Productivity Growth in the U.S. Shale Boom: Technology versus Location Choice

Mark Agerton*

December 29, 2019

Abstract

We often link increasing productivity in U.S. shale extraction to innovations in *how* firms drilled. Yet in Louisiana's Haynesville shale, economic incentives—falling output prices, firms' ability to learn about geology, and distortionary contracts—caused firms to shift *where* they allocated drilling efforts over the period 2008–2016. Once I control for this shift, residual productivity trends fall from 7% per year to just 2% per year. What firms learned about geology allowed them to allocate extraction slightly more efficiently over space, increasing resource rents. Mineral lease contracts—a mechanism for mineral owners to capture firms' information rents—caused larger shifts in where firms drilled, inducing misallocation that reduced profits and resource rents by economically significant quantities.

*This paper was previously circulated as "Drilling to Learn: Information, Royalty Rates and Drilling Decisions in the Haynesville Shale" and "Learning Where to Drill: Drilling Decisions and Geological Quality 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, and Dave Rapson (UC Davis); Jevgenijs Steinbuks, Xavier Gine, and Hanan Jacoby (World Bank); Wesley Blundell (CSU East Bay).

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.

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, the resource quality is usually not. When productivity increases, it is difficult to know whether firms got better at how they extracted, or whether they simply located extracted higher quality resources. Just like any other input into a production process, location and resource quality are choices that firms make based on prices and production technology.

Should firms' choices over location systematically vary with productivity innovations, and should we fail to account for it, we may over-estimate the importance of innovation in the production process. This may be particularly problematic in natural resources if extracting high-quality resources today means these are unavailable tomorrow. In such a case, apparent productivity improvements are merely intertemporal shifts of productive capacity. Should we confound productivity gains due to how and where firms extract, we run the risk of biased long-run forecasts.

Two economic factors are especially important in determining where firms locate extraction activities: contracts that govern property rights and information about resource quality. The owner and extractor of a resource are often different. The contracts used to assign property rights over the resource often distort firms' decisions about when and where to extract. The induced misallocation can diminish resource rents below their socially optimal level. If contract terms systematically change firms' incentives over time, this may look changes in productivity. Reimer, Abbott, and Wilen (2017) make this point forcefully in the context of fisheries. Firms' information about the spatial distribution of resource quality is another key determinant of where they locate extraction. As their information improves, firms can more efficiently allocate extraction over space. If aggregate improvements in information are large enough, this may also look like productivity changes.

In this paper, I study how firms choose where to drill and frac wells in unconventional shale formations, and both the productivity and welfare consequences of these choices. Using a structural model and data from Louisiana's

Haynesville shale, I distinguish between the productivity impacts of how and where firms drill. Naive estimates that fail to account for firms' location choices suggest that technology increased output per well by seven percent per year on average. Once I control for firms' location choice, this falls to just two percent.

I find that three economic factors caused firms to drill more productive Haynesville locations over time. First, natural gas prices fell. This increased the level of resource quality required for a well to break even. Second, mineral lease contracts specified use-it-or-lose-it deadlines that applied only to the initial well(s) in a location. Herrnstadt, Kellogg, and Lewis (2018) use a theoretical model to show how the deadlines partially offset the distortion caused by requiring firms to pay royalties, which are themselves a remedy to asymmetric information. Deadlines made firms willing to drill unprofitable locations in order to preserve the option to drill additional development wells. With deadlines met, firms focused drilling on the more productive—and profitable—locations. Third, firms learned about the spatial distribution of geological productivity through drilling. As their information improved, firms modestly improved their ability to target the more productive locations.

Model simulations imply that the distortions introduced by mineral lease contracts have much larger consequences for resource rents than do technological improvements or changes in the firms' information sets. I find that resource rents would have more than doubled were firms to have owned the resource. Asker, Collard-Wexler, and De Loecker (2019) have shown that market power in global oil markets has caused economically significant misallocation of extraction from high quality, low cost deposits to low quality, high cost ones. My findings demonstrate that other distortions cause misallocation in oil and gas extraction at an even finer scale absent market power. Consistent with Herrnstadt, Kellogg, and Lewis (2018), eliminating use-it-or-lose-it deadlines without eliminating royalty payments would have actually reduced resource rents. Eliminating technological innovations would have only decreased resource rents by 17%. Turning to information sets, were firms immediately able to access what they learn about the spatial distribution of geology by drilling without having to drill, rents would only rise around 12%. Eliminating all learning would have lowered rents by around 37%.

There are three reasons to study this issue in the context of shale oil and gas. First, shale extraction is important. The U.S. is the world's top producer of oil and gas[1], and the majority of the country's oil and gas production now comes from shale.[2]

Second, the standard narrative about shale extraction tends to be technology-centric, and largely neglects the role of resource quality (geology). The narrative goes something like this. In the early 2000s, the U.S. energy industry foresaw rising future demand for oil and natural gas, but worried that domestic production would be insufficient to meet demand. Anticipating physical scarcity and higher prices, gas producers innovated. They figured out how to combine horizontal drilling and hydraulic fracturing to extract large volumes of oil and gas from shale formations. Productivity innovations have increased supply. For the U.S. natural gas industry, a primary concern is now how to find demand sources for U.S. supply.

