RICE UNIVERSITY'S
BAKER INSTITUTE
FOR PUBLIC POLICY

Working Paper

Learning Where to Drill: Drilling Decisions and Geological Quality in the Haynesville Shale

Mark Agerton, Ph.D.
Nonresident Scholar, Center for Energy Studies

Learning Where to Drill: Drilling Decisions and Geological Quality in the Haynesville Shale*

Mark Agerton†

JEL Codes: D24, D25, L71, Q35, Q47

April 13, 2020

Abstract

We often link increasing productivity in resource extraction to innovation in *how* firms extract. Yet resource quality—*where* firms extract—is a key driver of productivity. Using a structural model and data from Louisiana's Haynesville shale, I disentangle the impacts of how and where firms extract natural gas. Mineral lease contracts, learning about geology, and prices actually explain more than half of growth in output per well—not just technological change. Neglecting this may lead to over-optimistic long-run supply forecasts. I also show that growth in output per well masked large distortions caused by mineral lease contracts, which reduced resource rents.

*This paper was previously circulated as "Drilling to Learn: Information, Royalty Rates and Drilling Decisions in the Haynesville Shale."

Special thanks to Edward B. Poitevent II (Stone Pigman Walther Wittmann LLC); Silas Martin and Lyndon Looger (Enverus, formerly Drillinginfo); Cullen Amend (Encino Energy); Jim McBride (JR McBride & Co); Gregory B. Upton and Mark J. Kaiser (LSU); Ryan Kellogg and Thomas Covert (University of Chicago); Peter R. Hartley, Xun Tang, Kenneth B. Medlock III, Mahmoud El-Gamal, Robin Sickles, Rossella Calvi, and Yunmi Cong (Rice); Nick Frazier (FDIC); Jim Griffin, Steve Puller, and Eric Lewis (Texas A&M); Jim Wilen, Bulat Gafarov, Kevin Novan, Katrina Jessoe, and Dave Rapson (UC Davis); Jevgenijs Steinbuks, Xavier Gine, and Hanan Jacoby (World Bank); Wesley Blundell (CSU East Bay); Lucija Muehlenbachs (University of Calgary and RFF); Alan Krupnick (RFF); Chuck Mason (University of Wyoming). The author gratefully acknowledges support from the Center for Energy Studies at Rice University's Baker Institute for Public Policy and from the Giannini Foundation. Thank you to Enverus (formerly Drillinginfo) and LSU's Center for Energy Studies for providing data.

†Assistant Professor, University of California Davis and Non-Resident Fellow, Center for Energy Studies at the Baker Institute for Public Policy. mjagerton@ucdavis.edu

Productivity in natural resource extraction is determined by both technology and resource quality, that is, *how* firms extract and *where* they extract. While the location of extraction activities may be observable, resource quality—determined by where firms extract—is usually not. When productivity increases, it is difficult to know whether firms got better at how they extracted, or whether they simply targeted higher quality resources.

Confounding changes in resource quality with productivity is particularly problematic when producing from high-quality resources today means that the resources are unavailable tomorrow. In this case, apparent productivity improvements might only be intertemporal shifts of productive capacity. Should we extrapolate apparent productivity gains into the future, we run the risk making overly optimistic supply projections.

As with any input into a production process, location and resource quality are inputs that firms choose based on economic factors, including prices and productivity. Since Marschak and Andrews (1944), economists have recognized that firms' behavior induces correlation between input choices and unobserved productivity shocks over time. This makes identifying productivity gains challenging. Resource extraction further complicates identification because resource quality is, in general, an unobservable choice variable, and it also varies over time.

In this paper, I disentangle the impacts of the economic forces that change *where* firms extract—prices, contracts, information, and depletion—with improvements in *how* they extract. To do this, I estimate a structural econometric model of firms' decisions to drill and extract natural gas. The setting is Louisiana's Haynesville shale over the period 2003–2016. I assemble a rich dataset that includes the terms of each mineral lease contract, the wells drilled on these leases, and the natural gas produced from each well. As Kellogg (2014), Levitt (2009), and Muehlenbachs (2015) do, I cast drilling as a Rust (1987)-style dynamic discrete choice model to drill a particular lease. I estimate the model jointly with equations for contract terms and production outcomes. The model incorporates all four economic forces that affect where firms drill.

Louisiana's Haynesville shale is an ideal setting to study the productivity, profit, and rent implications of how and where firms extract resources. The Haynesville is one of the major "shale plays" in the U.S where firms use horizontal drilling and hydraulic fracturing ("fracking") techniques to extract hydrocarbons. Firms determine production by choosing by when, where, and how to drill wells; they do not vary production from each well in response to prices (Anderson, Kellogg, and Salant 2018; Newell, Prest, and Vissing 2019). This means that I can study productivity using many individual production decisions. While information spillovers between adjacent locations are an issue in offshore drilling, their impact is likely to be much lower in lower-risk onshore shale drilling.[1] Common-pool externalities are also unlikely: hydrocarbons do not migrate easily through low-permeability shales.

My first main result is that three economic factors—prices, contracts, and learning about geology—induced systematic changes in where firms drilled. Even without technological progress, output per well would have risen. Naive estimates that fail to account for firms' choice of where to drill suggest that technology increased output per well by seven percent per year on average. Once I control for resource quality, this falls to just two percent.

The three economic factors worked in the following way. First, the price of output, natural gas, fell starting in 2009. At high prices, low-quality deposits were economic. When prices fell, firms "high-graded" extraction activities, raising average output per well. Second, use-it-or-lose-it deadlines in mineral lease contracts distorted firms' decisions about when and where to extract. Deadlines incentivized them to extract something from low productivity-locations right away, and then shift to high-quality ones. Third, firms were able to learn about the spatial distribution of resource quality by drilling. The value of information increased the economic payoff to drilling only one well in locations

[1] Spillovers of the sort studied by Hendricks and Kovenock (1989), Hendricks and Porter (1996), Hodgson (2018), and Lin (2013) should be limited in an onshore shale setting. Lower geological risk and lower costs in shale dampen the payoff to seeing information revealed by a neighbor. Mineral lease contracts limit the amount a firm can delay drilling an initial well. Locations usually accommodate several wells, and once an initial well is drilled, information about a neighbor's production will be of little value.

that initially appeared to be lower-quality. Subsequently, improved information allowed firms to target higher-quality locations. A fourth factor works against these three: depletion. As high-quality locations are depleted, firms will have to transition to worse ones.

My first result is important for several reasons. First, it provides new insights about the productivity of an important industry—U.S. unconventional oil and gas extraction. The U.S. recently became the world's top producer of oil and gas[2], and the majority of the country's oil and gas production comes from shale.[3] The boom in production has had significant economic impacts at the local, regional, and national level,[4] and has also impacted global energy markets (Baumeister and Kilian 2016; Hausman and Kellogg 2015; Kilian 2016, 2017).

Second, the narrative about on productivity in resource extraction tends to focus on technology (Cuddington and Moss 2001; Simpson 1999). A few papers consider the role of resource quality (Covert 2015; Managi et al. 2004; Montgomery and O'Sullivan 2017), but not firms' choices over the distribution of quality. More recent work has used shale extraction to study the underlying mechanisms by which firms learn about a production process (Covert 2015; Fetter et al. 2018; Fitzgerald 2015; Steck 2018). For these papers, minimizing the role of how firms choose where to drill is a necessary and reasonable simplifying assumption. However, is less benign for the purposes of understanding what drives productivity in shale or forecasting.

Third, I contribute to a small literature that pairs observed production data with a reduced-form model of the sampling process (drilling) to estimate an underlying resource distribution (Andreatta and Kaufman 1986; Bickel,

[2] `https://www.eia.gov/todayinenergy/detail.php?id=36292`

[3] In 2018, the U.S. Energy Information Administration (EIA) estimates that 59% and 72% of U.S. oil and gas production came from shale (6.5 mmbbl/d of oil and 60 bcf/d of gas).

[4] A review of the multitudinous studies on the economics impacts of the shale boom is not within the scope of this paper, but a few include Agerton et al. (2017), Çakir Melek, Plante, and Yucel (2018), Cosgrove et al. (2015), Decker, McCollum, and Jr (2018), Feyrer, Mansur, and Sacerdote (2017), Hausman and Kellogg (2015), Komarek (2016), Marchand and Weber (2017), and Upton and Yu (2019), and Marchand and Weber (2018) reviews several more.

Nair, and Wang 1992; Lee and Wang 1983; Meisner and Demirmen 1981; Smith 1980, 2018a; Smith and Ward 1981). Prior papers focus on the roles of output prices and depletion in determining unobserved resource quality. They do not allow for technological change, learning about geology, or mineral lease contracts. Using more detailed data afforded by U.S. shale activity and more structure, I show that additional economic factors matter a great deal to trends in output per well.

My second main result is that distortions from mineral lease contracts and improvements in firms' information about resource quality both impacted the discounted profits and resource rents more than improvements in technology over the period 2003–2016. Were firms to have owned the resource outright instead of paying royalties and facing use-it-or-lose-it deadlines, resource rents would have more than doubled. Improving or worsening firms' information about geology would have had more modest effects. Were firms to have had perfect information about the the spatial distribution of resource quality before drilling, rents would have only risen around 12%. Eliminating all learning would have lowered rents by around 37%. Somewhat surprisingly, I find that eliminating technological innovations would have only decreased resource rents by 17% and profits by 4%.

My result that mineral lease contracts lower resource rents adds empirical evidence to a recent set of papers examining how to individuals or firms should sell real options (Bhattacharya, Ordin, and Roberts 2018; Cong 2019; Herrnstadt, Kellogg, and Lewis 2018; Ordin 2019), as well as a older literature on how to to tax nonrenewable resources (reviewed by Lund (2009) and Smith (2013)). My result emphasizes that significant misallocation in oil and gas extraction does not require global market power as in Asker, Collard-Wexler, and De Loecker (2019): it also happens at the much smaller level of a private mineral lease contract.

My finding that learning about geology matters to resource rents contributes to a set of papers that study Hotelling-style models of nonrenewable resource extraction. This literature has identified two ways that new information from exploration increases welfare (Cairns 1990; Quyen 1991). First, discoveries increase the size of the resource stock. Second, discoveries resolve uncertainty about size the stock so that extraction can be more intertemporally efficient. I add a third purpose to new information about geology—enhancing the efficiency of how extraction gets allocated over space.

My results serve as a reminder that institutions matter in natural resources. We know that institutions shape resource management, determine rents, and drive economic performance of resource dependent countries (see reviews by Tarui (2015) and van der Ploeg (2011)). I show that institutions also shape the trajectory of productivity by determining where firms extract and, hence, resource quality. Ignoring how institutions determine unobservable resource quality introduces statistical bias in estimation of resource production functions—a point Reimer, Abbott, and Wilen (2017) also make in fisheries. The bias endangers the external validity of studies that use natural resource industries as settings to study broader economic questions. It also matters for forecasting resource production—something that industry and policy-makers in the world's largest producer of fossil fuels, the U.S., may care about.

1 Institutional details

Ownership of the mineral rights in the Haynesville is split among many private individuals.[5] Operators approach private mineral owners and negotiate bilateral mineral lease contracts with each. A lease grants the firm the option—but not obligation—to drill wells, extract minerals, and sell the production. In exchange, the firm agrees to pay the mineral owner an up-front, cash payment, the *bonus bid*, and a percentage of any revenue received from selling extracted minerals, the *royalty rate*. A record of the lease must be filed in the parish courthouse. Bonus bids are rarely reported, but most mineral lease records in the Haynesville specify the royalty rate. A high royalty rate can raise the landowner's revenue if the firm drills, but it also reduces the firm's incentive to drill.

[5] In the U.S. private individuals can own minerals, unlike most other countries, and State-owned minerals are a relatively small share of the Haynesville.

Mineral lease contracts expire after an initial *primary term*, usually three years. Should the firm drill and commence production within the primary term, the lease is considered to be *held by production*, and the operator maintains the right to drill as long as production continues in paying quantities (Lane, Freund, and McNab 2015; Smith 2018b). Many leases allow firms to extend the primary term in exchange for a cash payment. Such *lease extensions* normally last two years in the Haynesville. The first well drilled in a section holds all of the corresponding leases by production, even if a well is not physically drilled on each one.

The economic payoff to drilling an initial well can be larger than for later wells. An initial well provides the firm not only revenues but also a real option to drill more wells (usually seven) and new information about resource quality. The value of preserving the real option (Smith 2018b) and the value of new information can induce firms to drill financially unprofitable wells. Later development wells do not provide these additional economic payoffs, so we should expect that later wells will—if drilled—be more productive than initial wells. Firms also have more precise information about resource quality when drilling later wells. On average, this should raise output per well as firms target more profitable locations more effectively.

2 Data

Louisiana partitions the Haynesville into one square mile (640 acre) blocks called *sections* based on the 19th century PLSS grid (see Figure 1). Each section requires around eight wells to fully exploit. When a firm wants to drill a well and extract natural gas in a section, the State forms a *drilling unit* that usually coincides with the section.[6] While only one firm (the *operator*) is allowed to make decisions about a well, all parties with mineral interests in the unit must participate in the well. For my purposes, sections partition the shale into uniform sets of drilling opportunities with a single decision-maker. Because shale formations exhibit low permeability, wells in one section do not

[6] See Louisiana Revised Statutes of 1950, R.S. §3:9.

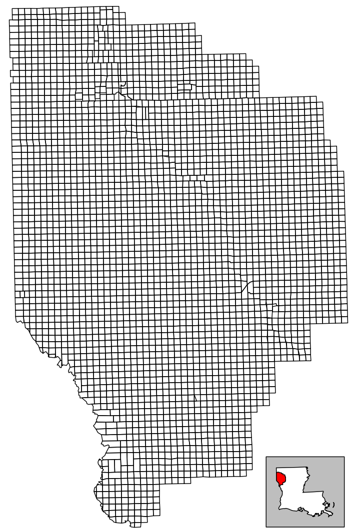


Figure 1: PLSS sections in Louisiana's Haynesville shale

drain hydrocarbons from a neighbor. This limits the scope for common-pool externalities in shale.

Firms in the Haynesville make investment decisions at the level of a section, so I take sections as my unit of observation. I observe three outcomes of firms' investment decisions on each section: the mineral lease contracts that firms sign, a sequence of drilling decisions, and a history of natural gas production from each well. Constructing my data involves merging these three datasets.

I define the geographic extent of Louisiana's Haynesville shale using a study on the geological quality of the Haynesville shale (Browning et al. 2015; Gülen et al. 2015). The study provides an estimated, spatial distribution of resource quality: "original gas in place" (OGIP). OGIP is based on coarse geological data like the thickness and total organic content of the shale.[7] Because it is calculated using geological fundamentals, not well production data, OGIP is not affected by firms' selection of where to drill. Firms had access to the sort of coarse geological information that OGIP is based on, so I assume that the variable is in their information set before they start leasing or drilling.

[7] Figure 6 in the Appendix shows a map of the OGIP measure over Louisiana's Haynesville.

I form sections by spatially merging Louisiana Department of Natural Resources (DNR) shapefiles of PLSS sections and Haynesville drilling units. I then spatially merge the following datasets to each section: the OGIP geology measure, land use characteristics and imperviousness from the U.S. 2001 National Land Cover Database, the urban/rural land classification from the 2010 U.S. Census, and the 2001–2006 average Census block-group characteristics from the American Community Survey (ACS).

I identify Haynesville shale wells from DNR data on their characteristics and spatial locations, and I merge them to sections.[8] The first well in my sample was drilled in September 2007, and the last, in October 2016. I gather well-level production data from commercial data provider Enverus.[9] I use futures prices from Bloomberg and deflate them to real terms using the PPI for final demand less food and energy.[10] I obtain lease locations and characteristics from Enverus and restrict attention to contracts that Enverus classifies as mineral leases, memorandums of lease, lease extensions, or lease amendments.[11] I spatially merge leases to sections. Sections usually contain many mineral leases. The first lease in my sample is signed in July 2003, and the last, in January 2016. Expiration dates go from January 2009 to November 2020. In sections that see at least one shale well drilled, I assume that neither

[8] I classify wells as "shale" wells if they lie within the geographic extent of the Haynesville as defined by the OGIP measure and are either permitted as a horizontal or Haynesville well by the DNR, or drilled into the Haynesville formation. I consider wells drilled into the shallower Fredericksburg or James Lime formations, any injection wells, and any wells with a vertical depth less than 8700' as non-shale wells. My definition of a shale well is very close to Herrnstadt, Kellogg, and Lewis (2018) but is slightly less restrictive. Most of the additional wells included in my sample are drilled by the operator Indigo. All of the wells that I classify as Haynesville wells access sands (formations) which wells in the Herrnstadt, Kellogg, and Lewis (2018) shale-well sample also extract from.

[9] Operators in Louisiana can report production by well or by groups of wells in the same lease or unit. Most Haynesville shale wells report production individually. Since some do not and instead report at the lease or unit level, I use production data from Enverus. For these cases, Enverus allocates lease and unit production volumes to the indiviual constituent wells by using drilling dates, well-test data, and models of production decline. Enerus was formerly known as Drillinginfo.

[10] BLS series WPSFD4131 from the FRED database.

[11] I exclude deeds that reflect outright transfer of mineral or royalty ownership, lease ratifications, lease options, lease assignments recorded when one firm transfers a lease to another firm, and any document classified as "Other" by Enverus.

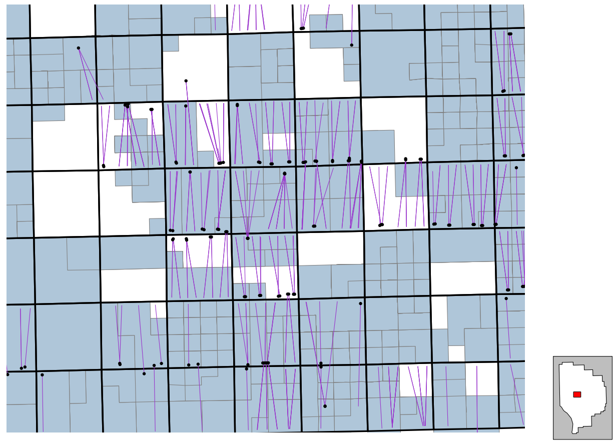


Figure 2: Wells, leases, and sections

leases which expired before the operator drilled the first shale well nor leases that start afterwards affected operators' decions. This assumption causes me to drop 14% of leases. In sections with no shale wells drilled, I do not have this issue.

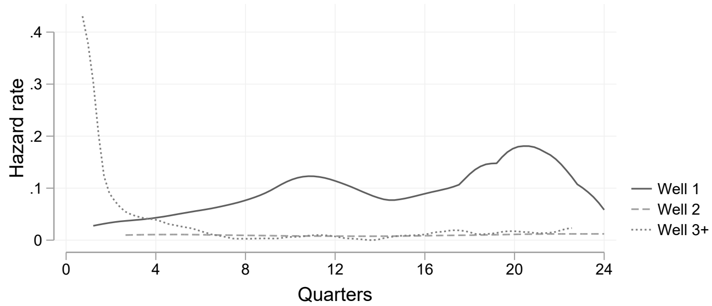
Figure 2 shows a map of how the data fit together in a small area within the Haynesville. The squares with heavy, dark outlines are the PLSS sections. The faint blue rectangles within each section represent the outlines of mineral leases of varying sizes. Leases generally fall within section-boundaries. Wells' surface locations are marked by round dots, and these are connected via the purple rays to the wells' terminus.

