municipality j in the past three months. Accumulated Acreage is the logarithm of the accumulated lease acreages that have been taken at municipality j up to month $t-4$. 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 $t < 47$.

The second and the third columns show the coefficient estimates and the standard errors for the log-linear model with fixed effects, using per capita income of the county where the municipality locates in the same year the lease is granted as the instrument for royalty rate. The fourth and the fifth columns show the estimation results for the same model without the operator fixed effects. The last two columns show the estimation results for the same model without instrumenting for royalty rate.