A number of recent papers study how firms learned from their own experiments with well inputs as well as their competitors' experiments (Covert 2015; Fetter et al. 2018; Hodgson 2018; Steck 2018). Others document the degree to which firms learned by doing (Fitzgerald 2015; Seitlheko 2016). These papers shed light on broader questions of how firms learn and how policies like mandatory disclosure can help or hinder the diffusion of knowledge. The studies on productivity innovations in shale control for well location using spatial fixed effects at the nine square mile level (Fetter et al. 2018; Fitzgerald 2015; Steck 2018) or techniques that model spatial correlation of output as geological productivity (Covert 2015; Montgomery and O'Sullivan 2017). None control for the process by which firms select where to drill. Montgomery and O'Sullivan (2017) do find that changes in resource quality have caused a rise in output per well, but they do not investigate why. These previous papers implicitly assume that—conditional on spatial controls—well location is random. In order

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

[2] 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).

to study the underlying mechanisms by which firms learn about a production process, this is a necessary and reasonable simplifying assumption. It is less benign for the purposes of understanding what drives productivity in shale or forecasting.

Third, the data on shale are particularly amenable to studying the relative importance of how and where firms choose to extract natural resources. This is especially true in Louisiana's Haynesville shale. Louisiana partitions the Haynesville into a regular grid of one square-mile *sections* based on a historical land grid. Firms make investment decisions at the level of a section, so I take these as my unit of observation. Each requires around eight wells to fully exploit. This implies that we observe which wells are not drilled, in addition to the ones that are. I assume that all wells in a section share the same geology. Because firms often drill wells in the same section wells years apart, I can identify improvements in technology from within-section growth in output per well. Between-location variation in output per well provides information on the cross-sectional distribution of unobserved geological productivity.

To overcome the fact that production data are not randomly sampled with respect to geology, I exploit data on two more outcomes—royalty rates and the timing of drilling decisions in relation to mineral lease contracts. I model the decision to drill in a Rust (1987)-style dynamic discrete choice framework, also used by Kellogg (2014), Levitt (2009), and Muehlenbachs (2015) to study oil and gas well investments. Tying together the three outcomes—royalty rates, drilling decisions, and production outcomes—allows me to identify the joint distribution of geology and firms' information about geology.

Pairing observed production data with a model of the sampling process (drilling) to estimate an underlying resource distribution is not new (Andreatta and Kaufman 1986; Bickel, Nair, and Paul C. C. Wang 1992; Lee and P. C. C. Wang 1983; Meisner and Demirmen 1981; Smith 1980, 2018a; Smith and Ward 1981). However, previous papers assume that systematic variation in average output per well is due to depletion of better locations. They do not allow for technological change. By exploiting highly detailed data on shale activity, I can accommodate technological change and learning about geology.

Prior work on the economics of learning in Hotelling-style models of nonrenewable resource extraction has identified two ways that new information about geology 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.

The model identifies learning in the following way. Consider a section that a firm has leased. Conditional on exogenous variation in mineral owner characteristics and natural gas prices, high royalty rates plus accelerated initial drilling imply that the firm believes the location is very productive. If the firm learns that location is, in fact, very productive, it should accelerate drilling of subsequent development wells. We will observe large production volumes from all wells in the section. The degree to which we see high royalty rates and accelerated initial wells paired with accelerated development drilling and large production volumes reflects the correlation in firms' initial signals about geological productivity and the actual productivity of a location. This correlation determines how much firms learn about geology.

I exclude any role for strategic information spillovers between adjacent drilling locations from my analysis. Hendricks and Kovenock (1989), Hendricks and Porter (1996), Hodgson (2018), and Lin (2013) all study these issues in offshore drilling. They find that spillovers create strategic incentives for firms to delay drilling until a neighboring competitor reveals new information by drilling. This is important for offshore locations that involve higher geological risk and investment costs.

Fortunately, the issue of information spillovers is likely to be limited in my setting. Geological risk in shale tends to be much lower, damping strategic incentives. Firms will delay only the first of the eight possible wells, and mineral lease expirations further limit the amount a firm can delay drilling. The potential bias introduced by ignoring information spillovers will cause my estimates to understate the true extent of learning and overstate the precision of firms' prior beliefs. The intuition for this is as follows. When firms'

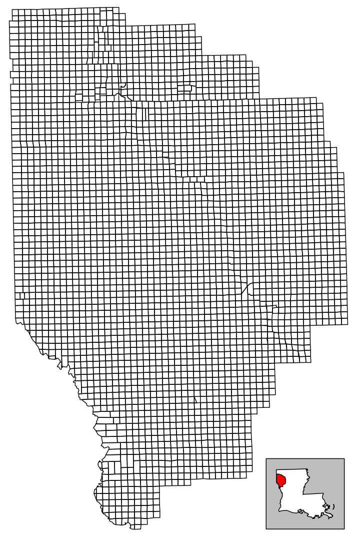


Figure 1: PLSS sections in Louisiana's Haynesville shale

priors are precise, the new information provided by the first well is less valuable. This lowers the economic payoff to drilling and causes firms to further delay exercising the option to drill a first well. We can rationalize such an empirical delay by overstating the precision of firms' signals and understating the informational gains from drilling (i.e., learning). The potential bias from informational spillovers implies that my estimates are a lower bound on the degree to which learning about geology increases average output per well.