Since I focus on firms' drilling decisions made at the level of a section, I aggregate royalty rates and primary terms from the level of a lease to the level of a section. Almost all of the royalty rates in my data fall into one of six discrete categories: 12.5%, 16.67%, 18.75%, 20%, 22.5%, and 25%. I compute the average royalty rate in a section, weighting each lease by its share of ownership in the unit.[12] Average royalty rates are close to the discrete ones, so I map average royalty rates back to the nearest discrete one.

Wells drilled within a short time of one another are unlikely to be the

[12] See Section A.3 in the Appendix for how I compute the share of a unit that each lease owns.

Figure 3: Drilling hazard by well-order for 3 year leases



The hazard rate is the probability of drilling in this quarter conditional on having not been drilled before.
Time 0 is when any lease in the section starts or the day after the prior well was drilled. Leases are weighted by their relative size.

result of separate investment decisions and new information. Instead, firms must plan ahead to secure suppliers and regulatory approval. Drilling a well usually takes two to eight weeks (Cochener 2010; EIA 2012; Redlinger, Lange, and Maniloff 2019), and well completion (hydraulic fracturing) takes additional time. When a firm drills a well at the end of one quarter and another at the beginning of the next quarter, it has likely made one large investment, not two smaller ones. To reflect this, I classify any well drilled within 8 weeks (less than 63 days) of another as belonging to the same drilling decision.[13] I then aggregate time-varying variables like prices and the number of wells drilled to a quarterly frequency.

My final sample consists of 1384 of 2738 sections in the Haynesville. I drop sections which have missing data, non-Haynvesville wells, non-standard lease terms, initial wells that cross multiple sections, and urban areas. These sections are likely to differ systematically from standard Haynesville sections in terms of cost, contract, or production process. Appendix A provides more detail about why I drop certain sections.

3 Descriptive evidence

I verify that mineral lease expirations do in fact change firms' behavior by estimating nonparametric drilling hazard rates over a sub-sample of leases with a three year primary term. Most of these also specify an optional two year extension. I separate my sample by the order in which wells were drilled—Well 1, Well 2, and Wells 3+. Since there are multiple leases per unit, I weight each lease by the share of the unit that it owns.[14] Figure 3 plots the estimated hazard rates. The probability of drilling an initial well peaks when most primary terms and lease extensions expire at quarters 12 and 20 (three and five years).[15] The hazard rates for Well 2 and Wells 3+ are quite different from Well 1. The Well 2 hazard rate is nearly constant, and is much lower in level terms, reflecting a long delay between when firms drill initially and when they drill again. The hazard rate for Wells 3+ suggests that firms tend to either drill immediately after drilling the prior well, or they delay and drill much later (as with Well 1).[16] Such a pattern is consistent with fixed costs of drilling, such as moving rigs. It also suggests that firms learn about geology from Well 1 but not not Well 2 or Wells 3+.

To get a sense as to how the output of wells has evolved in the Haynesville, I estimate three preliminary regressions. Each includes a linear time trend associated with the well's *spud* (initial drilling) date. The trend captures increases in ouptut per well over time. The dependent variable is cumulative gas production (scaled by the horizontal length of the wellbore) from well w

[13] Figure 9 in the Appendix shows the distribution of weeks since the previous well was drilled and where the 8-week cutoff lands.

[14] Figure 4 in the Appendix estimates these rates assuming that that the primary term starts with the first lease signed or, alternatively, the last lease.

[15] Herrnstadt, Kellogg, and Lewis (2018) find the same result, and they statistically verify that drilling hazard rates drop discontinuously after mineral lease expirations.

[16] There are fewer Well 2s in the sample compared to total number of Wells 3+, and they tend to be drilled after a longer delay. For this reason, the hazard rate of Well 2 begins in quarter 3, where the rate for Wells 3+ begins earlier. The cumulative failure rate is shown in Figure 5 in the Appendix. It does not suffer from these edge effects but makes it more difficult to visually distinguish the spike in drilling rates around lease expirations.

in section i after τ months of production:

$$\log\left(Q_{iw\tau}/len_{iw}\right) = \gamma_0 + \gamma_g g_i + \gamma_{yr} yr_{iw} + \gamma_\tau + \psi_i + \eta_{iw\tau}. \tag{1}$$

The term γ_τ is a fixed effect that nonparametrically captures natural well decline after τ months of production. The term ψ_i is a section-specific fixed-effect that includes the section's geological productivity. I assume that the error term, $\eta_{iw\tau}$, is uncorrelated with the other right hand side variables, which include OGIP (g_i) and the year the well is drilled (yr_{iw}). I cluster standard errors at the section level to correct for serial correlation of $\eta_{iw\tau}$ within wells iw and correlation between wells in the same section i. I estimate three specifications with progressively more controls. Table 1 displays estimates.

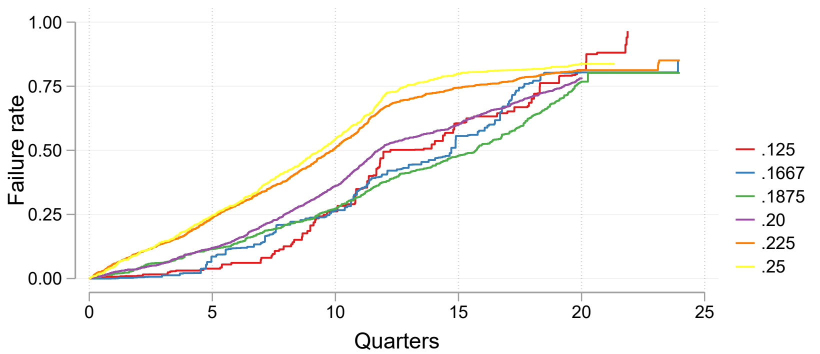
Table 1: Log linear model of cumulative production

	Naive OLS	OLS	Section FE
Spud date (years since July 2008)	0.07	0.04	0.00
	(0.01)	(0.01)	(0.01)
Log OGIP	0.53	0.37	
	(0.05)	(0.05)	
Was more than 1 well drilled in section?		0.20	
		(0.03)	
Average royalty rate		1.37	
		(0.41)	
Num. obs.	112714	112714	112714
Num wells	1799	1799	1799
Num units	1085	1085	1085

Dependent variable is the logarithm of cumulative production per foot from well w in section i after t months of production. Well length is measured as the lower minus the upper well perforation. The sample includes production months 4 through 72. Standard errors are clustered at the section-level to account for serial correlation and within-section correlation. Production month fixed effects control for a common well decline over time. Section fixed effects account for section-specific geology.

In the first specification, Naive OLS, I make the heroic assumption that unobserved section-specific geology, ψ_i, does not systematically change with the date wells are drilled. Model estimates imply a blistering 7% per year growth in output per well. The second model, OLS, includes an indicator variable for whether more than one well was drilled in the section and the average royalty rate in the section. The additional controls partially correct for correlation between resource quality, ψ_i, and the drilling date. Estimates imply that sections with multiple wells are around 20% more productive than sections with just one well. This suggests selection on unobserved quality is at work. Royalty rates are positively correlated with output per well. There are two possible explanations: firms might pay more for better locations, or higher royalty rates may eliminate drilling low-productivity locations. With additional controls, annual productivity growth estimates fall by nearly half, from 7% to 4%. Finally, in Section FE, I include section-specific fixed effects, ψ_i. This fully corrects for correlation between unobserved geological quality and the drilling date at the cost of removing all cross-sectional variation. Productivity changes are identified exclusively by comparing wells within the same section over time. The estimated trend in productivity falls to zero.

Figure 4: Cumulative probability of drilling Well 1 on 3 year leases by royalty rate



Failure rate is the probability that a section is drilled within t quarters of being leased.
Time 0 is when any lease in the section starts or the day after the prior well was drilled. Leases are weighted by their relative siz

Just as the number of wells in a section is informative about the productivity of the geology there, the timing of when firms drill is, too. We can exploit this fact to learn about the relationship between royalty rates and geology. If firms pay higher royalty rates in better locations, they will also accelerate drilling. If high royalty rates are independent of geology, firms delay or avoid drilling. Figure 4 plots nonparametric estimates of the cumulative probability of drilling Well 1 over time conditional the royalty rate (the failure function). With the notable exception of a small share of leases that have a 12.5% royalty

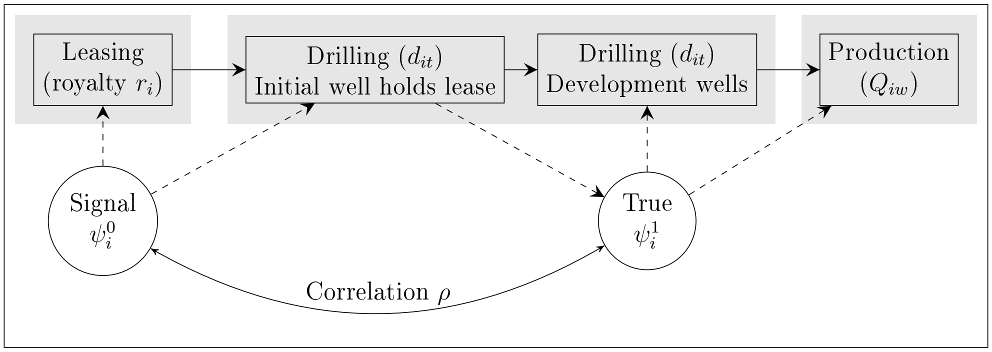


Figure 5: Signal (ψ^0) and true resource quality (ψ^1) link 3 observed outcomes in each section.

rate, the probability that a location is drilled sooner increases with the royalty rate. This suggests that firms pay higher royalty rates for better locations.[17]

4 Model

My goal is to evaluate how prices, mineral lease contracts and learning about geology affect drilling, average output per well, profits, and rents. To evaluate how these four outcomes would have evolved under different contracts or information sets, we need to know firms' drilling costs and their information sets. To identify these, I specify a model that combines leasing, drilling, and production in an economically consistent way.

Figure 5 diagrams the sequence of outcomes and the information structure in the economic model. Boxes at the top represent outcomes. Circles at the bottom represent firms' information. Dashed lines indicate how outcomes depend on information.

Upon arriving at section $i \in \{1, \dots, N\}$ to negotiate a lease, a firm receives

[17] The optimal contract derived by Herrnstadt, Kellogg, and Lewis (2018) implies that royalty rates rise with the degree of uncertainty about geology, not the quality of geology. The ability of small, private mineral owners to impose the optimal contract, however, relies on the assumption that they make take-it-or-leave-it offers to operators. The current and former landmen I have spoken with have suggested that it is normally operators who approach mineral owners and make offers. It is not unreasonable that actual mineral lease contracts deviate from the theoretical optimum.

two statistically independent pieces of information about section i's geological productivity. The first, g_i, is based on public information—the OGIP measure in Browning et al. (2015) and Gülen et al. (2015). Both the firm and I observe OGIP. The second is a noisy signal about section i's productivity, ψ_i^0. The firm knows ψ_i^0, but I do not. The firm uses the signal to form prior beliefs about section i's productivity. High signals can increase the firm's willingness to pay a higher royalty rate, and *vice versa*. Each quarter, the forward-looking firm decides how many wells to drill, d_{it}. High signals can cause the firm to accelerate when it drills one or more initial wells, and low signals, the opposite. Initial wells eliminate the mineral lease expiration (use-it-or-lose it deadline) and perfectly reveal section i's resource quality, ψ_i^1. Knowing ψ_i^1, the firm decides if and when to drill additional wells. Finally, ψ_i^1 and g_i together determine cumulative production, $Q_{iw\tau}$ for each well w in section i after τ months of production.

I assume that the signal and true productivity in section i have a joint standard-normal distribution with correlation ρ. The firm forms its prior beliefs about ψ_i^1 given ψ_i^0 as $F(\psi^1|\psi^0) = N(\rho\psi^0,\ (1-\rho^2))$. The correlation, $\rho \in (0,1)$ measures the precision of firms' initial signals. When signals are very noisy ($\rho \approx 0$), the value of information from initial wells is much larger than when signals are very precise ($\rho \approx 1$).

To estimate the economic model, I write the joint likelihood of the three outcomes—royalty rates, drilling decisions, and production—conditional on a joint realization of the signal, ψ^0, and true quality, ψ^1. I cannot observe the signal or true quality for each section, so I integrate them out of the likelihood by simulation and estimate their correlation, ρ.

4.1 Royalty rates

A royalty rate in section i is a discrete random variable $r_i \in \{\bar{r}_1, \ldots, \bar{r}_6\}$. It the outcome of a one-time negotiation between mineral owners and firms. Since we know little about the information structure of the game that the two play, I model the outcome in a way that allows—but does not require—firms' information to affect the royalty rate.

I assume that r_i is determined by a continuous latent variable r_i^*:

$$r_i^* = \underbrace{\beta_\psi \psi_i^0 + \beta_g g_i}_{\text{WTP}} + \underbrace{\beta_x^\top x_{ri}}_{\text{WTA}} + \nu_i. \tag{2}$$

The latent r_i^* is a linear combination of three sets of variables. The first set includes OGIP (g_i) and firms' signal about the location, ψ_i^0. Both can increase the firms' willingness to pay (WTP). The second set—mineral owner characteristics, x_{ri}—affect owners' willingness to accept drilling (WTA). These include median housing values, the imperviousness of a location's surface (a measure of development), and the share of minerals owned by out-of-state individuals.[18][19] I do not allow the payoff to drilling to depend on x_{ri}. This exclusion restriction rules out the possibility that landowners with low willingness to accept drilling impose restrictions that affect firms' drilling costs. The third set of variables only includes an i.i.d. bargaining shock, ν_i. Royalty rates take a discrete value $\bar{r}_l$ when r_i^* falls between two corresponding thresholds κ_{l-1} and κ_l: $r_i = \bar{r}_l \iff \kappa_{l-1} < r_i^* \leq \kappa_l$. The thresholds are ordered such that $-\infty = \kappa_0 < \kappa_1 < \ldots < \kappa_5 < \kappa_6 = +\infty$.

I assume that the the bargaining shock, ν_i, is normally distributed with variance normalized to one, and that it is statistically independent of the other right-hand side variables. Denote the CDF of the standard normal distribution $\Phi(\cdot)$. Then $\nu_i \sim F(\nu_i | g_i, x_{ri}, \psi_i^0) = \Phi(\nu_i)$. Under these assumptions, royalty rates can be estimated with an ordered probit regression that includes ψ_i^0 as a random effect. Denoting $\overline{r_i^*} \equiv \beta_\psi \psi_i^0 + \beta_g g_i + \beta_x^\top x_{ri}$, we can write the likelihood

[18] I include these characteristics based on the findings of Timmins and Vissing, who document that higher socio-economic status households have more leverage in negotiations with landmen (Timmins and Vissing 2014; Vissing 2015, 2016). Hitaj, Weber, and Erickson (2018) finds that absentee mineral owners behave differently than local mineral owners in leasing rural acreage.

[19] Time-varying variables do not enter this equation because it is the average royalty rate over all leases in a section that matters. Multiple leases imply that the point of time associated with a royalty rate is not well-defined.

of observing a particular royalty rate $r_i = \bar{r}_l$ as

$$L_i(r_i = \bar{r}_l | \psi_i^0, g_i, x_{ri}) = \Phi\left(\kappa_l - \overline{r_i^*}\right) - \Phi\left(\kappa_{l-1} - \overline{r_i^*}\right). \tag{3}$$

4.2 Drilling decision

In each section i and each quarter t, a firm decides how many wells to drill, d_{it}. Drilling is a dynamic decision: today's choice affects a firm's ability to drill tomorrow and (possibly) its information set.

Denote the endogenous state variable that determines the set of firms' choices as $s_{it} \in \mathcal{S}$. It includes information about the time remaining until a lease's primary term expires, the time remaining until its extension expires, and the cumulative number of wells drilled before period t, $D_{it} \equiv \sum_{s=0}^{t-1} d_{is}$. The firm cannot drill if the primary term or extension expire, or if it has drilled eight wells. I write the firms' action space as a correspondence Γ:[20]

$$\Gamma(s_{it}) = \begin{cases} \{0\} & \text{if lease extension expired} \\ \{0, 1, \ldots, 8 - D_{it}\} & \text{otherwise} \end{cases}.$$

All firms know OGIP, g_i, and their initial signal about the unobserved component of geological productivity, ψ_i^0. Firms choose whether to learn the true unobserved productivity, ψ_i^1, by drilling an initial well. Given the joint normality of ψ_i^0 and ψ_i^1, the state transition of the firm's information can be written as

$$F\left(\psi_{i,t+1} | \psi_{it}, D_{it}, d_{it}\right) = \begin{cases} N\left(\rho\psi_{it},\ (1-\rho^2)\right) & \text{if } D_{it} = 0 \text{ and } d_{it} > 0 \\ N(\psi_{it}, 0) & \text{otherwise} \end{cases}$$

where subscripts indicate the firm's information at time t and superscripts

[20] In specifying the state space, I make a simplifying assumption that if the option to extend is specified on the lease contract, then firms must either extend the lease or drill before the primary term expires. They cannot relinquish the lease after the primary term. This simplifies the modeling and avoids the problem that I cannot observe whether a firm actually pays to extend a lease. I can only observe if a firm drills during the extension or not.

denote whether there firm's information is a signal or true quality.

Firms take into account a vector of observable state variables, z_{it}, that affect the payoff drilling. These variables follow a first order Markov process with exogenous transitions. Group them into two components. The first, z_{1it}, is time-varying and contains real natural gas prices, p_t, and the state of technology: $z_{1it} = [p_t \quad yr_t]^\top$.[21] The second component, z_{2i}, is time-invariant and contains the average royalty-rate and the observable component of geology: $z_{2i} = [g_i \quad r_i]^\top$. Exogenous transitions means that $z_{i,t+1}$ is conditionally independent of the other state variables: $F(z_{i,t+1}|z_{it}, s_{it}, \psi_{it}, \epsilon_{it}, d_{it}) = F(z_{i,t+1}|z_{it})$. This does *not* rule out dependence between z_{it} and ψ_{it} because the royalty rate, r_i, may depend on ψ_i^0 through equation (2).

Finally, each period, the firm also receives a random vector of profitability shocks, ϵ_{it}, associated with each possible choice of how many wells to drill, d_{it}. Examples of these shocks include weather disruptions and availability of a suitable rig in the local area. I assume that shocks, ϵ_{it}, are i.i.d., and that the joint density of the state variables can be factored as

$$f(s_{i,t+1}, z_{i,t+1}, \psi_{i,t+1}, \epsilon_{i,t+1}|d_{it}, s_{it}, z_{it}, \psi_{it}, \epsilon_{it}) = \\ f_\epsilon(\epsilon_{t+1}) f_{s,\psi}(s_{t+1}, \psi_{i,t+1}|s_{it}, \psi_{it}, d_{it}) f_z(z_{i,t+1}|z_{it}).$$

Independence rules out serial correlation in ϵ. Instead, I allow for serial correlation in the unobserved component of profitability through ψ_{it}, which is updated once—from ψ_i^0 before the firm drills initial well(s) to ψ_i^1 after.

Drilling d wells yields a static payoff of

$$u(d, z_{it}, s_{it}, \psi_{it}, \epsilon_{it}) = \mathbb{E}[rev(d, z_{it}, s_{it}, \psi_i^1)|z_{it}, s_{it}, \psi_{it}] - cost(d, z_{it}, s_{it}) + \epsilon_{itd}. \quad (4)$$

Static payoffs are additively separable with respect to the choice-specific shocks, ϵ. This is standard in the dynamic discrete choice literature. Firms compute expected net revenues as the product of the number of wells, one minus the

[21] Appendix C.3 provides additional details on how I estimate and then discretize $F(z_{1i,t+1}|z_{1it})$.

royalty rate, natural gas prices less gathering charges, $gath$,[22][23] and EUR of the wells drilled, $Q(\cdot,\cdot,\cdot)$:

$$rev(d, z_{it}, s_{it}, \psi_i^1) = d(1-r_i)(p_t - gath)Q(g_i, \psi_i^1, yr_t). \tag{5}$$

The firm calculates EUR differently depending on whether it has drilled before ($D_{it} > 0$) and knows ψ_i^1 or whether the firm has not ($D_{it} = 0$) and must take a conditional expectation given its signal, ψ_i^0:[24]

$$Q(g_i, \psi_i^1, yr_t) = \exp\{\alpha_0 + \alpha_g g_i + \alpha_{yr} yr_t + \alpha_\psi \psi_i^1\} \tag{6}$$

$$\mathbb{E}[Q(g_i, \psi_i^1, yr_t)|\psi_i^0] = \exp\{\alpha_0 + \alpha_g g_i + \alpha_{yr} yr_t + \alpha_\psi \rho \psi_i^0 + \alpha_\psi^2(1-\rho^2)/2\} \tag{7}$$

Equation (7) makes clear that if correlation of ψ_i^0 and ψ_i^1, ρ, is close to one, then the signal ψ_i^0 changes behavior. If ρ is close to zero, then signals are uninformative and will not influence the probability of drilling. This implies that dispersion in the timing of initial wells across sections is informative of ρ. We obtain additional identification of ρ from the variance of well production across sections. If ρ is close to zero and signals are uninformative, then firms' targeting will be less precise, and variation in realized output across initial wells will be higher.

Equations (6) and (7) also include a common, linear technology trend to capture improvements in production know-how from year-to-year. A common trend is appropriate for this setting because shale producers do not drill wells themselves, rather, they use a common set of service companies that have

[22] I construct price p_t and natural gas gathering and processing charges $gath$ as a discounted flow of production revenues per unit of EUR, Q. Appendix C.2 describes how I do this using natural gas futures prices and a non-parametric estimate of production decline. This assumes that production decline rates are exogenous, consistent with industry practice as well as Anderson, Kellogg, and Salant (2018). In the paper, the authors show that the primary mechanism by which firms adjust to prices is via the decision to drill and complete a well—not the amount produced by a given well. This contrasts sharply with other, non-shale types of oil and gas extraction that may involve active injection of water or other gases to increase reservoir pressure and production.

[23] I set gathering charges to \$0.49 2009 USD per mcf following Gülen et al. (2015).

[24] The joint normality of ψ_i^1, ψ_i^0 and their independence from g_i and p_t imply the form of the conditional expectation.

developed many of the technological innovations in drilling and completion.

Drilling and completion costs are a function of the number of wells, d; the year yr_t; and an indicator function that takes the value one if the firm has to sign a lease extension and pay the mineral owner again, $ext(s_{it})$. There may be economies of scale to drilling multiple wells at once, so I allow average drilling costs to change by α_{2+} if a firm drills two or more wells. The function $h(yr_t; \alpha_h)$ captures variation in drilling and adjustment costs. In practice I use fixed effects for the years 2008–2012 with 2003–2007 and 2013–2016 having the same costs as 2008 or 2012.[25] The cost function is

$$cost(d, s_{it}, z_{it}) = d\Big\{h(yr_t; \boldsymbol{\alpha}_c) + \alpha_{2+}\mathbb{1}[d \geq 2]\Big\} + \alpha_{ext}ext(s_{it}). \tag{8}$$

Given a discount factor $\beta \in (0, 1)$, a firm's objective is to maximize the discounted sum of its static and dynamic payoffs. Dropping the i subscript and denoting $t+1$ with a trailing $'$, I write the firm's dynamic program as

$$V(s, z, \psi, \epsilon) = \max_{d \in \Gamma(s)} u(d, s, z, \psi, \epsilon) + \beta\, \mathbb{E}\left[V(s', z', \psi', \epsilon')|s, z, \psi, \epsilon, d\right]. \tag{9}$$

There are two absorbing states: when a lease expires before the firm drills, and when the firm drills all eight possible wells. In these states, the firm is unable to take further action, and I assume that the value of being in either is zero: $V(s, z, \psi, \epsilon) = 0$ for $s \in \{\text{expired}, \text{exhausted}\}$.

In estimation, I work with the firm's expectation of the value function (9) in $t+1$ given its choice in t:

$$\mathbb{E}\mathrm{V}(s', z, \psi) = \mathbb{E}\left[\max_{d \in \Gamma(s')} u_d(s', z', \psi') + \beta\, \mathbb{E}\mathrm{V}(s'', z', \psi') \middle| z, \psi\right]. \tag{10}$$

As in Kellogg (2014), I assume that firms do not account for future realizations of ϵ when calculating their expectations of future payoffs. Instead, ϵ is a surprise that firms learn each period. For this reason, $\epsilon_{d,it}$ does not appear in

[25] There is essentially no drilling before 2008, so time-varying fixed effects in 2003–2007 are not identified. Costs are fairly stable after 2012 if I use a third-order polynomial of time instead.

(9). [26] Define the choice-specific (alternative-specific) value function v_d as

$$v_d(s, z, \psi) = u_d(s, z, \psi) + \beta\,\mathbb{E}\mathrm{V}(s'(s, d), z, \psi). \tag{11}$$

To form the likelihood, I assume that vector of unanticipated choice-specific shocks ϵ is composed of random draws from a multivariate Type-I Extreme Value distribution with a location parameter equal to zero and scale parameter σ_ϵ.[27] The probability of observing action d conditional on all state variables except ϵ is a multinomial logit: $\Pr(d|s, z, \psi) = \frac{\exp\{v_d(s,z,\psi)\}}{\sum_{l\in\Gamma(s)}\exp\{v_l(s,z,\psi)\}}$.

Sections are usually associated with multiple leases $j = 1, \ldots, J_i$. Thus, there are potentially J_i pairs of mineral lease start and expiration dates, and J_i candidates for the section-level state variable s_{ijt} in each quarter. I assume that the firm chooses only one expiration date to matter, and that the probability a firm chooses a particular lease to matter, $\Pr(j|i)$, is equal to the share of the minerals in a section that the lease owns.[28] I therefore integrate over the set of possible state variables s_{ijt} implied by the leases. Lease expirations do not matter once the firm holds a section by production, so I only need to integrate over s_{ijt} for periods before an initial well is drilled. Denote $\bar{T}_{1i}$ as the first quarter in which the firm drills. Then likelihood of observing a sequence of

[26] Assuming firms do not anticipate future ϵ shocks has two benefits, though it does not substantially alter the signs and magnitudes of coefficients. First, it significantly improves the fit of the model and decreases the implied scale of ϵ. Second, when I assume that firms do take expectations over ϵ, the option value associated with these cost shocks represents much of the value of a well—not the financial payoffs from drilling. The option-value is especially inflated because of the relatively large number of alternatives that firms choose between (up to 9). See column T1EV in Table 2.

[27] An alternative formulation would be to assume that firms receive just one cost shock, and that the cost to drill d wells is $d(cost_t + \epsilon_{it})$. Because of the linearity of the payoff in d, however, such a model can only rationalize corner solutions.

[28] I estimate the model under a few alternative assumptions about which mineral leases matter to the firm (see Table 2). Results are essentially unchanged.

drilling decisions $\{d_{it}\}_{t=1}^{\bar{T}_i}$ in a section conditional on ψ_i^0 and ψ_i^1 is

$$L_i\left(\{d_{it}\}_{t=1}^{\bar{T}_i}\middle|\{z_{it}\}_{t=1}^{\bar{T}_i},\left\{\{s_{ijt}\}_{t=1}^{\bar{T}_i}\right\}_{j=1}^{J_i},\psi_i^0,\psi_i^1\right)=$$
$$\left[\prod_{t=T_{1i}+1}^{\bar{T}_i}\Pr(d_{it}|s_{it},z_{it},\psi_i^1)\right]\left[\sum_{j=1}^{J_i}\left(\prod_{t=1}^{T_{1i}}\Pr(d_{it}|s_{ijt},z_{it},\psi_i^0)\right)\Pr(j|i)\right]. \quad (12)$$

4.3 Production

The final component of the model consists of monthly production outcomes from each well. The expected profitability of a well is most closely linked to expected ultimate recovery (EUR), not month-to-month variations in output, so I focus on cumulative production, $Q_{iw\tau}$ from well w in section i after $\tau \in \{4, \ldots, 72\}$ months of production.[29] I assume that cumulative production, normalized by the horizontal length of the wellbore, is determined in a very similar way to the regression estimated earlier in Section 3:

$$\log(Q_{iw\tau}/len_{iw}) = \gamma_0 + \gamma_\tau + \alpha_g g_i + \alpha_{yr} yr_{iw} + \xi_{iw\tau} \quad (13)$$
$$\xi_{iw\tau} = \alpha_\psi \psi_i^1 + u_{iw} + \eta_{iw\tau}. \quad (14)$$

Equations (13) and (14) cast cumulative production $\log Q_{iw\tau}$ as a function the section's OGIP, g_i, the year the well was drilled yr_{iw}, and a common decline curve, γ_τ. The random effect, $\xi_{iw\tau}$, has three components. The first one is the true quality of a section, ψ_i^1. This is shared between all wells in a section. The second and third are also i.i.d. normal well-specific shocks $u_{iw} \sim_{iid} N(0, \sigma_u^2)$ and section-well-month output shocks $\eta_{iw\tau} \sim_{iid} N(0, \sigma_\eta^2)$. Random effects implies that the joint CDF of u, η is $F(u_{iw}, \eta_{iw\tau}|\psi_i^1, g_i, yr_{iw}) = \Phi(u_{iw}/\sigma_u)\,\Phi(\eta_{iw\tau}/\sigma_\eta)$. Conditional on ψ_i^1, the likelihood of observing a T_{iw}-

[29] Male et al. (2015) and Herrnstadt, Kellogg, and Lewis (2018) both note that the initial three months of production data are particularly noisy, so I drop these from the data. I drop observations after month 72 as these add little information.

length vector of cumulative production is

$$
\begin{aligned}
L\left(\{\log\left(Q_{iw\tau}/len_{iw}\right)\}_{\tau=1}^{T_{iw}}\middle|\psi_i^1, g_i, yr_{iw}; \gamma_\tau\right) = \\
&-\frac{1}{2}\left[T_{iw}\log(2\pi) + (T_{iw}-1)\log\sigma_\eta^2 + \log(\sigma_\eta^2 + \sigma_u^2 T_{iw})\right] \\
&-\frac{1}{2\sigma_\eta^2}\left[\sum_\tau (u_{iw} + \eta_{iw\tau})^2 - \frac{\sigma_u^2}{\sigma_\eta^2 + \sigma_u^2 T_{iw}}\left(\sum_\tau (u_{iw} + \eta_{iw\tau})\right)^2\right] \quad (15)
\end{aligned}
$$

where $u_{iw} + \eta_{iwt}$ is defined according to equations (13) and (14).

The coefficients α_g, α_{yr}, and α_ψ are the same for expected revenue (6, 7) and production (13, 14). This restriction imposes consistency between firms' decisions and well outcomes. Given a firm's marginal tax rate, I can identify σ_ϵ, the scale of the Type-I Extreme Value cost shocks using equation (4). Identification is based on equating firms' beliefs about EUR $Q(g_i, \psi_i^1)$ in equation (4) with actual cumulative production in equation (13).[30] A bit of algebra implies that we can compute $\hat{\sigma}_\epsilon$ as

$$
\hat{\sigma}_\epsilon = \exp\left\{\hat{\gamma}_0 + \hat{\gamma}_{240} + (\hat{\sigma}_u^2 + \hat{\sigma}_\eta^2)/2 + \log len_{50\%} + \log(1 - tax) - \hat{\alpha}_0\right\}. \quad (16)
$$

I describe how I estimate γ_{240} in Appendix C.1.

4.4 Model likelihood

Omitting exogenous variables, write the likelihood conditional on the noisy signal and true quality, ψ_i^0 and ψ_i^1 as

$$
L(history_i|\psi_i^0, \psi_i^1) = L\left(r_i\middle|\psi_i^0\right) L\left(\vec{d}_i\middle|\psi_i^0, \psi_i^1\right)\prod_{w=1}^{W_i} L\left(\log\vec{Q}_{iw}/len_{iw}\middle|\psi_i^1\right). \quad (17)
$$

Because I cannot observe ψ_i^0 and ψ_i^1, I integrate them out by simulation. Given M draws of (ψ_i^0, ψ_i^1), the simulated likelihood is $SL(history_i) = \frac{1}{M}\sum_{m=1}^M L_i\left(history_i|\psi_{im0}, \psi_{im1}\right)$.

[30] To be specific, consistency implies that $\exp\{\alpha_0 + \alpha_g g_i + \alpha_\psi \psi_i^1 + \alpha_t yr_{iw}\} = \mathbb{E}[Q_{iw,240}](1 - tax)/\sigma_\epsilon$ where the left-hand side is $Q(g_i, \psi_i^1, yr_t)$ from (6) and the $\mathbb{E}[Q_{iw,240}]$ on the right hand side is the expectation of (13).

The final statistical assumption I make is that all unobserved shocks are uncorrelated across sections. This includes the signal and true productivity, ψ_i^0 and ψ_i^1; royalty-rate shocks in (2), ν_i; choice specific shocks in (4), ϵ_{it}; well-specific production shocks, u_{iw}; and well-month production shocks, $\eta_{iw\tau}$. The assumption rules out the possibility of informational spillovers between neighboring sections and, consequently, any cause for strategic interactions. The simulated likelihood of the entire dataset is $SL(data) = \prod_i SL(history_i)$.

5 Estimation

I calibrate the the firm's nominal annual discount factor to be $\beta^{nom} = 1/1.125$ and scale it by inflation, which is 1.98% over the sample period. The real discount factor $\beta \approx 0.901$ is close to the values used by Covert (2015), Kellogg (2014), and Muehlenbachs (2015).[31] I estimate the model in three steps. First, I take production decline $\hat{\boldsymbol{\gamma}}_{\boldsymbol{\tau}}$ estimates from production-month fixed effects estimated in equation(1). While there are many of these coefficients, they are estimated precisely. I use these to calculate the present value of an additional unit of production (see Appendix C.1) and $\hat{\gamma}_{240}$.

In the second step, I estimate the parameters that characterize the exogenous processes real natural gas prices follow ($\log p_t$). I cannot reject the null hypothesis that the $\log p_t$ follows a random walk. I take a difference and estimate $\hat{\sigma}_p = 0.0900$. I discretize $\log p_t$ over an even grid of 51 points that extend $\pm \log 5$ beyond the minimum and maximum prices I observe.[32] I create a sparse transition matrix based on Tauchen (1986). Many elements of the matrix are small, so I zero out probabilities less than 10^{-5} to ease computation. To further reduce the dimension of the state space, I assume that the technology year transition is random: each quarter the firm believes yr_t will increase one unit and cause output per well to increase by α_{yr} until 2016, when technology is fixed. The sample ends in 2016, so productivity changes beyond

[31] See Appendix C.2 for further discussion.

[32] Kellogg (2011) similarly uses 51 grid points for log oil prices and extends the grid $\pm \log 5$ beyond the minimum and maximum observed.

this would not be identified from the data.

In the third step, I estimate the structural model using the Rust (1987) Nested Fixed Point (NFXP) algorithm. I use 2000 Halton draws to integrate out ψ^0 and ψ^1 and calculate standard errors using the Fisher Information matrix. Appendix C contains more details on computation.[33]

6 Results

Table 2 contains parameter estimates for a baseline specification plus five robustness checks. Signs of coefficients from the royalty-rate equation (2) are as expected. The impact of firms' initial signal, ψ_i^0, is positive and statistically significant, indicating that royalty rates are correlated with unobserved heterogeneity in geology. The lack of significance for the log OGIP variable, g_i, raises the possibility that public geological information affects royalty rates differently than potentially private signals ψ_i^0. Coefficients for variables affecting landowners' willingness to accept have the expected signs. Areas with higher housing prices and out-of-state owners require higher royalty payments. Locations with a greater share of permeable surface (less concrete and development) require lower royalty rates.

Equations for drilling (6) and production (7) share the same coefficients for log OGIP, unobserved resource quality, and time: α_g, α_ψ, and α_t. Because the variance of log OGIP, g_i, is just 0.33[34] versus 1 for ψ_i^1, unobserved resource quality explains more variation in well output than does observable variation in log OGIP. The estimated time-trend coefficient, $\hat{\alpha}_t = 0.022$, is lower than the Naive OLS and OLS estimates in Table 1 but still larger than the Section FE estimates that eliminate cross-sectional variation in the data. The difference between Section FE and structural estimates demonstrates the value of being able to include cross-sectional variation in the structural model. I estimate the correlation of firms' initial signals, ψ_i^0, with actual quality, ψ_i^1, to be $\hat{\rho} = 0.66$.

[33]Estimation routines are available publicly at `https://github.com/magerton/ShaleDrillingLikelihood.jl`

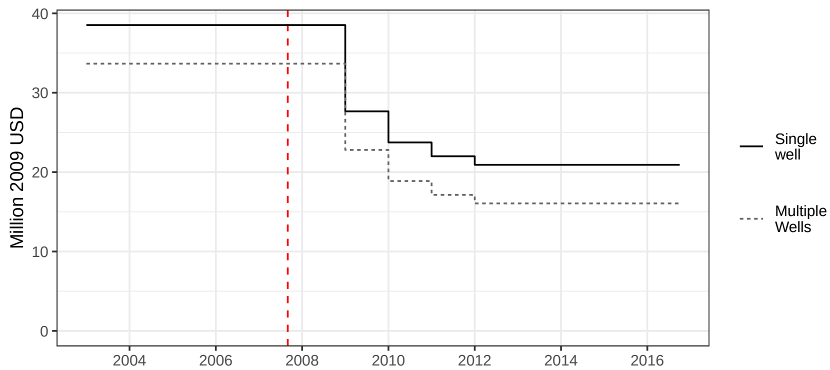
[34] See section-level summary statistics in Table 2 in the Appendix.