1 Institutional details

Louisiana partitions the Haynesville into one square mile (640 acre) blocks called *sections*. Each section requires around eight wells to fully exploit. The partition is based on the Public Land Survey System (PLSS) grid created during the 19th century. Figure 1 shows the PLSS, and the inset map shows the Haynesville's location in northwest Louisiana. When a firm wants to drill a well and extract natural gas section, the State usually forms a *drilling unit* that coincides with the section.[3] While only one firm is allowed to make decisions about a well (the *operator*), all parties with mineral interests in the unit must participate in the well, be they original mineral owners or firms who have leased the mineral rights. The pre-defined, square-mile sections partition the shale into uniform sets of investment opportunities that have one decision-maker. Because shale formations exhibit low permeability, hydrocarbons do not flow into wells from very far away. Thus, wells in one section do not drain hydrocarbons from a neighboring section. This limits the scope for common-pool externalities to affect drilling behavior.

Operators can only drill wells that originate on surface locations under which they have leased mineral rights. Ownership of the mineral rights within a Haynesville section is generally split among multiple private individuals. State-owned minerals are a relatively small share of the Haynesville. Operators normally attempt to lease the majority of a section before drilling.[4] An operator will approach mineral owners and negotiate bilateral mineral lease contracts with each, either directly or through a third-party *landman*. A lease gives the firm the right—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. Rates in the Haynesville range from 12.5% to 25%, with more recent leases tending to be in the 20–25% range.[5] A higher royalty rate can raise the landowner's revenue if the firm drills, but it also reduces the firm's incentive to drill.

Mineral lease contracts and the firm's right to drill a well expire after an initial *primary term*, usually three to five years. Should the firm drill and commence production within the primary term, the lease is considered to

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

[4]Special thanks to Edward B. Poitevent II (Stone Pigman Walther Wittmann L.L.C.), Silas Martin (Drillinginfo), and Cullen Amend (Encino Energy) for assistance with the institutional background of mineral leasing.

[5]Figure 20 in the Appendix shows the distribution of royalty rates in my sample.

be *held by production*, and it enters into an indefinite *secondary term*. The operator maintains the right to drill during the secondary term 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. Since all mineral interests in a drilling unit must participate in each well, each lease in the unit will be held by production, even a well is not physically drilled on each one.

The ability to hold a lease by production implies that the economic payoff to drilling an initial well can be quite large. By drilling an initial well, the firm gains production revenue and new information about geology. If the aggregate improvement in firms' information is large and allows firms to more efficiently allocate extraction over space, it could increase average output per well. The firm also gains the option to drill several more wells at any point in the future. Smith (2018b) shows that even absent informational gains from drilling, the option value of holding a lease by production will induce firms to drill unprofitable locations. Once an operator drills an initial well and holds a lease, the opportunity cost of drilling addtional well increases sharply, and an operator should only drill again if the location is especially productive. The aggregate shift from drilling to hold leases by production to drilling development wells may also increase average output per well.

2 Data

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: 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 University of Texas Bureau of Economic Geology study on the geological quality of the Haynesville shale (Browning et al. 2015; Gülen et al. 2015). The authors estimate a spatial distribution of resource quality that they call "original gas in place" (OGIP). OGIP is based on coarse geological data like the thickness and total organic content of the shale.[6] 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.

The Louisiana Department of Natural Resources (DNR) provides shapefiles of PLSS sections and the areas designated as Haynesville drilling units. I combine these two sets of polygons and use them to partition space into sections. 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 use DNR data on the characteristics and locations of wells drilled in the Haynesville and merge them to sections, and I gather well-level production data from commercial data provider Enverus.[7][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.[9] I use futures prices from Bloomberg, and I follow Herrnstadt, Kellogg, and Lewis

[6]Figure 18 in the Appendix shows a map of the OGIP measure over Louisiana's Haynesville.

[7]Formerly known as Drillinginfo.

[8]Operators in Louisiana can report production by well or by groups of wells in the same lease or unit. Rather than separate out aggregate production to individual wells, I use Enverus' production data, which is allocated to indiviual wells using drilling and well-test data.

[9]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 are drilled by the operator Indigo. All of the wells that I classify as Haynesville wells access sands (formations) which Herrnstadt, Kellogg, and Lewis (2018)-designated wells access.

(2018) and use RigData's index of dayrates for 1000-1499 horsepower drilling rigs in the Arkansas–Louisiana–Texas region. While the cost of drilling and completing a well includes more than renting a drilling rig, movements in the dayrate should be representative of how other costs move over time.[10] I deflate prices and costs to real terms using the PPI for final demand less food and energy.[11] 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.[12] I spatially merge leases to sections. Sections usually contain many mineral leases. In sections that see at least one shale well drilled, I assume that leases which expired before the operator drilled the first shale well or leases that start afterwards did not affect 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. Wellheads (the surface location of the vertical wellbores) are marked by round dots, and these are connected via the purple rays to bottom-holes (the end of the horizontal portion of the well).

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.[13] Average royalty rates are close to the discrete ones, so I map average royalty rates back to the nearest discrete one.

[10] This index closely tracks the BLS PPI for drilling oil and gas wells, PCU213111213111.

[11] Specifically, I use BLS series WPSFD4131 from the FRED database.

[12] 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.