Table 2: Estimates for full model

		Use only 1 lease per section				
	Baseline	First	First, restr	Last	With rigs	T1EV
			Leasing			
ψ^0	0.113	0.118	0.191	0.216	0.116	0.190
	(0.052)	(0.050)	(0.052)	(0.065)	(0.049)	(0.096)
Log median house value	0.599	0.595	0.581	0.586	0.597	0.605
	(0.076)	(0.076)	(0.077)	(0.077)	(0.076)	(0.077)
Out-of-state owners (share)	1.183	1.182	1.188	1.182	1.184	1.195
	(0.138)	(0.138)	(0.139)	(0.140)	(0.138)	(0.142)
Pct impervious	-1.698	-1.697	-1.755	-1.735	-1.705	-1.720
	(0.510)	(0.508)	(0.520)	(0.525)	(0.511)	(0.513)
Log OGIP	0.140	0.140	0.143	0.144	0.140	0.142
	(0.096)	(0.096)	(0.097)	(0.097)	(0.096)	(0.097)
0.125 | 0.1667	3.868	3.828	3.605	3.667	3.843	3.900
	(1.034)	(1.036)	(1.052)	(1.056)	(1.037)	(1.040)
0.1667 | 0.1875	4.203	4.163	3.943	4.007	4.178	4.239
	(1.046)	(1.047)	(1.063)	(1.068)	(1.048)	(1.051)
0.1875 | 0.2	5.056	5.017	4.805	4.875	5.032	5.102
	(1.055)	(1.057)	(1.073)	(1.078)	(1.058)	(1.061)
0.2 | 0.225	5.955	5.917	5.716	5.790	5.931	6.011
	(1.059)	(1.060)	(1.077)	(1.082)	(1.061)	(1.066)
0.225 | 0.25	6.530	6.492	6.298	6.374	6.506	6.593
	(1.060)	(1.061)	(1.078)	(1.083)	(1.062)	(1.067)
			Drilling			
$\alpha_{2003--08}$	-12.489	-12.693	-10.763	-9.328	-10.487	-9.889
	(0.211)	(0.192)	(0.198)	(0.178)	(0.342)	(0.212)
α_{2009}	-8.965	-8.847	-8.749	-7.269	-7.102	-6.558
	(0.156)	(0.136)	(0.149)	(0.145)	(0.289)	(0.148)
α_{2010}	-7.696	-7.532	-7.812	-6.423	-5.772	-5.730
	(0.149)	(0.132)	(0.144)	(0.137)	(0.309)	(0.131)
α_{2011}	-7.131	-6.842	-7.339	-6.237	-4.960	-5.691
	(0.153)	(0.136)	(0.146)	(0.140)	(0.344)	(0.134)
$\alpha_{2012--16}$	-6.782	-6.605	-7.049	-6.241	-4.627	-5.659
	(0.140)	(0.125)	(0.134)	(0.123)	(0.349)	(0.118)
$\alpha_{d>1}$	1.576	1.554	1.557	1.356	1.583	1.502
	(0.074)	(0.071)	(0.070)	(0.071)	(0.074)	(0.068)
α_{rig}					-1.349	
					(0.222)	
α_{ext}	-1.495	-0.903	-0.753	-1.010	-1.591	-2.044
	(0.118)	(0.084)	(0.090)	(0.083)	(0.127)	(0.142)
α_0	-2.709	-2.629	-2.646	-3.008	-2.875	-3.442
	(0.221)	(0.215)	(0.215)	(0.216)	(0.239)	(0.241)
α_g	0.597	0.569	0.602	0.606	0.637	0.628
	(0.050)	(0.049)	(0.048)	(0.047)	(0.053)	(0.053)
α_ψ	0.340	0.340	0.346	0.341	0.358	0.351
	(0.011)	(0.011)	(0.010)	(0.010)	(0.012)	(0.009)
α_t	0.022	0.028	0.024	0.026	0.014	0.018
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
ρ	0.664	0.674	0.699	0.568	0.710	0.458
	(0.066)	(0.058)	(0.051)	(0.065)	(0.064)	(0.133)
			Production			
Intercept	-14.781	-14.655	-14.814	-14.810	-14.962	-14.863
	(0.241)	(0.236)	(0.231)	(0.226)	(0.256)	(0.252)
σ_η	0.097	0.097	0.097	0.097	0.097	0.097
	(1.852e-05)	(1.851e-05)	(1.847e-05)	(1.851e-05)	(1.855e-05)	(1.857e-05)
σ_u	0.320	0.319	0.321	0.313	0.317	0.297
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)
σ_ϵ	1.993	2.085	1.810	2.605	1.961	3.793
Avg drilling cost for 2+ wells	17.4	17.7	16.4	20.3	16.8	25.1
Log lik	93388.40	93383.92	94119.07	93413.32	93391.46	93175.53
Num z gridpoints	51	51	51	51	17	51
Num ψ gridpoints	51	51	51	51	19	51
Num simulations (M)	2000	2000	2000	2000	2000	2000

Mean well costs are the average drilling cost for multiple wells over the period 2009–2016. These are measured in million 2009 USD. This is calculated as $\frac{\sigma_\epsilon}{1-\tau_k}\frac{cost(2,s_{it},z_{it})}{2}$ where σ_ϵ is computed from (16). The effective marginal corporate income tax is 40.2%, and the marginal tax rate on capital investment is $\tau_k \approx 37.7\%$. Estimates 2–4 vary the set of mineral leases used for each unit: the first lease signed, the first lease signed with the restriction that the firm cannot drill until the last lease is signed, and the last lease signed. With rigs adds the rig dayrate as a regressor and requires coarsening the grid to keep computation feasible. The last column assumes that firms anticipate the Type I Extreme Value shocks.

Figure 6: Drilling costs



Horizontal lines show estimated average cost per well. Vertical red line marks when first well drilled in Sep 2007.

This means that while firms' initial beliefs are informative, they are by no means perfect, and the information initial wells provide can be valuable.

I calculate $\hat{\sigma}_\epsilon$ using (16) and use it to compute the cost to drill a single well and the average cost to drill more than one well.[35] These are plotted in Figure 7. My estimated average costs are higher than the drilling and completion costs of \$9–11 million and \$10.5 million reported by Kaiser and Yu (2014) and Gülen et al. (2015). However, my estimates include the full opportunity cost of the well—not just direct financial costs of drilling and completion. This includes operating expenses like disposal of produced water and future decommissioning costs. It also includes any other opportunity costs the firm incurs. Operators often take positions in multiple shale plays. If firms faced capital constraints or managers had limited attention as in Brown, Maniloff, and Manning (2018), drilling for cheap natural gas in the Haynesville would have detracted from the firm's ability to drill for more valuable oil elsewhere and increased the opportunity cost of drilling. That said, it is also

[35] Substituting in the median well length of 4428' into equation (16) (see Table 3 in the Appendix) and an effective corporate marginal income tax rate of $tax = 40.2\%$ supplied by Gülen et al. (2015), I estimate that $\hat{\sigma}_\epsilon = 1.99$. Drilling costs are capital expenditures and therefore taxed differently than production revenues. Again following Gülen et al. (2015), I assume that 80% of firms' drilling costs are expensable as intangibles, and that the remaining nominal 20% are depreciated at a constant rate over the following seven years. This implies that the effective corporate marginal tax rate for drilling expenditures is $tax_k = 37.7\%$. I multiply costs in equation (8) by $\hat{\sigma}_\epsilon/(1 - tax_k)$ to convert them into pre-tax dollars.

possible that I over-estimate drilling costs. In this case, percent changes in drilling, profits, and resource rents are still meaningful.

Figure 6 shows a remarkable decline in drilling costs between 2008 and 2009 as the fixed effects drop from $\hat{\alpha}_{2003-08}$ to $\hat{\alpha}_{2009}$. High opportunity costs in early years rationalize why firms did not drill when gas prices were at their peak. There are few explanations for high opportunity costs in 2008–2009. The period coincides with a financial crisis that generated significant economic uncertainty and may have limited access to capital to pay for drilling. The year 2008 was also the peak of a mineral-rights rush in the Haynesville.[36] Focused primarily on leasing minerals during a land rush, firms may not have had the capacity to additionally implement large drilling programs.[37] Industry executives I spoke with also described how operators needed time to overcome technical challenges associated with drilling Haynesville. The formation is deeper than the Barnett shale where firms started shale development, and it exhibits higher pressures and temperatures.

The final component of cost is the cost firms must pay to extend a mineral lease. The estimate of this, α_{ext}, is negative and highly significant. Scaled by $\hat{\sigma}_\epsilon/(1 - tax)$ and converted from dollars per section to dollars per acre,[38] it implies that costs to extend mineral leases were approximately \$5837/acre. Costs to extend mineral leases tend to track bonus payments. My extension costs lie within the range of bonus payment assumptions used in Gülen et al. (2015) and Kaiser (2012) (\$3000/acre and \$5000–25,000/arcre).

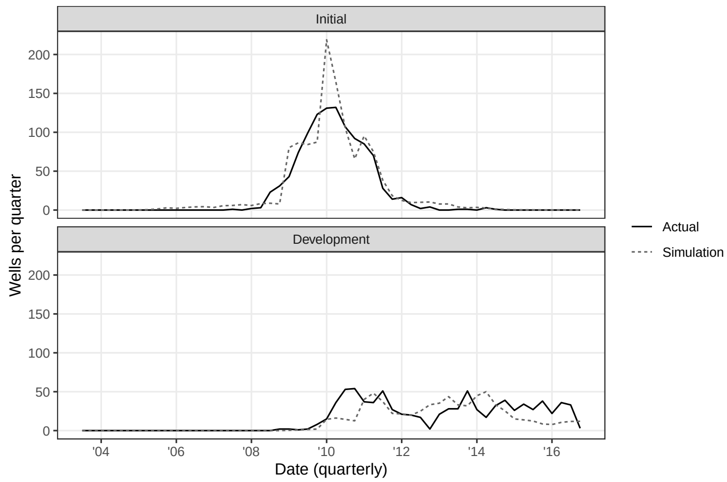
The right five columns of Table 2 are robustness checks. Columns "First" and "Last" do not integrate over the set of possible expiration dates. Instead, they assume either the first or last lease and its expiration date mattered to the firm. The "First, restr." estimate assumes that the first lease's expiration date matters, but the firm cannot drill until the last lease is signed. Parameter estimates are all qualitatively similar to the baseline specification. The second-

[36] See Figure 1 in the Appendix.

[37] One former landman described to me how his firm experienced drilling delays not because of insufficient equipment, but because of a regional shortage in capacity to verify title to the firm's mineral leases.

[38] Recall that there are 640 acres per section.

Figure 7: Model fit for drilling rates of initial and later development wells



Actual drilling rates computed directly from data. Simulation computed from the model using estimated parameters and taking an expectation conditional on royalty rates and prices.

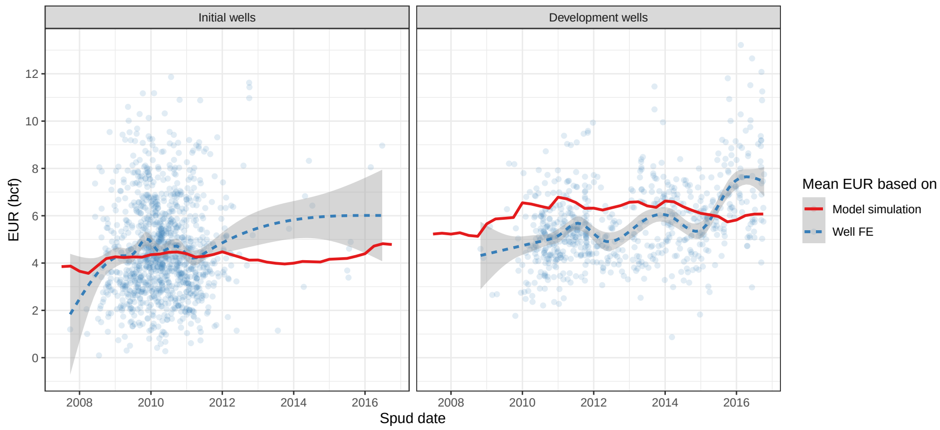
to-last column adds rig dayrates to better capture costs.[39] The likelihood improves mildly, but at a significant computational cost and a reduction in the number of grid points for prices, dayrates, and ψ_{it}. Finally, the last column assumes that firms anticipate the Type I Extreme Value shocks. While model fit is substantially worse and the implied σ_ϵ is much bigger, parameter signs and magnitudes are largely unchanged.

6.1 Model fit

To assess model fit, I compare actual drilling rates for initial and development wells with drilling rates predicted by the model. I simulate drilling given initial conditions and prices: $\sum_i \mathbb{E}\left[d(z_{it}, s_{ijt}, \epsilon_{it}) \middle| \{z_{s=0}^t\}, \{s_{ij0}, \Pr(j|i)\}_{j=1}^{J_i} x_{ir}, r_i\right]$. As royalty rates are correlated with ψ_{i0} and ψ_{i1}, I take care to integrate with respect to $dF(\psi_{i0}, \psi_{i1}|x_{ri}, r_i)$, not $dF(\psi_{i0}, \psi_{i1})$. Figure 7 shows that model predictions track actual drilling behavior. The coarse annual fixed effects for costs mean that the fit of drilling rates is poor at a quarterly level in 2010, but

[39] As Herrnstadt, Kellogg, and Lewis (2018) do, I use an index of dayrates purchased from market intelligence firm RigData for 1000-1499 horsepower drilling rigs in the Arkansas–Louisiana–Texas region. This index closely tracks the BLS PPI for drilling oil and gas wells, series PCU213111213111.

Figure 8: Model fit for mean EUR of initial wells and later development wells

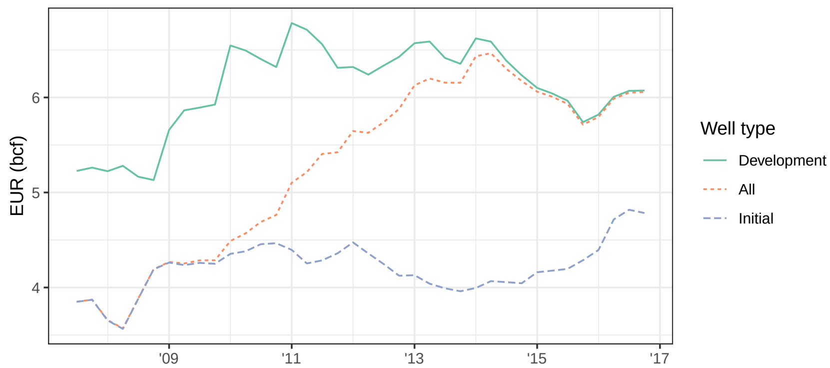


Red 'Model simulation' line represents mean EUR predicted by model given royalty rates and prices. Blue points represent EURs calculated based on estimated well fixed−effects from production data, and blue 'Well FE' line is an estimate of the mean of the points. Gray band is 95% confidence interval.

close when averaged over the entire year.

In Figure 8, I compare mean EUR estimates based on actual production data (blue) with the path of mean EUR generated by simulating the model given prices and initial conditions (red).[40] Simulating mean EUR over time involves computing the expectation of EUR with respect to the distribution of ψ_i^1 conditional on royalty rates, prices, and lease terms: $\mathbb{E}\left[Q_{iw,240} \middle| x_{ri}, r_i, \{z_{it}\}_{t=1}^{T_i}, \{s_{ij0}, Pr(j|i)\}_{j=1}^{J_i}\right]$. This is a very demanding test of model fit. The simulation uses only leasing and price information (not drilling decisions) to predict both drilling and production outcomes. Model-predicted mean EUR is represented as a red line in Figure 9. For the initial wells, the model-predicted mean EUR is close to the actual mean EUR computed using well fixed effects. The model predicts that mean EUR for development wells is higher than what we see empirically, indicating that production data provide important additional statistical information.

Figure 9: Model-predicted mean EUR over time



Simulations are based on estimated parameters. They condition on actual royalty rates and the path of prices.

6.2 Why mean EUR rose

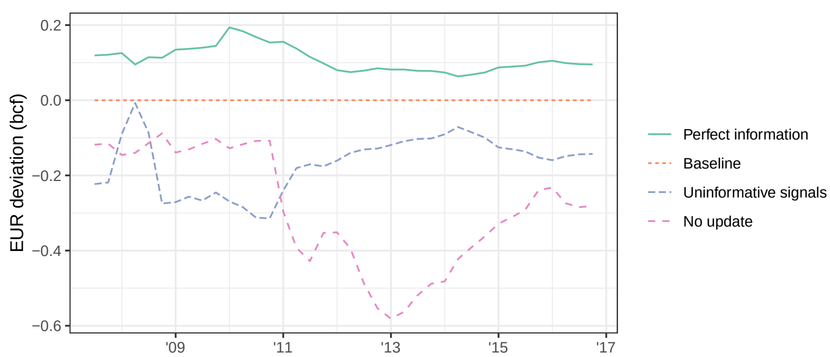
Figure 9 shows the the large effect that selection can have on mean EUR. The top and bottom lines represent the simulated mean EUR of development and initial wells. They trend upward, but at a moderate pace. The middle line represents the mean EUR of all wells—initial plus development. Its rise of 1.5–2 bcf reflects a pure selection effect: a one-time transition from initial to development drilling. The transition reflects mineral lease expirations and learning about geology. The separate panes of Figure 9 suggest that technological progress causes only a mild rise in output per well.

To further understand the way learning about geology impacts overall mean EUR, I simulate three counterfactual informational environments. In the first, firms have perfect information, so the correlation of signal and actual productivity is perfect: $\rho(\psi^0, \psi^1) = 1$. In the second, firms have totally uninformative signals ($\rho = 0$) and learn the maximum amount upon drilling. In the third, firms are unable to update their signals: drilling provides no new information, and firms are stuck with $\psi_{it} = \psi_i^0 \; \forall t$.

In Figure 10, I plot the deviation of the three counterfactual mean EUR

[40] I compute EURs using a common nonlinear cumulative production trend and well-specific fixed effects (see Appendix C.1). Blue points in Figure 9 represent each well on the date it was drilled (spudded) versus its EUR. The blue line is a smoothed mean of these well-specific EURs.

Figure 10: Deviation of counterfactual mean EUR under alternate informational environments from baseline

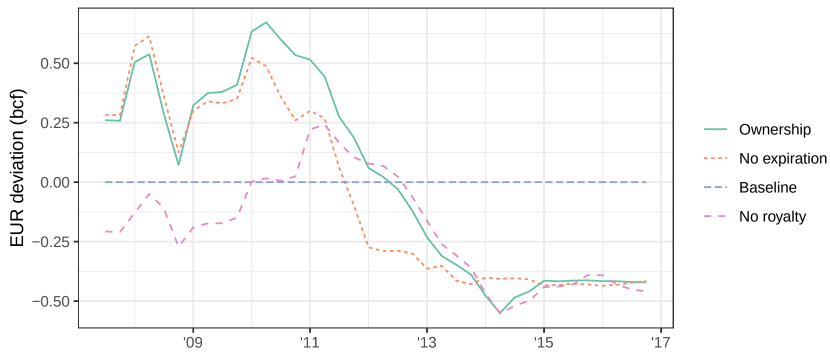


Simulations shown are in deviations from baseline simulations with estimated parameters.
All simulations condition on actual royalty rates and the path of prices.

paths from the baseline mean EUR path ("All" in Figure 9). Positive values imply that counterfactual mean EUR lies above baseline estimates, and *vice versa*. Similarly, positive slopes imply that mean EUR is rising faster than baseline estimates. Changes to firms' information about geology modify the path of mean EUR. Providing firms perfect information raises mean EUR in every period compared to the baseline world. The overall rise happens because noisy signals make firms drill bad locations in search of good ones. Firms also fail to drill some good ones they believe to be bad. In the second case of uninformative signals, firms learn more about geology from drilling an initial well. Mean output per well increases slightly faster over 2009–2014 compared to the baseline scenario. Finally, when firms can make no update to their initial signals, mean EUR rises more slowly starting in 2010 than in the baseline scenario, and ends up a little more than 0.1 bcf lower—a minor difference. Out of the different information scenarios, the no update scenario differs the most from the baseline scenario. Even this change, however, can only explain a small portion of the total predicted increase in mean EUR over the 2008–2016 period.