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

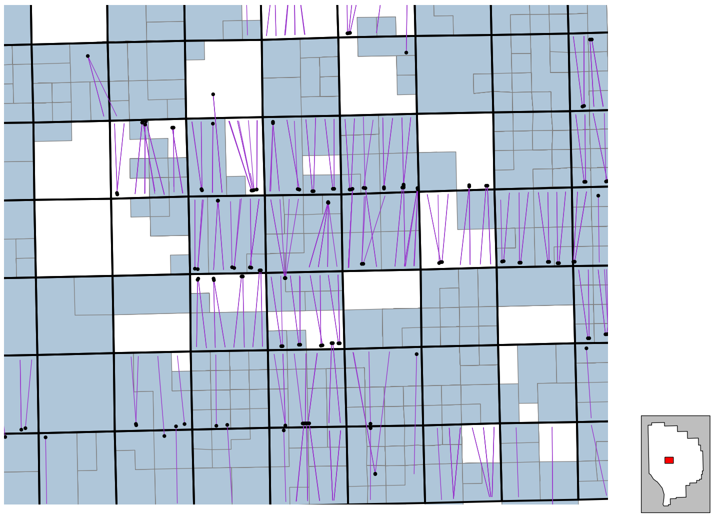


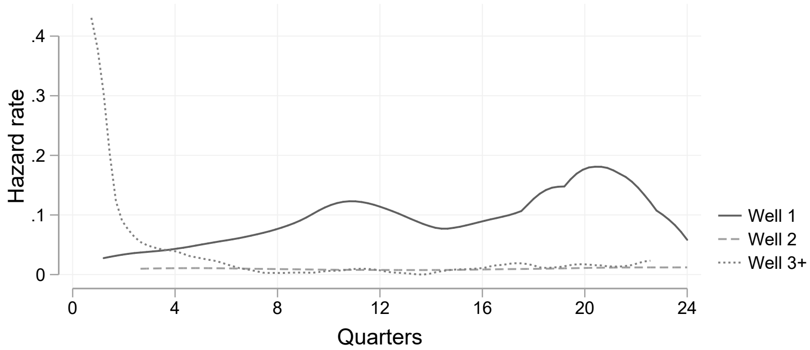
Figure 2: Wells, leases, and sections

Wells that are drilled sequentially and within a short time of one are unlikely to be the result of separate investment decisions and new information. Instead, firms must plan ahead to secure suppliers and regulatory approval. Drilling a well tends to take from two to four weeks, 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 denote any well drilled within 8 weeks (less than 63 days) of another as belonging to the same drilling decision.[14] 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 discusses sample selection in greater depth.

[14]Figure 21 in the Appendix shows the distribution of weeks since the previous well was drilled and where the 8-week cutoff lands.

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.

3 Descriptive evidence

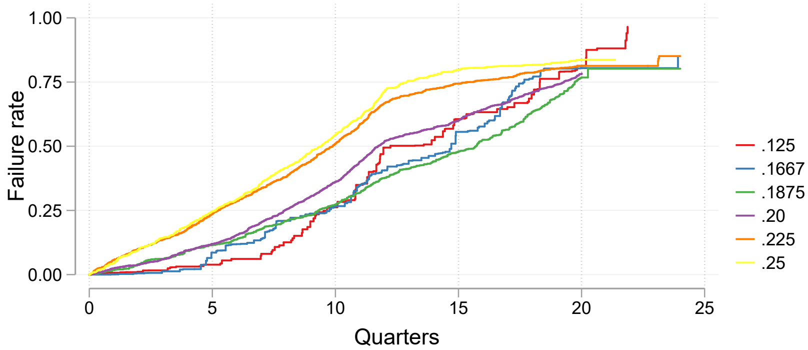
I verify that mineral lease expirations do in fact change firms behavior by estimating nonparametric drilling hazard rates for Well 1, Well 2, and Wells 3+. Since there are multiple leases per unit, I weight each lease by the share of the unit that it owns.[15] Figure 3 shows these estimates. The probability of drilling an initial well peaks when most primary terms and lease extensions expire at quarters 12 and 20 (three and five years). Herrnstadt, Kellogg, and Lewis (2018) find the same result, and they statistically verify that drilling hazard rates drop discontinuously after mineral lease expirations. In contrast, the Well 2 hazard rate is nearly constant. The hazard rate for Wells 3+ suggests that firms tend to either drill immediately after drilling the prior well, or they delay drilling.[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 Wells 2+.

To get a sense as to how well productivity has evolved in the Haynesville, I estimate three preliminary regressions. Each includes a linear time trend associated with the well's spud (drilling) date. The trend captures increases

[15]Figure 16 in the Appendix estimates these rates assuming that that the primary term starts with the first lease signed or, alternatively, the last lease.

[16]This is even more evident in the cumulative failure rate, shown in Figure 17 in the Appendix.

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 months 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

in ouptut per well over time. The dependent variable is cumulative gas production (scaled by the horizontal length of the wellbore) from well w in section i after τ months of production:

$$\log\left(Q_{iw\tau}/len_{iw}\right) = \gamma_0 + \gamma_x^\top x_{iw} + \gamma_\tau + \psi_i + \eta_{iw\tau}. \tag{1}$$

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. The vector x_{iw} includes the date the well was drilled and the OGIP measure from Browning et al. (2015) and Gülen et al. (2015). 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.

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

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.

average royalty rate in the section. The additional variables partially correct for correlation between ψ_i and the drilling date. Estimates imply that sections with multiple wells are 20% more productive than sections with just one well. 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. The additional controls reduce productivity growth by nearly half, falling 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. Now, productivity changes are identified exclusively by comparing wells within the same section over time. The estimated trend in productivity falls to zero.

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.

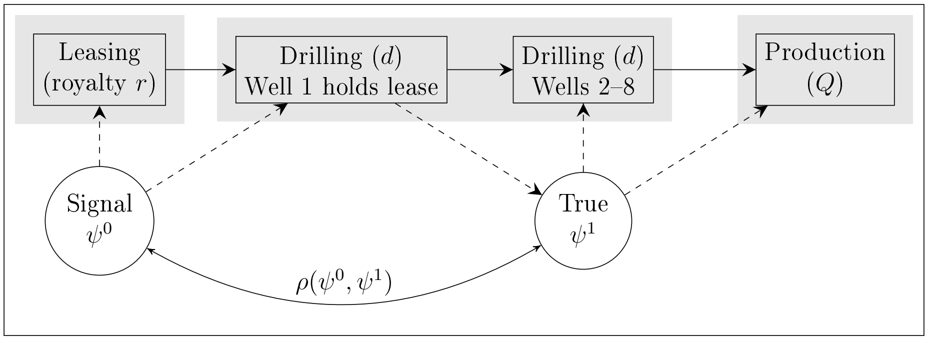


Figure 5: Signal (ψ^0) and true quality (ψ^1) link 3 outcomes in each section

If firms pay higher royalty rates in better locations, they will also accelerate drilling. If high royalty rates are negotiated independent of geology, they will cause firms to 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 rates, the probability that a location is drilled sooner generally 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 mineral lease contracts and learning about geology affect drilling, average output per well, and welfare. To evaluate how these three 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.

[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.

Figure 5 diagrams the sequence of outcomes and the information structure. 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 two statistically independent pieces of information about section i's geological productivity. The first 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. Each quarter, the firm decides how many wells to drill. High signals can cause the firm to accelerate when it drills one or more initial wells. An initial well eliminates the mineral lease expiration (use-it-or-lose it deadline) and perfectly reveals section i's true productivity, ψ_i^1. Knowing ψ_i^1, the firm decides if and when to drill additional wells. Finally, ψ_i^1 and OGIP together determine lifetime production for each well w in section i.

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 firm learns more from drilling initial wells than if signals are precise ($\rho \approx 1$).

4.1 Royalty rates

A royalty rate in section i is a discrete random variable $r_i \in \{\bar{r}_1, \dots, \bar{r}_6\}$. 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 and firms' signal about the location, ψ_i^0. Both can increase the firms' willingness to pay. The second set—mineral owner characteristics, x_{ri}—affect owners' willingness to accept drilling. These include median housing values, the imperviousness of a location's surface (a measure of urbanization), 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 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 $\nu_i \sim F(\nu_i | g_i, x_{ri}, \psi_i^0) = \Phi(\nu_i)$. This effectively means that rates can be modeled 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 of observing a particular royalty rate $r_i = \bar{r}_l$ as

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

[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] That 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.

4.2 Drilling decision

In each section i and each quarter t, a firm decides how many wells to drill, d_{it}. Drilling today affects a firm's ability to drill in the future and potentially its future information.

Denote the endogenous state variable that determines the set of firms' choices as s_{it}. 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 to date, $D_{it} \equiv \sum_{s=0}^{t-1} d_{is}$. 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 transition of the firm's information is

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

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, and the calendar year: $z_{1it}^{\top} = [p_t \quad yr_t]$.[21] The second component, z_{2i}, is time-invariant and

[20] In specifying the state space, I make a simplyfing 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.

[21] Appendix C.3 provides additional details on how I estimate and then discretize

contains the average royalty-rate and the observable component of geology: $z_{2i}^{\top} = [g_i \quad r_i]$. 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, associated with each possible choice of how many wells to drill, d. Examples of these shocks include weather disruptions and availability of a suitable rig in the local area. I assume that shocks, ϵ, are iid, 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 exclusively through ψ_{it}, which is updated once—after the firm drills an initial well.

Given a choice to drill d wells in section i in period t, the firm's static payoff to drilling is

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

Static payoffs are additively separable with respect to the choice-specific shocks. This is standard in the dynamic discrete choice literature. Firms compute net drilling as the product of the number of wells, one minus the royalty rate, natural gas prices less gathering charges, $gath$,[22] and EUR of the wells drilled:

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

$F(z_{1i,t+1}|z_{1it})$.

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

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:[23]

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

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

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 they 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 wells will be higher.

Equations (6) and (7) also include the linear "technology" trend to capture exogenous improvements in production know-how from year-to-year. As in Steck (2018), because technology—and therefore revenue—may be larger in the future, firms have an incentive to delay drilling.

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 prior and subsequent years having

[23] 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.

the same cost as 2008 or 2012, respectively.[24] The cost function is

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

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 $'$ to declutter notation, I write the firm's dynamic program as

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

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 not with the original value function, but with the firm's expectation of the value function in $t + 1$ given its choice in t:

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

Most dynamic discrete choice models assume that firms incorporate future choice-specific shocks, ϵ, into their expectations of future payoffs. I do not. I assume that firms do not anticipate ϵ as Kellogg (2014) does.[25] Define the choice-specific (alternative-specific) value function v_d as

$$v_d(s, z, \psi) = u_d(s, z, \psi) + \beta\,\mathbb{E}\mathrm{V}(s'(s, d), z, \psi). \tag{9}$$

[24] 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.