Distortions induced by mineral lease contracts matter far more to average output per well than does learning about geology. I compare mean EUR under three counterfactual lease contract structures with baseline mean EUR that use

Figure 11: Deviation of counterfactual mean EUR under alternate mineral lease contracts from baseline



Simulations shown are in deviations from baseline simulations with estimated parameters.
All simulations condition on actual royalty rates and the path of prices.

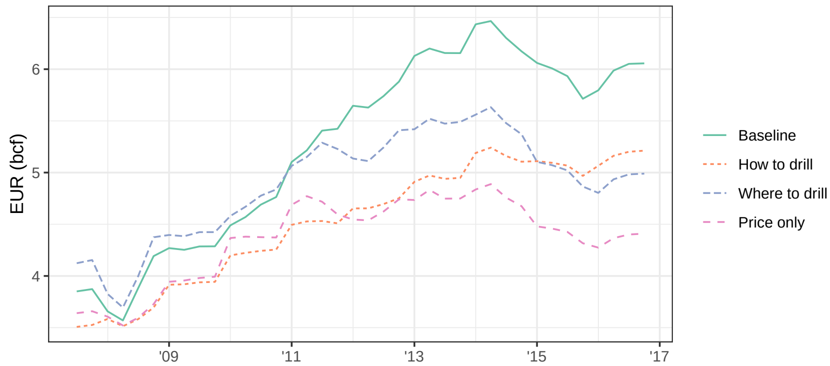
actual mineral lease contracts. Figure 11 shows the deviation of counterfactual mean EUR from the baseline scenario. In the first counterfactual, firms have full ownership of the minerals: no royalty rates or lease expirations distort their incentives.[41] In the second counterfactual, firms pay royalty rates but leases do not expire. In both of these scenarios, mean EUR rises more slowly than in the baseline scenario: mean EUR starts higher compared to baseline and ends 0.4 bcf lower. In the third counterfactual, I eliminate royalty rates. The level of mean EUR generally decreases as firms are able to drill lower-quality locations.

To summarize the relative importance of changes in *where* firms drilled (resource quality) and *how* firms drilled (technology) for the path of mean EUR, I compare four scenarios. For a reference point, I simulate the path of mean EUR under a price only scenario that eliminates learning about geology, mineral lease expirations (but not royalty rates), and technological progress.[42] The top, baseline scenario (corresponding to the middle line in Figure 9) produces the maximum increase in mean EUR by including learning, lease expirations, and technology. Together, the changes in where and how firms drilled raised

[41] Operationally, I remove expiration dates by modifying the transition function for the leasing-drilling state, s_{it}.

[42] Specifically, I eliminate learning by by disallowing updates to firms' noisy signals so that $\psi_{it} = \psi_i^0 \forall t$.

Figure 12: Effects of where vs how firms drill on mean EUR



Simulations are based on estimated parameters. They condition on actual royalty rates and the path of prices.

mean EUR by over 1.5 bcf relative to the price only world. The third path simulates a where-to-drill world in which learning about geology and lease expirations affect firms' choices, but technology is fixed at 2007 levels ($\alpha_t = 0$). In this scenario, mean EUR initially increases rapidly along with the baseline scenario. In 2011, the increase slows and mean EUR peaks at a little more than 0.5 bcf above the reference price only scenario. Finally, I simulate a how-to-drill world that allows for technological progress ($\hat{\alpha}_t = 0.022$) but eliminates learning about geology and lease expirations. In this fourth simulation, mean EUR ends up a little more than 0.75 bcf higher than the price-only world.

6.3 Profit and rent implications

In addition to affecting the path of mean EUR, learning about geology, mineral lease contracts, and technological progress also affected firms' profits and realized resource rents. I simulate profits and rents through the last quarter of 2016 and compute their present value using firms' discount rate. I also assume that the demand for gas and the supply of drilling inputs are both perfectly elastic, so that the path of prices is unchanged. Profits are the expectation of (4) times $\hat{\sigma}_\epsilon$, and they include the expected value of the choice-specific shocks,

Table 3: Counterfactual profits, resource rents, and drilling relative to Baseline

	Billion 2009 USD		Wells drilled		
	Profit	Rent	Initial	Development	Total
Baseline scenario	**-1.62**	**6.14**	**1267**	**699**	**1965**
	Difference from baseline				
	Percent		Wells		
	Profit	Rent	Initial	Development	Total
No technology ($\alpha_t = 0$)	-4%	-17%	-176	-163	-339
Information changes					
Perfect information ($\psi_{it} = \psi_i^1$)	31%	12%	-87	-33	-121
Uninformative signals ($\rho = 0$)	-62%	-27%	113	9	122
No update ($\psi_{it} = \psi_i^0$)	-57%	-37%	51	-161	-109
Contract changes					
No expiration or royalties (ownership)	386%	117%	-357	-8	-365
No expiration	200%	-25%	-733	-481	-1214
No royalty	307%	230%	329	800	1129

Baseline and counterfactual simulations are computed using estimated parameters and realized prices, and they integrate with respect to the distribution of ψ^0, ψ^1 conditional on royalty rates. Firm profits and resource rent are present values measured in billion 2009 USD. Profits are after taxes and royalties, and they include all elements of (4). The rent calculation adds taxes and royalties paid to mineral owners. Wells drilled is the expected number of wells drilled by the end of 2016 Q4. Percent changes are calculated as $(x - x_0)/|x_0|$ to account for negative baseline values.

$\mathbb{E}[\epsilon]$.[43] Rents are pre-tax revenues, plus royalty payments, less pre-tax drilling costs. I include $\mathbb{E}[\epsilon]/(1-tax)$ in the rents. I do not include the cost to extend leases since it is a transfer. Table 3 shows the present value of profits, rents, and the number of wells drilled over the sample period 2003 Q3 to 2016 Q4 for eight simulations. The top row of the table displays the baseline estimates in levels, while the bottom rows display deviations from the baseline. My large estimated drilling costs raise concerns about the estimate of σ_ϵ. Therefore, I calculate profit and rent deviations in percentages that are not affected by σ_ϵ.

When I shut down technology ($\alpha_t = 0$), the present value of profits and resource rents fall a surprisingly mild amount given the focus on productivity innovations in fracking. In the second set of counterfactual simulations, I assess the role of firms' information about geology. When firms have perfect information, drilling falls modestly (121 total wells), but profits and rents rise. When firms receive the noisiest possible uninformative signals ($\rho = 0$) but

[43] I compute expected values of ϵ as $\mathbb{E}[\epsilon] = \log \sum \exp\{v_d\} - \sum_{d \in \Gamma(s)} v_d \Pr(d)$ where choice-specific value functions v_d are defined by (11).

can learn about geology, drilling rises as firms search for good locations, but profits and rents fall. When firms can make no update to their initial signals ($\psi_{it} = \psi_i^0 \forall t$), profit, rents, and drilling all fall.

In the third set of counterfactual simulations, I alter mineral lease contracts by eliminating royalty payments, mineral lease expirations, or both. All three changes cause large impacts on profits and rents compared to changes in firms' information or technological progress. Mineral lease expirations and royalty rates increase profits and rents. The increases come from very different places, however. When lease expirations but not royalty rates are removed, firms reduce drilling by more than half, and rents decrease because drilling falls precipitously. This is consistent with what Herrnstadt, Kellogg, and Lewis (2018) find. When royalties are removed, profit and rent increases come from much higher levels of development drilling. Finally, when we make firms the mineral owners profits increase the most. Rents increase less than if we simply eliminate royalties but leave expiration dates.

6.4 Selection correction

The final exercise I conduct is to see how including a selection correction term affects estimates of productivity time trends in a model of log cumulative production (normalized by well length): $\log Q_{iw\tau}/len_{iw}$. The appropriate selection correction is the conditional expectation of ψ_i^1 given royalty rates and the history of drilling: $\mathbb{E}\left[\psi_i^1 \middle| \{d_{it}, z_t\}_{t=0}^T, \{s_{ij0}, Pr(j|i)\}_{j=1}^{J_i}, x_{ir}, r_i\right]$. I return to the initial regression model (1) and I re-estimate it including the selection correction. Results are in Table 4. The first column reproduces the Naive OLS estimates from Table 1, with 7% annual growth in output per well. In the second column, I include the selection correction term in and leave coefficients for OGIP and $\mathbb{E}[\psi_i^1|royalty, drilling]$ unrestricted. The time trend falls from 7% to 5%. Once I impose the restriction that α_g and α_ψ are the same as the structural estimates in Table 2, the unrestricted time trend, α_t, falls to 1% per year—slightly less than what I estimate using the structural model. For comparison, I repeat the results with section-specific fixed effects that suggests no improvement in technology. This emphasizes the danger of not accounting for unobserved resource quality when estimating productivity.

Table 4: Log linear model of cumulative production with selection correction

		With correction		
	Naive OLS	Unrestricted	Impose α_g, α_ψ	Section FE
Spud date (years since July 2008)	0.07	0.05	0.01	0.00
	(0.01)	(0.01)	(0.01)	(0.01)
Log OGIP	0.53	0.43		
	(0.05)	(0.05)		
$\mathbb{E}[\psi_1 \mid royalty, drilling]$		0.06		
		(0.02)		
Num. obs.	112714	112714	112714	112714
Num wells	1799	1799	1799	1799
Num units	1085	1085	1085	1085

Dependent variable is the logarithm of cumulative production per foot from well w in section i after t months of production. Well length is measured as the lower minus the upper well perforation. The sample includes production months 4 through 72. Standard errors are clustered at the section-level to account for serial correlation and within-section correlation. Production month fixed effects control for a common well decline over time. Section fixed effects account for section-specific geology. Estimated parameters $\hat{\alpha}_g = 0.6$ and $\hat{\alpha}_\psi = 0.34$ are from Table 2.

7 Conclusion

Innovation in the production process—how firms extract—certainly played a key role in sparking the U.S. shale boom: it has increased output per well and lowered costs. The focus on studying innovation in the shale extraction process plays into a broader narrative. Innovation offsets the physical limits of natural resources. In other words, technology vanquishes Malthus.

I show that systematic changes in where firms choose to extract shale resources have also played an important role in increasing output per well. These changes are driven by economic fundamentals—prices, mineral lease contracts, and information about the resource distribution.

While mineral lease contracts distort firms' incentives and reduce resource rents, the structure of most private mineral leases is fairly efficient from the perspective of a revenue-maximizing, liquidity constrained principal (Herrnstadt, Kellogg, and Lewis 2018). It seems doubtful that some kind of policy intervention to remove this distortion is warranted. Improving firms' information sets would increase resource rents, but the effect of doing this would be small compared to changing mineral lease contracts.

The key policy insight of this paper is a cautionary tale for forecasters who might extrapolate past increases in output per well into the future. It is difficult to replicate natural resource quality across space. Should we implicitly assume that we can, our forecasts may be overly optimistic. It is possible that Malthus might bite back.

References

Agerton, Mark, Peter R. Hartley, Kenneth B. Medlock, and Ted Temzelides (2017). "Employment Impacts of Upstream Oil and Gas Investment in the United States". *Energy Economics* 62, 171–180.

Anderson, Soren T., Ryan Kellogg, and Stephen W. Salant (2018). "Hotelling under Pressure". *Journal of Political Economy* 126.3, 984–1026.

Andreatta, G. and G. M. Kaufman (1986). "Estimation of Finite Population Properties When Sampling Is without Replacement and Proportional to Magnitude". *Journal of the American Statistical Association* 81.395, 657–666.

Asker, John, Allan Collard-Wexler, and Jan De Loecker (2019). "(Mis)Allocation, Market Power, and Global Oil Extraction". *American Economic Review* 109.4, 1568–1615.

Baumeister, Christiane and Lutz Kilian (2016). "Understanding the Decline in the Price of Oil since June 2014". *Journal of the Association of Environmental and Resource Economists* 3.1, 131–158.

Bhattacharya, Vivek, Andrey Ordin, and James W Roberts (2018). "Bidding and Drilling Under Uncertainty: An Empirical Analysis of Contingent Payment Auctions".

Bickel, Peter J., Vijayan N. Nair, and Paul C. C. Wang (1992). "Nonparametric Inference Under Biased Sampling from a Finite Population". *The Annals of Statistics* 20.2, 853–878.

Brown, Jason, Peter Maniloff, and Dale Manning (2018). *Effects of State Taxation on Investment: Evidence from the Oil Industry*. SSRN Scholarly Paper ID 3246052. Rochester, NY: Social Science Research Network.

Browning, John, Svetlana Ikonnikova, Frank Male, Gürcan Gülen, Katie Smye, Susan Horvath, Carl Grote, Tad Patzek, Eric Potter, and Scott W. Tinker (2015). "Study Forecasts Gradual Haynesville Production Recovery before Final Decline". *Oil & Gas Journal* 113.12, 64–71.

Cairns, Robert D. (1990). “The Economics of Exploration for Non-Renewable Resources”. *Journal of Economic Surveys* 4.4, 361–395.

Çakir Melek, Nida, Michael Plante, and Mine K. Yucel (2018). “The U.S. Shale Oil Boom, the Oil Export Ban and the Economy: A General Equilibrium Analysis”. *The Federal Reserve Bank of Kansas City Research Working Papers.*

Cochener, John (2010). “Quantifying Drilling Efficiency”, 16.

Cong, Lin William (2019). *Auctions of Real Options*. SSRN Scholarly Paper ID 2136359. Rochester, NY: Social Science Research Network.

Cosgrove, Brendan M., Daniel R. LaFave, Sahan T. M. Dissanayake, and Michael R. Donihue (2015). “The Economic Impact of Shale Gas Development: A Natural Experiment along the New York / Pennsylvania Border”. *Agricultural and Resource Economics Review* 44.2, 20–39.

Covert, Thomas (2015). “Experiential and Social Learning in Firms: The Case of Hydraulic Fracturing in the Bakken Shale”. SSRN Scholarly Paper. SSRN Scholarly Paper. Rochester, NY.

Cuddington, John T. and Diana L. Moss (2001). “Technological Change, Depletion, and the U.S. Petroleum Industry”. *The American Economic Review* 91.4, 1135–1148.

Decker, Ryan A, Meagan McCollum, and Gregory B Upton Jr (2018). “Firm Dynamics and Local Economic Shocks: Evidence from the Shale Oil and Gas Boom”.

EIA (2012). *Pad Drilling and Rig Mobility Lead to More Efficient Drilling - Today in Energy - U.S. Energy Information Administration (EIA)*. Today in Energy.

Fetter, T. Robert, Andrew L Steck, Christopher Timmins, and Douglas Wrenn (2018). *Learning by Viewing? Social Learning, Regulatory Disclosure, and Firm Productivity in Shale Gas*. Working Paper 25401. National Bureau of Economic Research.

Feyrer, James, Erin T. Mansur, and Bruce Sacerdote (2017). “Geographic Dispersion of Economic Shocks: Evidence from the Fracking Revolution”. *American Economic Review* 107.4, 1313–1334.

Fitzgerald, Timothy (2015). “Experiential Gains with a New Technology: An Empirical Investigation of Hydraulic Fracturing”. *Agricultural and Resource Economics Review* 44.2, 83–105.

Gülen, Gürcan, Svetlana Ikonnikova, John Browning, Katie Smye, and Scott W. Tinker (2015). “Production Scenarios for the Haynesville Shale Play”. *SPE Economics & Management* 7.04, 138–147.

Hausman, Catherine and Ryan Kellogg (2015). “Welfare and Distributional Implications of Shale Gas”. *Brookings Papers on Economic Activity*, 71–125.

Hendricks, Kenneth and Dan Kovenock (1989). “Asymmetric Information, Information Externalities, and Efficiency: The Case of Oil Exploration”. *The RAND Journal of Economics* 20.2, 164.

Hendricks, Kenneth and Robert H. Porter (1996). “The Timing and Incidence of Exploratory Drilling on Offshore Wildcat Tracts”. *The American Economic Review* 86.3, 388–407.

Herrnstadt, Evan, Ryan Kellogg, and Eric Lewis (2018). “Royalties and Deadlines in Oil and Gas Leasing: Theory and Evidence”.

Hitaj, Claudia, Jeremy G. Weber, and Ken Erickson (2018). *Ownership of Oil and Gas Rights: Implications for U.S. Farm Income and Wealth.*

Hodgson, Charles (2018). “Information Externalities, Free Riding, and Optimal Exploration in the UK Oil Industry”.

Kaiser, Mark J. (2012). “Profitability Assessment of Haynesville Shale Gas Wells”. *Energy* 38.1, 315–330.

Kaiser, Mark J. and Yunke Yu (2014). “HAYNESVILLE UPDATE—2: North Louisiana Drilling Costs Vary Slightly 2007-12”. *Oil & Gas Journal* 112.1, 54–61.

Kellogg, Ryan (2011). “Learning by Drilling: Interfirm Learning and Relationship Persistence in the Texas Oilpatch”. *The Quarterly Journal of Economics*, qjr039.

— (2014). “The Effect of Uncertainty on Investment: Evidence from Texas Oil Drilling”. *American Economic Review* 104.6, 1698–1734.

Kilian, Lutz (2016). “The Impact of the Shale Oil Revolution on U.S. Oil and Gasoline Prices”. *Review of Environmental Economics and Policy* 10.2, 185–205.

— (2017). “The Impact of the Fracking Boom on Arab Oil Producers”. *The Energy Journal* 38.01.

Komarek, Timothy M. (2016). “Labor Market Dynamics and the Unconventional Natural Gas Boom: Evidence from the Marcellus Region”. *Resource and Energy Economics* 45, 1–17.

Lane, J. Thomas, Britt A. Freund, and J. Breton McNab (2015). “Ch. 25 Held By Production Leases: When Are They Actually Held?” In: *Energy & Mineral Law Foundation.* Vol. 36, 965–1013.

Lee, P. J. and P. C. C. Wang (1983). “Probabilistic Formulation of a Method for the Evaluation of Petroleum Resources”. *Journal of the International Association for Mathematical Geology* 15.1, 163–181.

Levitt, Clinton J. (2009). “Learning through Oil and Gas Exploration”.

Lin, C.-Y. Cynthia (2013). “Strategic Decision-Making with Information and Extraction Externalities: A Structural Model of the MultiStage Investment Timing Game in Offshore Petroleum Production”. *Review of Economics and Statistics* 95.5, 1601–1621.