[25] Doing this 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. This value is especially inflated because of the relatively large number of alternatives that firms choose between (up to 9). See the last column in Table 2.

Given parameters $\boldsymbol{\alpha}$ and ρ, I evaluate $\mathbb{E}$V numerically.

To form the likelihood, I assume that vector of 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.[26] 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, $\Pr(j)$, is equal to the share of the minerals in a section that the lease owns.[27] It is then straightforward to 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 quarters 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 drilling decisions $\{d_{it}\}_{t=1}^{\bar{T}_i}$ in a section conditional on ψ_i^0 and ψ_i^1 is

$$L_i\left(\{d_{it}\}_{t=1}^{\bar{T}_i}\middle|\{z_{it}\}_{t=1}^{\bar{T}_i}, \left\{\{s_{ijt}\}_{t=1}^{\bar{T}_i}\right\}_{j=1}^{J_i}, \psi_i^0, \psi_i^1\right) =$$
$$\left[\prod_{t=T_{1i}+1}^{\bar{T}_i} \Pr(d_{it}|s_{it}, z_{it}, \psi_i^1)\right]\left[\sum_{j=1}^{J_i}\left(\prod_{t=1}^{T_{1i}} \Pr(d_{it}|s_{ijt}, z_{it}, \psi_i^0)\right)\Pr(j)\right]. \quad (10)$$

4.3 Production

The final component of the model consists of monthly production outcomes from each well. The profitability of a well is most closely linked to expected ultimate recovery (EUR), not month-to-month variations in output, so I focus

[26] 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.

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

on cumulative production, $Q_{iw\tau}$ from well w in section i after $\tau \in \{4, \ldots, 72\}$ months of production.[28] I assume that cumulative production, normalized by the horizontal length of the wellbore, is determined a very similar to the regression estimated earlier in Section 3. The estimating equation is

$$\log\left(Q_{iw\tau}/len_{iw}\right) = \gamma_0 + \gamma_\tau + \alpha_g g_i + \alpha_{yr} yr_{iw} + \xi_{iw\tau} \tag{11}$$

$$\xi_{iw\tau} = \alpha_\psi \psi_i^1 + u_{iw} + \eta_{iw\tau}. \tag{12}$$

Equations (11) and (12) 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. Finally, $\xi_{iw\tau}$ is a random effect with three components. The first component 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 $F(u_{iw}, \eta_{iw\tau} | \psi_i^1, g_i, yr_{iw}) = \Phi\left(u_{iw}/\sigma_u\right) \Phi\left(\eta_{iw\tau}/\sigma_\eta\right)$. The likelihood of observing a T_{iw}-length vector of cumulative production is then

$$\begin{aligned} L\left(\left\{\log\left(Q_{iw\tau}/len_{iw}\right)\right\}_{\tau=1}^{T_{iw}} \middle| \psi_i^1, g_i, yr_{iw}; \gamma_\tau\right) = \\ -\frac{1}{2}\left[T_{iw}\log(2\pi) + (T_{iw}-1)\log\sigma_\eta^2 + \log(\sigma_\eta^2 + \sigma_u^2 T_{iw})\right] \\ -\frac{1}{2\sigma_\eta^2}\left[\sum_\tau (u_{iw} + \eta_{iw\tau})^2 - \frac{\sigma_u^2}{\sigma_\eta^2 + \sigma_u^2 T_{iw}}\left(\sum_\tau (u_{iw} + \eta_{iw\tau})\right)^2\right] \end{aligned} \tag{13}$$

where $u_{iw} + \eta_{iwt}$ is defined according to equations (11) and (12).

The coefficients α_g, α_{yr}, and α_ψ are shared by the revenue and production equations (6) and (7). 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 in equation (4). Identification comes from equating firms' beliefs about EUR $Q(g_i, \psi_i^1)$ in equation (4)

[28] 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.

with actual cumulative production in equation (11).[29] A bit of algebra implies that we can compute $\hat{\sigma}_\epsilon$ as

$$\hat{\sigma}_\epsilon = \exp\left\{\hat{\gamma}_0 + \hat{\gamma}_{240} + (\hat{\sigma}_u^2 + \hat{\sigma}_\eta^2)/2 + \log len_{50\%} + \log(1 - tax) - \hat{\alpha}_0\right\}. \quad (14)$$

4.4 Model likelihood

Omitting exogenous variables to reduce notational clutter and replacing sequences of outcomes with vectors, we can write the likelihood conditional on the noisy signal and true quality, ψ_i^0 and ψ_i^1 as

$$L(history_i|\psi_i^0, \psi_i^1) = L\left(r_i\middle|\psi_i^0\right) L\left(\vec{d}_i\middle|\psi_i^0, \psi_i^1\right) \prod_{w=1}^{W_i} L\left(\log \vec{Q}_{iw}/len_{iw}\middle|\psi_i^1\right). \quad (15)$$

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

$$SL(history_i) = \frac{1}{M} \sum_{m=1}^{M} L_i\left(history_i|\psi_{im0}, \psi_{im1}\right). \quad (16)$$

The final statistical I 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 of the sort examined by Hendricks and Kovenock (1989) Hendricks and Porter (1996), Lin (2013), or Hodgson (2018). The simulated likelihood of the entire dataset is $SL(data) = \prod_i SL(history_i)$.