Lund, Diderik (2009). “Rent Taxation for Nonrenewable Resources”. *Annual Review of Resource Economics* 1.1, 287–308.

Male, Frank, Akand W. Islam, Tad W. Patzek, Svetlana Ikonnikova, John Browning, and Michael P. Marder (2015). “Analysis of Gas Production from Hydraulically Fractured Wells in the Haynesville Shale Using Scaling Methods”. *Journal of Unconventional Oil and Gas Resources* 10, 11–17.

Managi, Shunsuke, James J. Opaluch, Di Jin, and Thomas A. Grigalunas (2004). “Technological Change and Depletion in Offshore Oil and Gas”. *Journal of Environmental Economics and Management* 47.2, 388–409.

Marchand, Joseph and Jeremy Weber (2017). “The Local Effects of the Texas Shale Boom on Schools, Students, and Teachers”. *SSRN Electronic Journal*.

— (2018). “Local Labor Markets and Natural Resources: A Synthesis of the Literature”. *Journal of Economic Surveys* 32.2, 469–490.

Marschak, Jacob and William H. Andrews (1944). “Random Simultaneous Equations and the Theory of Production”. *Econometrica* 12.3/4, 143–205.

Meisner, J. and F. Demirmen (1981). “The Creaming Method: A Bayesian Procedure to Forecast Future Oil and Gas Discoveries in Mature Exploration Provinces”. *Journal of the Royal Statistical Society. Series A (General)* 144.1, 1–31.

Montgomery, J. B. and F. M. O’Sullivan (2017). “Spatial Variability of Tight Oil Well Productivity and the Impact of Technology”. *Applied Energy* 195, 344–355.

Muehlenbachs, Lucija (2015). “A Dynamic Model of Cleanup: Estimating Sunk Costs in Oil and Gas Production”. *International Economic Review* 56.1, 155–185.

Newell, Richard G., Brian C. Prest, and Ashley B. Vissing (2019). “Trophy Hunting versus Manufacturing Energy: The Price Responsiveness of Shale Gas”. *Journal of the Association of Environmental and Resource Economists* 6.2, 391–431.

Ordin, Andrey (2019). “Investment and Taxation: The Case of Oil and Gas in the Permian Basin”. Job market paper. Job market paper. Duke University.

Quyen, N. V. (1991). “Exhaustible Resources: A Theory of Exploration”. *The Review of Economic Studies* 58.4, 777–789.

Redlinger, Michael, Ian Lange, and Peter Maniloff (2019). “Interfirm Learning in Environmental Safety: Evidence from the Bakken”. *Applied Economics* 51.28. _eprint: https://doi.org/10.1080/00036846.2018.1564121, 3081–3090.

Reimer, Matthew N., Joshua K. Abbott, and James E. Wilen (2017). "Fisheries Production: Management Institutions, Spatial Choice, and the Quest for Policy Invariance". *Marine Resource Economics* 32.2, 143–168.

Rust, John (1987). "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher". *Econometrica* 55.5, 999–1033.

Simpson, Ralph David (1999). *Productivity in Natural Resource Industries: Improvement through Innovation.* Washington, D.C.: Resources for the Future.

Smith, James L. (1980). "A Probabilistic Model of Oil Discovery". *The Review of Economics and Statistics* 62.4, 587–594.

— (2013). "Issues in Extractive Resource Taxation: A Review of Research Methods and Models". *Resources Policy* 38.3, 320–331.

— (2018a). "Estimating the Future Supply of Shale Oil: A Bakken Case Study". *Energy Economics* 69, 395–403.

— (2018b). "How and Why Petroleum Leases Are Held by Production: Analysis of a Compound Option". *Land Economics* 94.1, 52–72.

Smith, James L. and Geoffrey L. Ward (1981). "Maximum Likelihood Estimates of the Size Distribution of North Sea Oil Fields". *Journal of the International Association for Mathematical Geology* 13.5, 399–413.

Steck, Andrew L (2018). "Industry Dynamics with Social Learning: Evidence from Hydraulic Fracturing".

Tarui, Nori (2015). "The Role of Institutions in Natural Resource Use". In: *Sustainable Economic Development.* Elsevier, 125–136.

Tauchen, George (1986). "Finite State Markov-Chain Approximations to Univariate and Vector Autoregressions". *Economics Letters* 20.2, 177–181.

Timmins, Christopher and Ashley Vissing (2014). "Shale Gas Leases: Is Bargaining Efficient and What Are the Implications for Homeowners If It Is Not?" Department of Economics, Duke University.

Upton, Gregory and Han Yu (2019). *Local Labor Market Shocks and Earnings Differentials: Evidence From Shale Oil and Gas Booms.* SSRN Scholarly Paper ID 3104693. Rochester, NY: Social Science Research Network.

Van der Ploeg, Frederick (2011). "Natural Resources: Curse or Blessing?" *Journal of Economic Literature* 49.2, 366–420.

Vissing, Ashley (2015). "Private Contracts as Regulation: A Study of Private Lease Negotiations Using the Texas Natural Gas Industry". *Agricultural and Resource Economics Review* 44.2, 120–137.

— (2016). "One-to-Many Matching with Complementary Preferences: An Empirical Study of Natural Gas Lease Quality and Market Power".

ONLINE APPENDIX
Learning Where to Drill: Drilling Decisions and Geological Quality in the Haynesville Shale

Mark Agerton

March 27, 2020

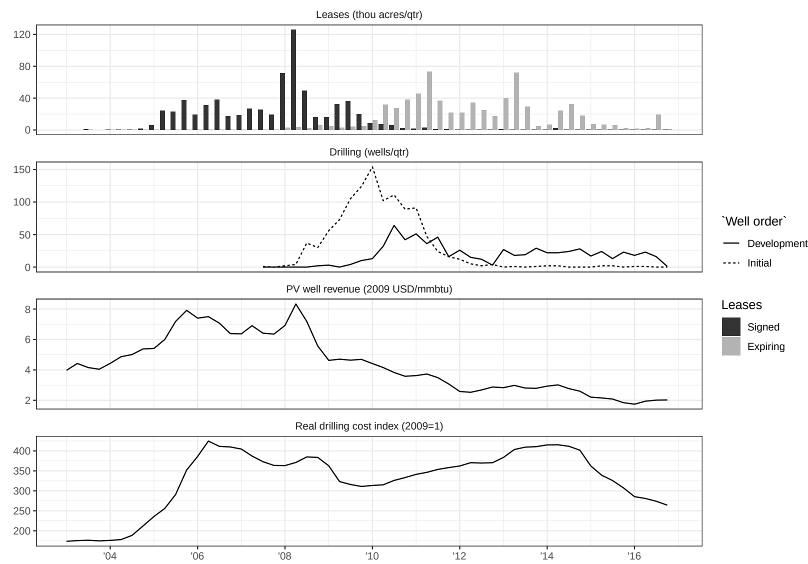
A Data construction

A.1 Merging data

The DNR website has separate shapefiles for the PLSS grid and the drilling units in the Haynesville. Since not all sections have been unitized, I merge these two datasets. Drilling unit polygons tend to fall on a more regular grid compared to the PLSS sections, so I make some small modifications to the PLSS grid so that it aligns better with the Haynesville drilling units. This is done programatically so as to be replicable.

Of the quarter-million wells in the DNR SONRIS database, 29,458 fall within my geographic definition of the Haynesville, which is taken from (Browning et al. 2015; Gülen et al. 2015). I remove 20,469 wells drilled before January 1, 2000, leaving 8,993 wells to be considered. I define wells to be shale wells if the DNR SONRIS database codes them as a "Haynesville well" (a tax designation) or a horizontal well, or if the well is included in the DNR's "Haynesville wells" shapefile. The Haynesville shale formation and the associated unconventional wells are quite deep, so I further exclude wells

Figure 1: Haynesville development over time



shallower than 8700’ as well as those drilled into the shallower Fredericksburg or James Lime formations. I also exclude expired permits to drill, injection wells, and abandoned wells as these will not hold leases by production. I exclude several wells that appear to be double-counted or that appear to be associated with one firm targeting the Cotton Valley in a section when another firm is targeting the Haynesville in the same section. Finally, I exclude two dry wells from my sample. Though this introduces a small bias upwards in production estimates, this is small compared to the more than 1000 wells in my final sample, and these dry wells cannot hold leases by production. This leaves 3,619 Haynesville wells that I will consider.

Merging wells to sections involves matching the overlap of units with the line segments that connect wellheads (the location of the vertical part of the well) and bottom-holes (which terminate at the end of the horizontal part of the well). There are no rules for how firms name their wells, but many name them according to the drilling unit names. I also use this information to merge wells and sections. For all but a very few cases, the name and spatial merges concur, and I examine the others on a case-by-case basis. This method of merging is more accurate than using the wellhead location alone since, as Figure 2 shows, the vertical portion of a well may sit in one section when the horizontal wellbore is actually underneath a neighboring section.

I merge production data from commercial provider Enverus to each well based on the well's API number. While the DNR reports production data, it does so at varying levels of aggregation: the lease, unit, or well. Enverus allocates production streams to appropriate wells accounting for whether multiple wells contribute to the same production stream, natural well decline, and well test volumes.

With the mineral leasing information, I keep 68,795 contracts classified by Enverus as a Lease, Lease amendment, Lease extension, or Memo of Lease. I remove 2,434 contracts classified as Assignment, Lease option, Lease ratification, Mineral Deed, Other, or Royalty Deed.

A.2 Sample Selection

I do not use all of the possible sections in the Haynvesille in my sample. Some of these are missing data, and others appear to differ systematically from sections with drilling that targest the Haynesville. Table 1 tabulates the reasons I drop certain sections, and Figure 2 displays this information visually.

I am missing data for 578 sections: demographics, production or well data, or a royalty rate. The lack of well or production information is unlikely to be random: wells with missing data are likely to be conventional or uncompleted, so I drop these sections. For 1188 sections, I have concerns that firms are not drilling Haynesville wells, or that the lease contracts differ from standard ones. In these sections, firms' decisions do not meet assumptions of my structural model. The first set of reasons I drop sections are that lease terms are nonstandard (or are missing). I drop 331 sections that have leases with

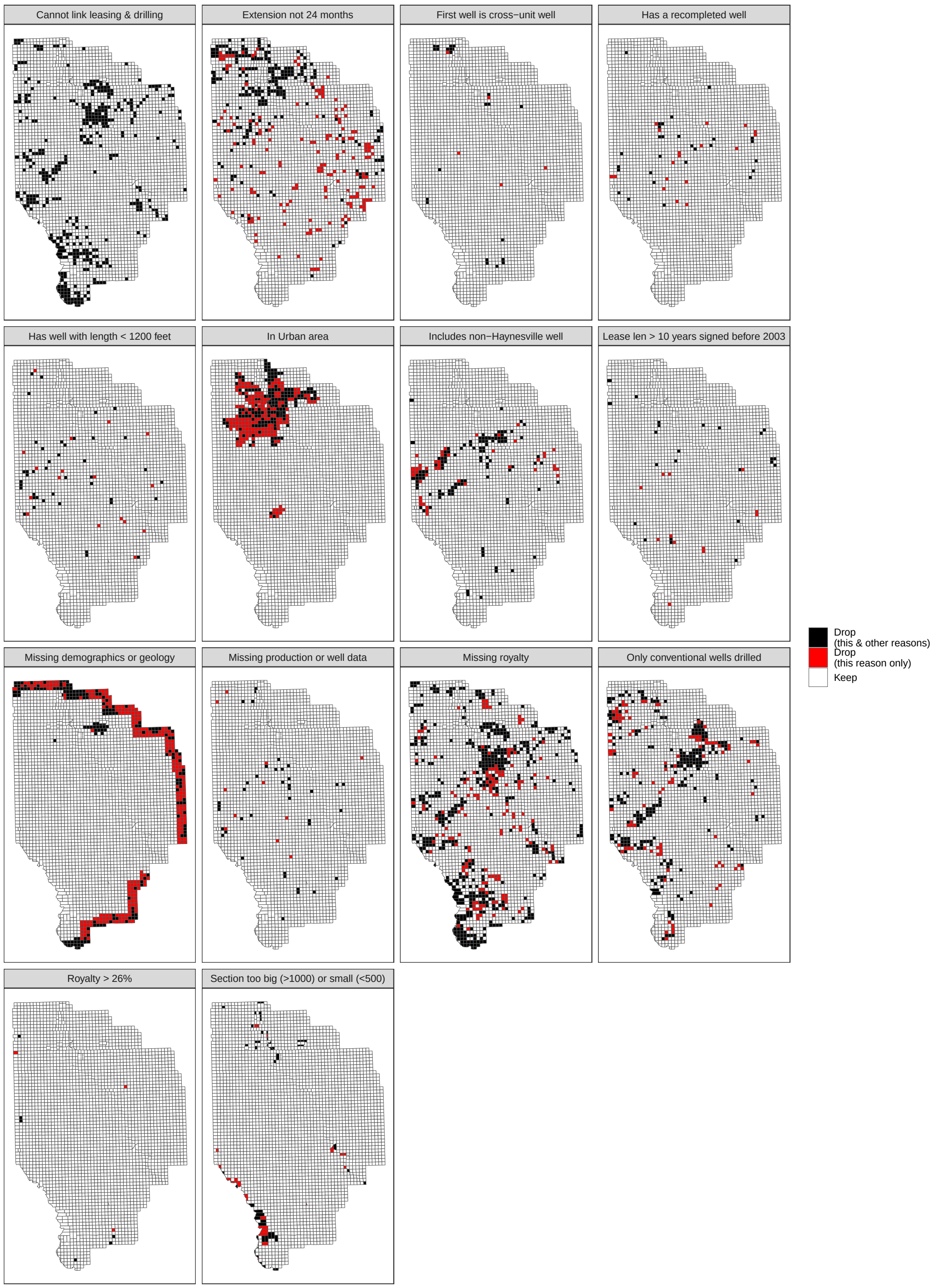


Figure 2: Sections dropped from final sample

extensions that are not 24 months. The vast majority of lease extension are 24 months: landmen talk about a standard "three year lease with a two year 'kicker."' On a practical level, handling additional extension lengths requires significantly enlarging the state-space of the value function I compute and adds to the computational burden. 29 sections have leases longer than 10 years or leases that were signed before 2003. Longer leases are uncommon, and they tend to be on property owned by the government or other large institutions which can more easily place additional requirements on firms. I also exclude the pre-2003 leases, as these pre-date most shale-related activity nation-wide and not likely to be intended for shale development. I remove 330 sections in which the first shale well is not drilled during an identifiable primary term or extension, and 6 leases with unusually high royalty rates (greater than 26%).

The second set of reasons I drop sections are that drilling costs may be quite different, or the firm may not be targeting the Haynesville. I drop 330 sections where only conventional wells are drilled and another 153 in which the shale wells I identify target a formation besides the Haynesville according to Enverus. For 59 sections, at least one well has a lateral that is less than 1200.' This is much shorter than the median 4428' and may also mean the firm is not targeting the Haynesville. I also drop 46 sections with wells that are recompleted after their initial hydraulic fracturing.

The third set of reasons I drop wells is that the incentives to drill may be quite different. I drop 327 sections that are in Shreveport and Mansfield and classified as being in urban areas by the 2010 Census. Urban sections have higher royalty rates and lower drilling activity than the rest of the sample. Drilling in them likely to be more costly than in rural locations, and mineral ownership patterns are likely to be more fragmented. 70 sections are either much larger or smaller than 640 acres. These primarily occur along the border with Texas or in urban areas, and incentives for firms to hold the section with production will be different. For 24 sections, the initial shale well that would

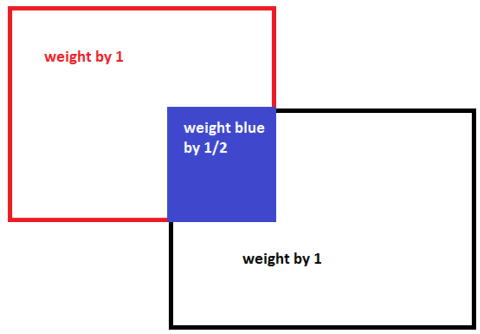


Figure 3: Lease weighting method

hold them with production spans multiple units (a "cross-unit" well). These wells present two challenges. First, they are likely to have different costs and payoffs compared to single wells. Second, they imply spatial correlation between neighboring sections that I do not model, and it is unclear whether I should treat the multiple sections as a single unit before the initial well is drilled.

A.3 Overlapping leases

Lease polygons from Enverus often overlap. There are two reasons for this. First, when multiple grantors sign a lease (say, siblings who inherited mineral rights from deceased parents), Enverus records each lease separately. Second, Enverus draws lease polygons in Louisiana with a minimum area of 40 acres. So, to compute the area of a section that corresponds to a lease, I first compute all spatial intersections of all leases in the section. Then for each lease, I sum over its constituent intersections, weighting each by one over the number of leases also containing that intersection. Figure 3 shows a visual example of this.

B Descriptive statistics

B.1 History of shale activity

For many years, firms knew that gas deposits existed in the Haynesville shale formation but were not able profitably extract the gas. Then, in the early-to-mid 2000s, new technologies allowed firms to start producing gas from a similar, nearby formation, Texas' Barnett shale. Soon, firms' attention turned east towards the Haynesville, and by 2008, a "land-rush" (actually, a mineral rights rush) was on. The panes of Figure 1 plot the history of investment from 2003 to 2016. The top pane shows quarterly mineral leasing when leases expire.[1] The second pane breaks out the number of wells drilled per month by whether a well is the first in its section, or whether it is drilled subsequently. The third and fourth panes show the expected real revenue from an additional unit of total production and a real drilling cost index.

The frenzy of leasing in 2008 coincided with a peak in gas prices, which are shown in the third pane. By the time drilling picked up in 2009, gas prices were falling quickly. While drilling costs dipped as well, the decline was much milder than the fall in gas prices.[2] Despite the fall in output prices, firms increased drilling of initial wells and, to some extent, wells 2–8. Both mineral lease expirations and the value of information provided by initial wells may have have incentivized initial drilling, even if it was unprofitable. The fact that firms did not drill when prices were at their peak suggests that they may have initially faced high internal costs to ramping up a new industrial activity in a new location.

B.2 Descriptive figures

[1] Specifically, it shows when the primary term expires if there is no option to extend in the lease, or when the extension expires if there is one.

[2] The bottom pane shows the PPI for drilling, which generally tracks the proprietary RigData dayrate index.

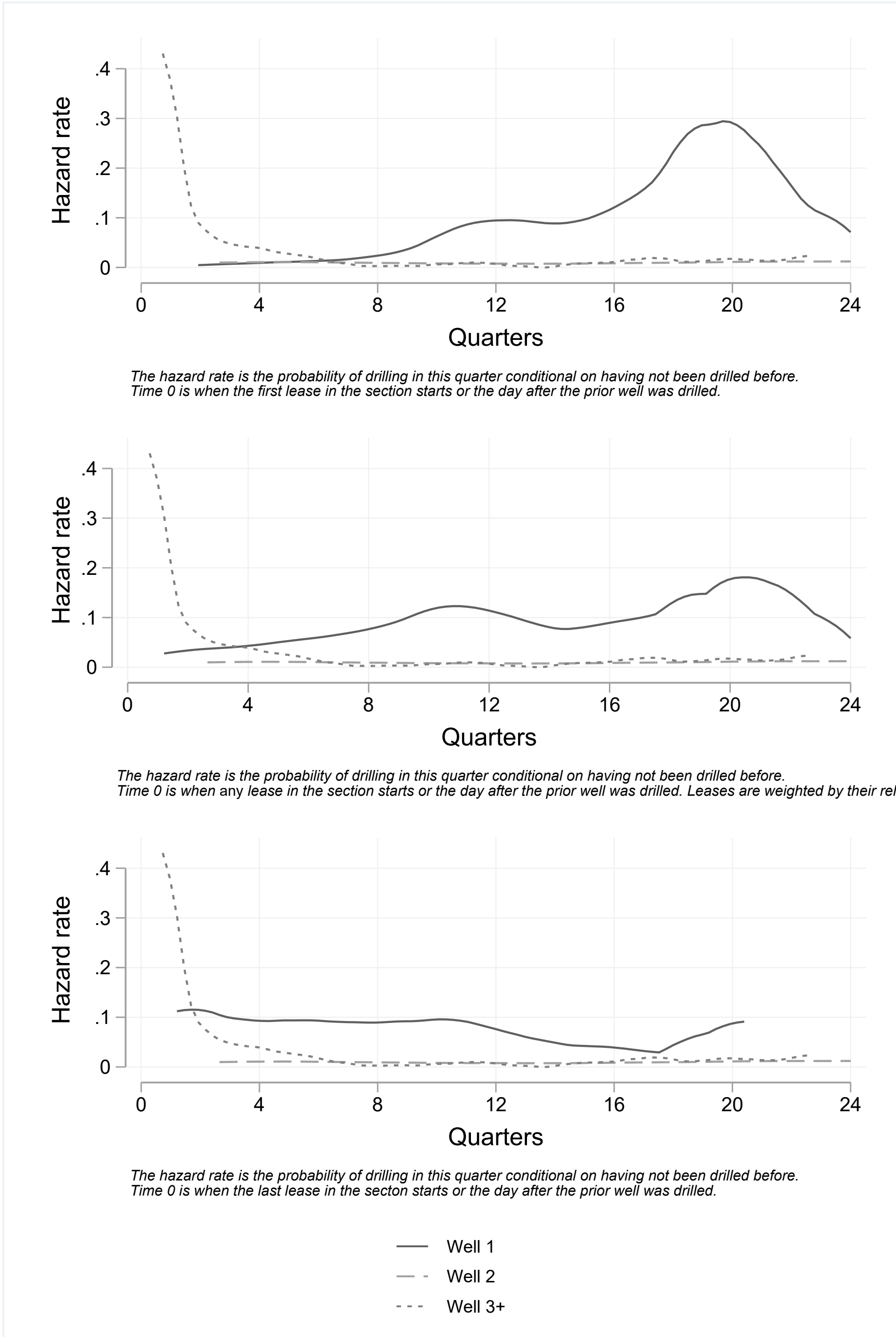
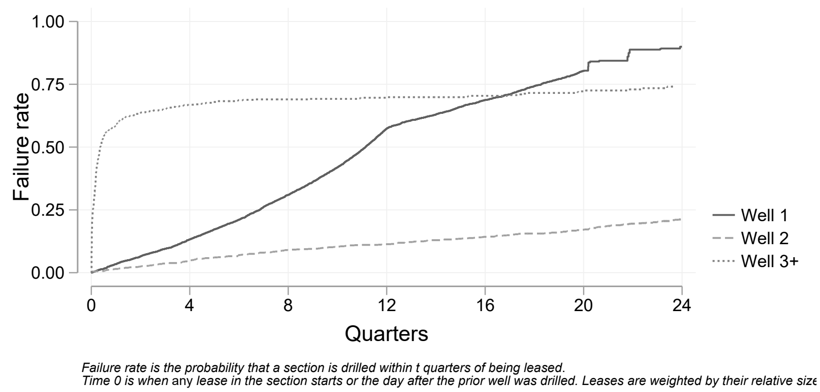


Figure 4: Drilling hazard rates when just the first lease is used, all leases are used, and just the last lease is used

Figure 5: Cumulative weekly failure rate by well-order for 36-month leases



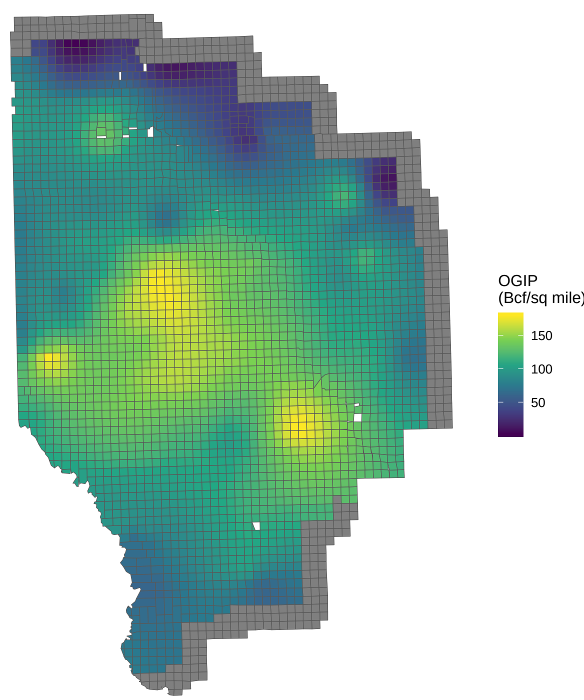


Figure 6: Original gas in place (Gülen et al. 2015)

Table 1: Reasons sections are dropped

	Count	Share
Missing demographics or geology	20	0.01
Missing production or well data	49	0.02
Missing royalty	532	0.19
Dropped for missing data	578	0.21
Unusual leasing		
Extension not 24 months	331	0.12
Lease length > 10 years signed before 2003	29	0.01
No lease when first shale well drilled	330	0.12
Royalty > 26%	6	0.00
Unusual drilling		
Only conventional wells drilled	330	0.12
Well targets Cotton Valley or Other formation	153	0.06
Has well with length < 1200 feet	59	0.02
Has a recompleted well	46	0.02
Unusual incentives		
In Urban area	327	0.12
Section size $\notin (500, 1000)$ acres	70	0.03
First well is cross-unit well	24	0.01
Dropped becuase section history is unusual	1188	0.43
Total dropped	1354	0.49
Total kept	1384	0.51

Shares of reasons why sections are dropped do not sum to one since many sections are dropped for multiple reasons.

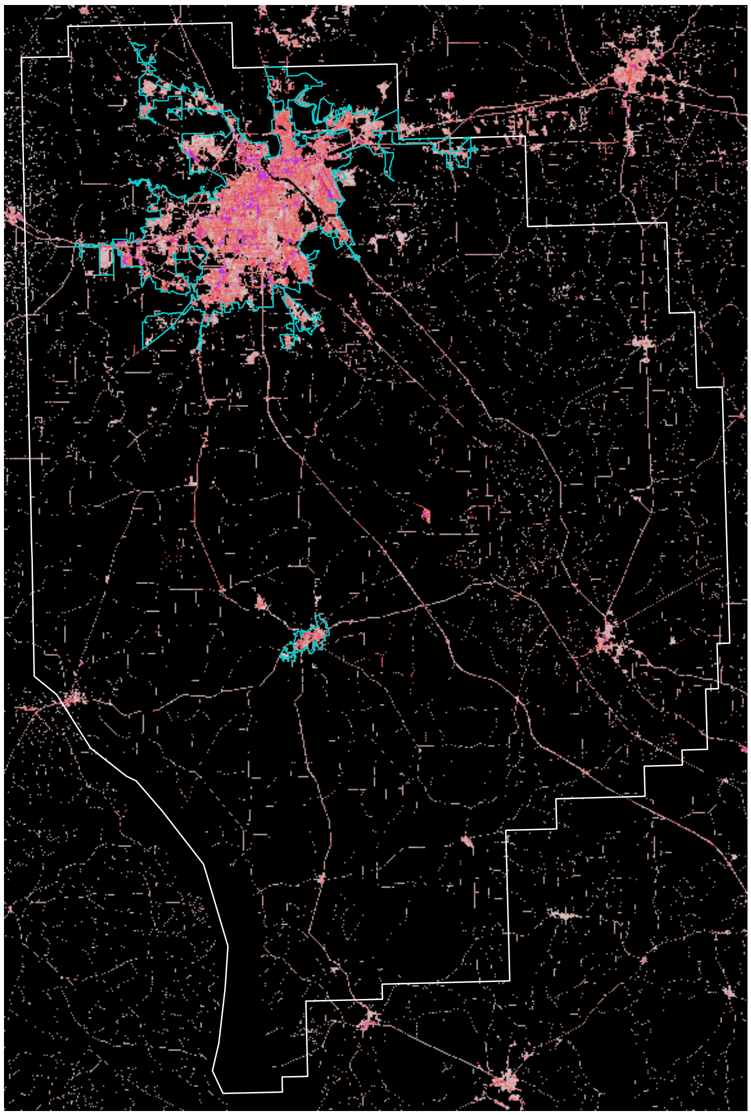


Figure 7: Imperviousness (pink) and urban areas (blue outline)

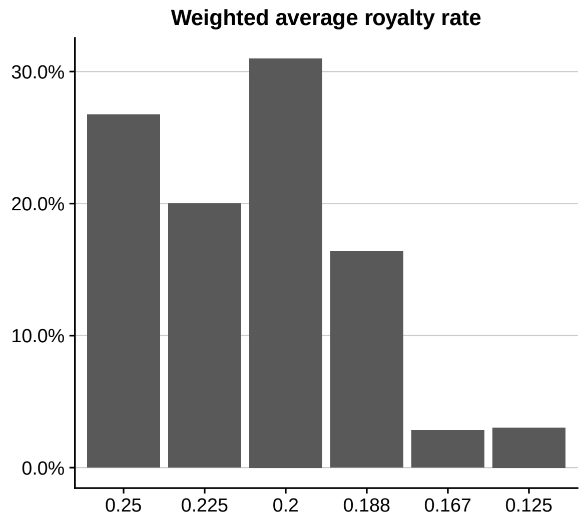


Figure 8: Distribution of discretized, averaged royalty rates r_i (unit-level)

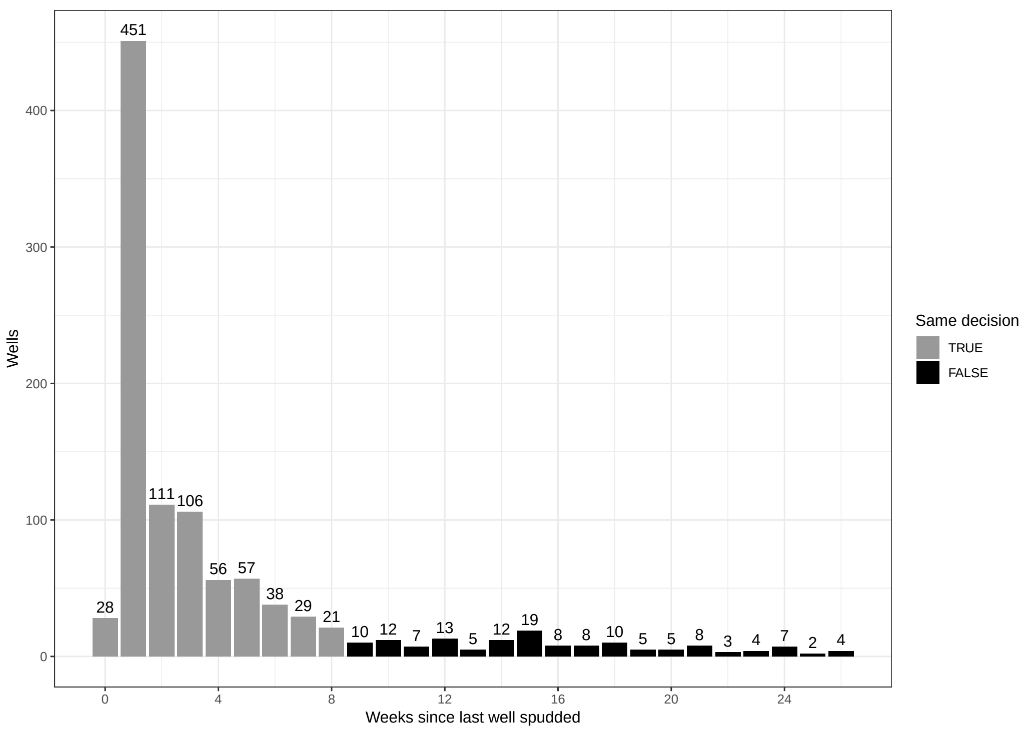
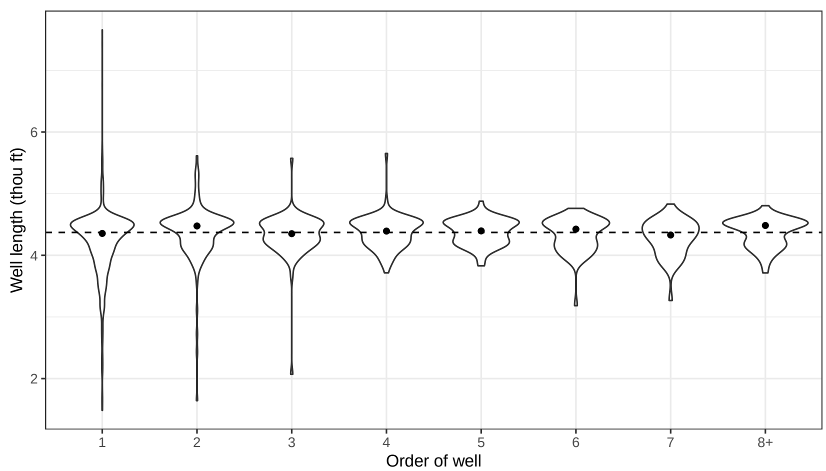


Figure 9: Weeks since previous well drilled

Figure 10: Distribution of well-length



Dots mark category specific medians. Dashed lines mark sample median. Sample excludes cross−unit wells.

Table 2: Summary: Sections

	N	Mean	SD	Min	Q1	Median	Q3	Max
Acres	1384	644.95	37.86	501.98	635.69	642.84	649.48	962.92
Num shale wells	1384	1.40	1.80	0.00	1.00	1.00	1.00	13.00
0 wells	1384	0.23	0.42	0.00	0.00	0.00	0.00	1.00
1 well	1384	0.59	0.49	0.00	0.00	1.00	1.00	1.00
2+ wells	1384	0.18	0.38	0.00	0.00	0.00	0.00	1.00
Number conventional wells	1384	0.62	2.07	0.00	0.00	0.00	0.00	24.00
First lease signed (year)	1384	2006.63	1.25	2003.50	2005.50	2006.50	2007.75	2014.25
Last lease signed (year)	1384	2009.14	1.51	2003.50	2008.25	2009.00	2010.00	2016.00
Number of leases signed	1384	18.58	27.13	1.00	5.00	11.00	22.00	405.00
Blended royalty rate	1384	0.21	0.03	0.12	0.20	0.20	0.25	0.25
Log OGIP	1384	4.67	0.33	2.47	4.53	4.71	4.90	5.19
Log median housevalue	1384	11.22	0.38	9.79	11.04	11.23	11.38	12.60
Log pop. density	1384	2.05	0.90	0.80	1.36	1.88	2.66	5.39
Share of permeable land	1384	0.96	0.05	0.40	0.94	0.97	0.99	1.00
Share of mineral owners OUT of state	1384	0.10	0.19	0.00	0.00	0.00	0.11	1.00
Share of mineral owners IN of state	1384	0.22	0.27	0.00	0.00	0.10	0.39	1.00
Share of mineral owners with address unkown	1384	0.68	0.33	0.00	0.43	0.78	1.00	1.00

Table 3: Summary: Wells

	N	Mean	SD	Min	Q1	Median	Q3	Max
Horizontal well length (ft)	1799	4492.23	905.95	1484.00	4134.00	4428.00	4570.00	9912.00
OGIP (bcf/sq mi)	1799	124.96	26.64	27.08	106.36	125.77	145.83	179.43
Mean royalty rate	1799	0.21	0.03	0.12	0.20	0.20	0.25	0.25
Num units spanned	1799	1.12	0.34	1.00	1.00	1.00	1.00	3.00
1 unit only	1799	0.88	0.32	0.00	1.00	1.00	1.00	1.00
2 units only	1799	0.11	0.31	0.00	0.00	0.00	0.00	1.00
3 units	1799	0.01	0.07	0.00	0.00	0.00	0.00	1.00
Year drilled	1799	2011.31	1.93	2007.67	2010.00	2010.75	2011.75	2016.75
Initial well (vs dev't)	1799	0.61	0.49	0.00	0.00	1.00	1.00	1.00
Haynesville well tax designation	1799	0.95	0.21	0.00	1.00	1.00	1.00	1.00
Permitted as cross-unit well	1799	0.11	0.31	0.00	0.00	0.00	0.00	1.00
DrillingInfo formation = 'Haynesvile'	1799	0.97	0.18	0.00	1.00	1.00	1.00	1.00
Total production (bcf)	1799	4.40	2.04	0.04	3.00	4.11	5.50	15.69
Months of production	1799	84.92	25.35	4.00	73.00	93.00	103.00	127.00
First month of production data (date)	1799	2011.83	1.98	2008.42	2010.50	2011.33	2012.42	2018.17
First month of production data (month)	1799	1.00	0.00	1.00	1.00	1.00	1.00	1.00
Last month of production data (date)	1799	2018.92	1.08	2010.17	2019.17	2019.17	2019.25	2019.25
Last month of production data (month)	1799	84.92	25.35	4.00	73.00	93.00	103.00	127.00

Table 4: Summary: Periods

	N	Mean	SD	Min	Q1	Median	Q3	Max
Before 1st well (Initial drilling)								
Time remaining (including extension)	277320	12.09	5.91	0.00	8.00	12.00	17.00	40.00
Observation is during lease extension	277320	0.16	0.36	0.00	0.00	0.00	0.00	1.00
Num wells drilled this month	277320	0.07	0.27	0.00	0.00	0.00	0.00	8.00
After 1st well (Development wells)								
Drilling last period	27915	0.05	0.22	0.00	0.00	0.00	0.00	1.00
Num wells drilled this month	27915	0.03	0.34	0.00	0.00	0.00	0.00	7.00