[29] 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 (11).

5 Computation

I calibrate the the firm's nominal annual discount factor to be $\beta = 1/(1+.025)$ 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).[30] 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).

In the second step, I estimate the parameters that characterize exogenous processes for real natural gas prices ($\log p_t$). I cannot reject the null hypothesis that the logarithm of natural gas prices follows a random walk: $\log p_{t+1} = \log p_t + u_{pt}$, so I estimate that $\sigma_p = 0.0900$ and discretize $\log p_t$ over an even grid of 51 points that extend $\pm \log 3$ beyond the minimum and maximum prices I observe. I create a sparse transition matrix based on Tauchen (1986). Many transition are small, so I zero out probabilities less than 1e-5 to minimize 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.

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.[31]

6 Results

Table 2 contains parameter estimates for the full model under six sets of assumptions. The left column displays baseline estimates on which I base coun-

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

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

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			
α_{2008}	-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)
α_{2012}	-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	51	51	51	51	17	51
Num ψ	51	51	51	51	19	51
Num simulations	2000	2000	2000	2000	2000	2000

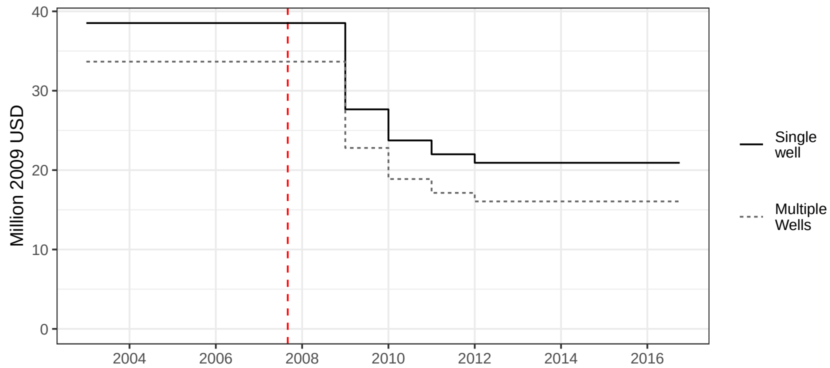
Mean well costs are the average drilling cost for multiple wells over the period 2009–2016. These are measured in million 2009 USD. This is calculated as $\frac{\sigma_\epsilon}{1-\tau_k}\frac{cost(2,s_{it},z_{it})}{2}$ where σ_ϵ is computed from (14). The effective marginal corporate income tax is 40.2%, 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.

terfactual simulations. The right five columns contain a variety of 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 qualitatively similar across these four specifications. The fifth column adds rig dayrates to better capture costs. The liklihood improves mildly, but at a significant computational cost. Finally, the last column assumes that firms anticipate the Type I Extreme Value shocks. Model fit is substantially worse, though signs and magnitudes are largely unchanged. I now restrict attention to the baseline estimates.

The signs of coefficients from the royalty-rate equation, Equation (2), are as expected. The impact of firms' prior signal, ψ_i^0 is positive and statistically significant, so royalty rates are indeed correlated with unobserved heterogeneity in geology. In light of this, it is somewhat surprising that the OGIP coefficient, while positive, is not statistically significant. Coefficients for variables affecting landowners' willingness to accept variables 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. The equations for drilling and production, (6) and (7), share the same coefficients for OGIP, unobserved geological productivity, and time: α_g, α_ψ, and α_t. While the estimated OGIP coefficient $\hat{\alpha}_g = 0.60$ is nearly twice $\hat{\alpha}_\psi = 0.34$, the variance of $\log OGIP_i$ is just 0.33,[32] whereas the variance of ψ_i^1 is 1 by construction. This implies that unobservable differences in geology matter more to drilling and production outcomes than do observable differences in the OGIP measure. The estimated time-trend coefficient, $\hat{\alpha}_t = 0.022$, is lower than the Naive OLS and OLS estimates in Table 1 on page 15 but still larger than the Section FE estimates that throw out much of the variation in the data. The estimated standard deviation of well-specific productivity shocks, $\hat{\sigma}_u = 0.320$, is about the same as the standard deviation of section-specific shocks, $\hat{\alpha}_\psi = 0.340$. This means

[32] See section-level summary statistics in Table 6 in the Appendix.

Figure 6: Drilling costs



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

that geological productivity within a section varies about as much as geological productivity across sections. Firms still face considerable uncertainty about well output even if they perfectly know ψ_i^1. I estimate the correlation of firms' prior signals, ψ_i^0, with actual quality, ψ_i^1, to be $\hat{\rho} = 0.66$. Though firms' initial beliefs are informative, they are by no means perfect, and the information initial wells provide can be valuable.

Substituting in the median well length of 4428' into equation (14),[33] and an effective corporate marginal income tax rate of $tax = 40.2\%$ supplied by Gülen et al. (2015), I estimate that $\hat{\sigma}_\epsilon = 2.00$. 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.