Table 5: Summary: Leases

	N	Missing	Mean	SD	Min	Q1	Median	Q3	Max
Is an initial lease	20730	0	0.05	0.22	0.00	0.00	0.00	0.00	1.00
Start (year)	20730	0	2008.21	1.59	2003.50	2007.00	2008.33	2009.42	2016.00
Primary end (year)	20730	0	2011.26	1.62	2006.75	2010.08	2011.33	2012.50	2024.25
Has extension	20730	0	0.79	0.41	0.00	1.00	1.00	1.00	1.00
Extension end (year)	16345	4385	2013.18	1.62	2009.00	2011.83	2013.25	2014.50	2020.83
Primary term (months)	20730	0	36.51	4.84	3.00	36.00	36.00	36.00	120.00
Extension (months)	16345	4385	24.00	0.00	24.00	24.00	24.00	24.00	24.00
Primary + Extension (months)	20730	0	55.44	10.08	3.00	60.00	60.00	60.00	120.00
Has royalty	20730	0	0.23	0.42	0.00	0.00	0.00	0.00	1.00
Royalty	15890	4840	0.22	0.03	0.02	0.19	0.20	0.25	0.75
Royalty $<$ 0.20	15890	4840	0.26	0.44	0.00	0.00	0.00	1.00	1.00
Royalty = 0.20	15890	4840	0.31	0.46	0.00	0.00	0.00	1.00	1.00
Royalty = 0.25	15890	4840	0.41	0.49	0.00	0.00	0.00	1.00	1.00
Is Lease	20730	0	0.79	0.41	0.00	1.00	1.00	1.00	1.00
Is Memo	20730	0	0.20	0.40	0.00	0.00	0.00	0.00	1.00
Is Other Type	20730	0	0.01	0.10	0.00	0.00	0.00	0.00	1.00
Units per lease	20730	0	1.37	1.94	1.00	1.00	1.00	1.00	132.00
Lease within 1 unit	20730	0	0.80	0.40	0.00	1.00	1.00	1.00	1.00
Lease within 2 units	20730	0	0.15	0.35	0.00	0.00	0.00	0.00	1.00
Spatially weighted acreage	20730	0	40.15	222.29	0.20	3.19	8.81	26.91	19067.22
Legal acreage specified on lease	18998	1732	65.97	203.00	0.00	3.16	20.00	60.00	7872.00
Mineral owner is OUT of state	20730	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Mineral owner is IN of state	20730	0	0.28	0.45	0.00	0.00	0.00	1.00	1.00
Mineral owner address unkown	20730	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 6: Summary: Total drilling by geology and royalty

	Original gas in place (Bcf/sq mi)			Royalty						
Total wells drilled	(11.8,100]	(100,125]	(125,179]	0.125	0.167	0.188	0.2	0.225	0.25	All
0	185	62	67	5	10	63	107	59	71	315
1	268	295	257	31	23	118	253	168	227	820
2	19	27	27	5	3	13	17	12	23	73
3	8	6	20	0	0	3	12	12	7	34
4	0	11	20	0	1	6	9	6	9	31
5	1	18	10	0	0	5	7	6	11	29
6	0	5	16	0	0	4	5	6	6	21
7	0	4	18	0	1	6	5	0	10	22
8	1	3	29	0	0	8	12	7	6	33
9	0	0	3	0	1	0	1	1	0	3
10	0	1	0	1	0	0	0	0	0	1
11	0	0	1	0	0	1	0	0	0	1
13	0	1	0	0	0	0	1	0	0	1
All	482	433	468	42	39	227	429	277	370	1384

C Computation

C.1 Production-based EUR calculations

I assume that production from all Haynesville wells shares a common decline curve. In the paper, I compute monthly production decline, cumulative production, and well-specific estimates for EUR. Denote the number of months that a well has produced as τ and estimate a common production decline curve using all months of well production data as

$$\log q_{iw\tau} = \gamma^q_\tau \tau + \gamma^q_{\min\{\tau,72\}} + u^q_{iw} + \eta^q_{iw\tau}. \tag{1}$$

Equation (1) accounts for production decline nonparametrically until month 72, and then assumes a linear decline for months 72–240 following Patzek, Male, and Marder (2013). I am most interested in EUR for each well, which is related to cumulative production, $Q_{iw\tau} \equiv \sum_{s=1}^{\tau} q_{iws}$. Equation (1) implies that cumulative production can be expressed as

$$\begin{aligned} \log Q_{iw\tau} &= u^q_{iw} + h(\tau; \boldsymbol{\gamma}^q, \boldsymbol{\eta}^q_{iw}) \\ h(\tau; \boldsymbol{\gamma}^q, \boldsymbol{\eta}^q_{iw}) &= \log \sum_{s=1}^{\tau} \exp\left\{\gamma^q_\tau s + \gamma^q_{\min\{s,72\}} + \eta^q_{iws}\right\} \end{aligned}$$

Unfortunately, there is no closed for expression for $\mathbb{E}[h(\tau; \boldsymbol{\gamma}^q, \boldsymbol{\eta}^q)]$, even under the assumption that the vector $\boldsymbol{\eta}^q$ is a vector of i.i.d. log normal variables. So, taking the coefficient vector $\hat{\boldsymbol{\gamma}}^q$ from the above estimation, I ignore η_{iwt} and estimate

$$\log Q_{iw\tau} = \gamma_0 + \gamma_h h\left(\tau; \widehat{\boldsymbol{\gamma}^q}, \mathbf{0}\right) + \gamma_{\min\{\tau,72\}} + u_{iw} + \eta_{iw\tau}$$

where $\gamma_{\min\{\tau,72\}}$ and u_{iw} are month-specific and well-specific fixed effects for cumulative production. By including cumulative production month fixed effects, $\gamma_{\min\{\tau,72\}}$, I ensure that errors in my apprximation to $\mathbb{E}\left[h\left(\tau; \widehat{\boldsymbol{\gamma}^q}, \boldsymbol{\eta}^q\right)\right]$

do not affect the quality of my estimates for cumulative production trends over months $\tau \leq 72$. At the same time, I also gain a way to approximate $Q_{iw\tau}$ for future, out-of-sample values under scientifically-based linear production decline. To check the validity of my approximation to well decline over later months 73–240, I test the hypothesis that that $\gamma_h = 1$. I cannot reject it even at the 10% level.

Having verified that my decline curve is valid, I use only months $\tau \in 4, \ldots, 72$ to estimate

$$\log Q_{iw\tau} - h(\tau; \hat{\boldsymbol{\gamma}}^q, 0) = \gamma_{\min\{\tau,72\}} + u_{iw} + \eta_{iw\tau}$$

using production-month fixed effects ($\gamma_{\min\{\tau,72\}}$) and well-specific fixed effects (u_{iw}). The nonlinear trend in cumulative production is

$$f(\tau; \boldsymbol{\gamma}^q, \boldsymbol{\gamma}) = h\left(\tau; \widehat{\boldsymbol{\gamma}^q}, \mathbf{0}\right) + \widehat{\gamma_{\min\{\tau,72\}}},$$

and EUR for well w in section i is simply

$$\mathbb{E}\left[Q_{iw,240} \middle| \{Q_{iw,\tau}\}_{\tau=1}^{T_{iw}}\right] = \exp\{f(240; \widehat{\boldsymbol{\gamma}^q}, \widehat{\gamma}) + \hat{u}_{iw} + \hat{\sigma}_\eta^2/2\}. \qquad (2)$$

C.2 Constructing prices

When evaluating the financial profitability of a well, what firms care about is not the current price of natural gas, but the present value of the price at which the gas will be sold when it is produced. Operators often sell gas production forward, hedging against future price drops and locking in revenues when production commences.[3] Thus, I use a weighted average of the forward curve that incorporates both well decline and time-discounting to capture firms expected production revenue. Let $F(t, t+\tau)$ be the monthly average futures price at time t for gas delivered at time $t+\tau$ where both t

[3] One could also justify this by assuming that the futures market accurately reflects firms' expectations about future prices.

and τ are measured in months. Following Covert (2015), I assume that a shale gas well produces for 20 years. The median number of months between spud date and first production is five, so the relevant wellhead gas price for the firm is a weighted and discounted average of futures prices less costs for gathering, treatment, and compression \$0.49[4] respectively:

$$p_t = \sum_{s=5}^{245} \left\{ \frac{\exp\{\hat{\gamma}_\tau^q (s-5) + \hat{\gamma}^q_{\min\{s-5,72\}}\}}{\sum_{\tau=1}^{240} \exp\{\hat{\gamma}_\tau^q(\tau-5) + \hat{\gamma}^q_{\min\{\tau-5,72\}}\}} \tilde{\beta}^{s/12} \left[F(t,t+s) - 0.49\right] \right\} \tag{3}$$

where $\tilde{\beta}$ is the nominal discount factor, and production decline parameters are estimated using equation (1). The variable p_t then represents the marginal value of an additional unit of expected ultimate recovery (EUR).

Reliable measures of forward prices, $F(t,t+\tau)$, are only available for τ up to 5 years. To account for this, I replace $F(t,t+\tau)$ for years 6–24 with the average 5-year futures price, $\overline{F(t, 5\ year)} = \frac{1}{12}\sum_{m=1}^{12} F(t, 48+m)$. Rather than estimate β, I set it exogenously as is typical in empirical dynamic discrete choice papers. I follow Kellogg (2014), who assumes a nominal discount rate of 12.5% based on a survey of the Society of Petroleum Evaluation Engineers. I also compute average inflation from the average change in the logarithm of the PPI for final goods less energy and food over the sample period Jan 2003–Oct 2016. This is 1.98%. Combining the two, this gives me an annual nominal discount factor of $\tilde{\beta}^{nom} = 1/1.125 \approx 0.89$ and an annual real discount factor of $\beta = 1.0198/1.125 \approx 0.91$, which is close to the value 0.9 used by Covert (2015) and Muehlenbachs (2015) for similar applications, as well as the real discount rate used in Kellogg (2014).

C.3 Transitions for prices

An important element that determines firms' value function is the set of transition probabilities for the time varying exogenous variables, z_{1it}. The firm

[4] I take these from Gülen et al. (2015).

uses these to compute the Emax function, equation (10). I form the transition probabilities in two steps. First, I estimate the parameters that characterize the underlying time series process. Second, I discretize the variable over an evenly spaced grid and create a Markov transition matrix.

I fail to reject unit roots in the logged weighted average price of natural gas, $\log p_t$ computed using (3), and logged drilling dayrate, $\log c_t$. I therefore assume they follow random walks[5] and estimate their covariance matrix Σ_{pc} directly from $\Delta \log p_t$ and $\Delta \log c_t$ using my sample period. The estimated standard deviations are $\hat{\sigma}_p = 0.09005$ and $\sigma_c = 0.06977$, and the correlation of $\Delta \log p_t$ and $\Delta \log c_t$ is $\hat{\rho}_{pc} = 0.3099$.

When I use only gas prices, p_t—not dayrates, c_t—I discretize prices on an evenly spaced grid of 51 points that goes from one-fifth the lowest price in my dataset to five times the highest price.[6] When I include dayrates, $\log c_t$, the size of the state space increases exponentially. This causes difficulties in terms of memory and computational time. So, when I include both gas prices and rig rates, I use only 17 grid points for each dimension allow the grid to extend only $\pm log(2.5)$ beyond the minimum and maximum prices observed.

For the transition matrices for both prices (gas prices and rig rates, if included) and for $\Pr(\psi^1|\psi^0)$, I use the Tauchen (1986) procedure. Many of the elements in the transition matrix for z_{1it} are very small, so I zero out any that are less than 10^{-5}. This allows me to use sparse matrices and helps considerably with computation. I do not zero out elements of the transition matrix for ψ^1.

C.4 Nested fixed point routine

I use a Rust (1987)-style nested fixed point (NFXP) routine to estimate the model. In the inner NFXP loop, I solve the integrated value function by

[5] While diagnostics suggest that $\Delta \log c_t$ has more structure, including a lagged value would expand the state space beyond what is computationally feasible for me to handle. This simplification is unlikely to make much difference in estimation.

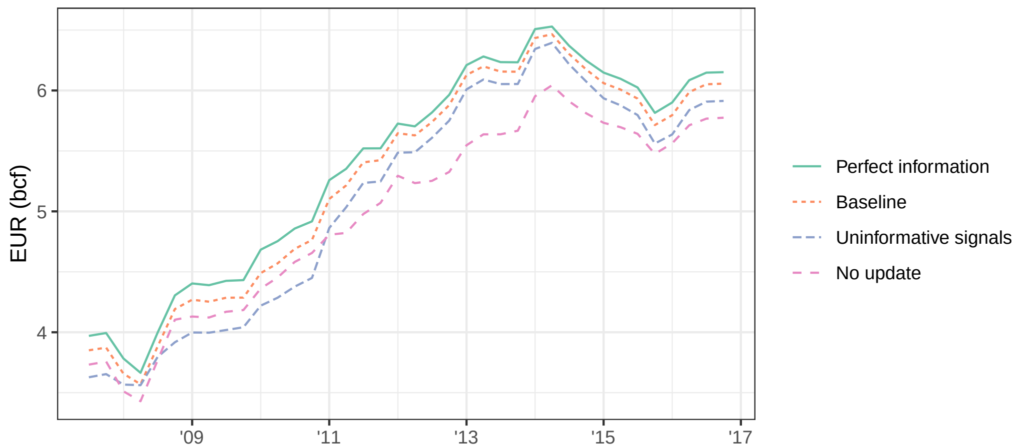
[6] This is the same as in Kellogg (2014).

backwards induction one leasing-drilling state at a time. The leasing-drilling state s_{it} is a tuple $s_{it} = (\tau_{0it}, \tau_{1it}, D_{it})$ where τ captures time-to expiration and D, cumulative prior drilling. These are sorted lexicographically by $-\tau_1$, $-\tau_0$, and D. The implication of this is that the integrated value functions at s_i depend on s_j when $i < j$ but not vice versa. The last element in S, $s_{|S|}$, is the the terminal state at which the firm cannot drill, either because the lease expired or all of the possible wells have been drilled. As stated previously, this is normalized to zero: $\mathbb{E}\mathrm{V}(s_{|S|}, z, \psi) = 0 \quad \forall z, \psi$. Computing $\mathbb{E}\mathrm{V}$ at all s involves computing $\mathbb{E}\mathrm{V}$ at $s_{|S|-1}$, then computing $\mathbb{E}\mathrm{V}$ at $s_{|S|-2}$ using $\mathbb{E}\mathrm{V}$ at $s_{|S|-1}$, and so on.

At all leasing-drilling states s_i with $i < |S|$, the firm's problem is finite horizon if the firm cannot remain at s_i by not drilling. Conversely, it is an infinite-horizon problem if the firm can. I solve finite-horizon problems by value function iteration, and infinite horizon problems by a hybrid iteration algorithm that involves a few initial value function iterations and subsequent policy function iterations until convergence (see Rust (1994)). For each section i, I compute the value function given its time-invariant characteristics, geology and royalty-rates. The state space is large, with between 2 and 8 million elements.

The outer NFXP loops involve searching over the simulated likelihoods for a maximum. The log likelihood of each action depends on the flow-payoffs and the integrated value function that correspond to each action in the action space. I parallelize computation over units. For each action, I re-compute the flow-payoffs given the state variables and evaluate the value function at the appropriate state values. While I discretize random variables to compute the value function, they are, in fact, continuous. When computing payoffs to each action, I interpolate between grid points using quadratic B-splines. For end point conditions, I require continuous second derivatives at the second-from-last knot. I use Monte Carlo integration with two Halton (1960) sequences of bases two and three to integrate out the independent standard normal

Figure 11: Counterfactual mean EUR under alternate informational environments



Simulations are based on estimated parameters. They condition on actual royalty rates and the path of prices.

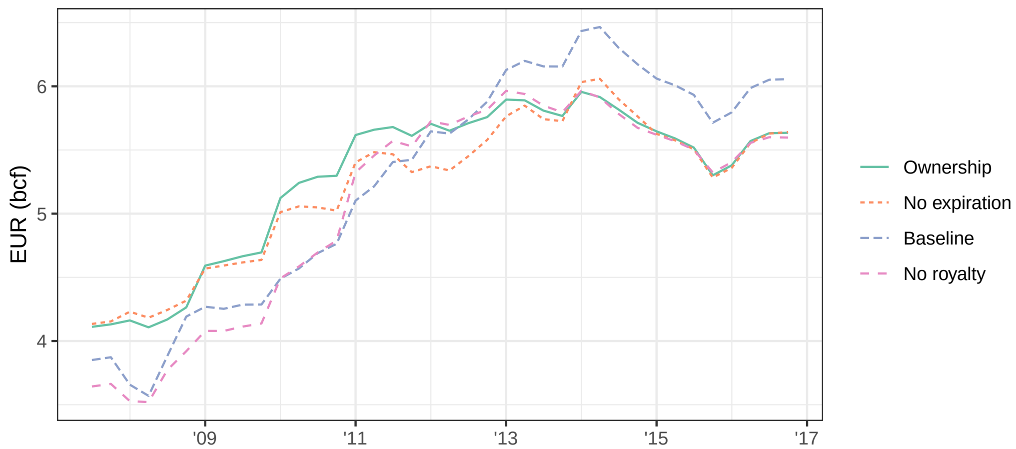
variables u and v. After discarding the first 5000 observations, for each unit i, I draw 2000 pairs of shocks. Results do not change meaningfully if I increase (or decrease) the number of simulated draws.

I obtain starting values by separately estimating each component of the model and then combining them. Closed-form gradients are available for each component of the likelihood, so I use the BFGS Quasi-Newton optimization routine. I calculate standard errors by using the Fisher information matrix.

All of the structural estimation code is publicly available at `https://github.com/magerton/ShaleDrillingLikelihood.jl`. The package includes an extensive set of unit tests to verify accuracy. Outputs are available at `https://github.com/magerton/ShaleDrillingResults`.

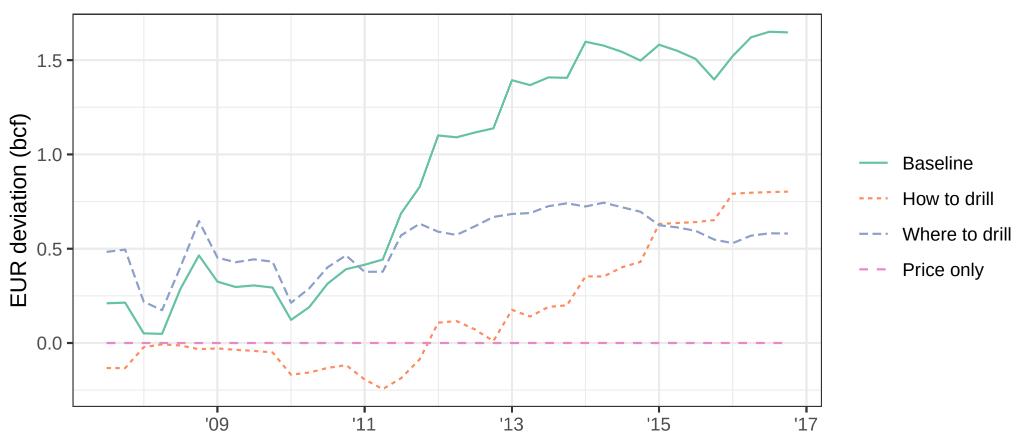
D Simulations: additional figures

Figure 12: Counterfactual mean EUR under alternate mineral lease contracts



Simulations are based on estimated parameters. They condition on actual royalty rates and the path of prices.

Figure 13: Effects of where vs how firms drill on mean EUR (deviations from baseline)



Simulations shown are in deviations from 'Price only' simulations with estimated parameters.
All simulations condition on actual royalty rates and the path of prices.