Figure 6 plots the cost to drill a single well and the average cost to drill more than one well based on $\hat{\sigma}_\epsilon$. 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). My estimates include the full

[33]See summary statistics in Table 7.

opportunity cost of drilling—not just direct financial costs. 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. That said, it is also possible that I over-estimate drilling costs. In this case, percent changes in drilling, profits, and resource rents are still meaningful.

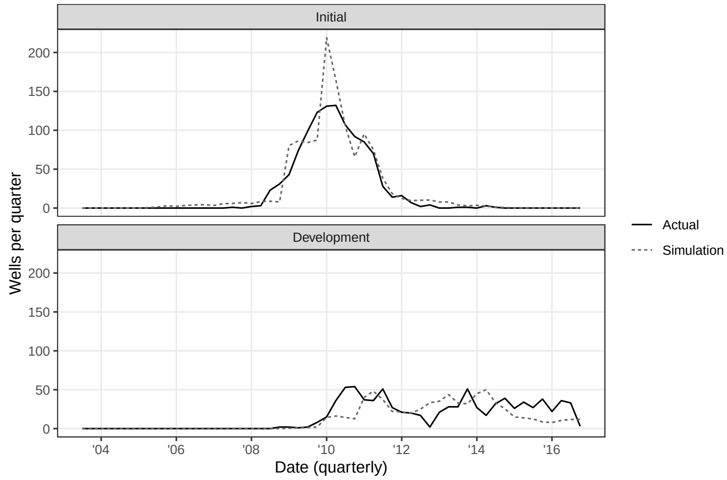
Figure 6 shows a remarkable decline in drilling costs between 2008 and 2009 as the fixed effects drop from $\hat{\alpha}_{2003-08}$ to $\hat{\alpha}_{2009}$. High opportunity costs in 2008–2009 are required to rationalize why firms did not drill when gas prices were at their peak. The year 2008 was the peak of a mineral-rights rush in the Haynesville (see Figure 13 in the Appendix). Focused primarily on leasing minerals during a land rush, firms may not have had the capacity to additionally implement large drilling programs. Industry executives I spoke with described how operators faced new operational challenges as they moved from drilling in Texas' Barnett shale to Louisiana's Haynesville. Compared to the Barnett, the Haynesville is much deeper and characterized by higher pressures and temperatures. Operators needed time to adapt to the difficult new environment. Such difficulties would be captured by a shadow cost in excess of a purely financial cost of drilling.

When I include time-varying rig dayrates, variation in costs due to time fixed effects swamps variation due to rig dayrates. Some of the productivity gains in shale have involved reducing costs, not increasing output per well. For example, firms have dramatically redced the time required to drill a well, and they have learned to schedule drilling and completions more efficiently. These falls in drilling costs would not be reflected in drilling rig dayrates, but they would be reflected in time fixed effects.

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, as expected. Scaled by $\hat{\sigma}_\epsilon/(1 - tax)$ and converted from dollars per section to dollars per acre,[34] it implies that costs to extend mineral leases were approx-

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

Figure 7: Model fit for drilling rates of initial Well 1s and later Wells 2+



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.

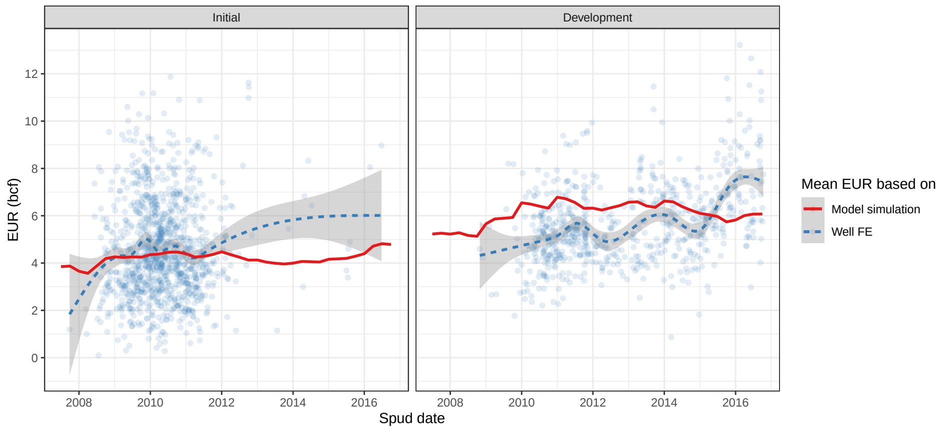
imately \$4805/acre. Costs to extend mineral leases tend to track bonus payments. Gülen et al. (2015) calibrate bonus payments to \$3000/acre, and Kaiser (2012) calibrates them to the \$5000–25,000 range, so my estimate for lease extension costs is within the expected range.

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 these given initial conditions and prices, that is $\sum_i \mathbb{E}\left[d(z_{it}, s_{ijt}, \epsilon_{it}) \middle| \{z^t_{s=0}\}, \{s_{ij0}\}_{j=1}^{J_i} x_{ir}, r_i\right]$. Because 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 the model predictions track actual drilling behavior. Because of the coarse annual time-scale for the technology level, the fit of drilling rates is poor at a quarterly level in 2010, though it is reasonable at an annual level.

Next, I compare EUR estimates based on actual production data with the path of mean EUR generated by simulating the model given prices and initial conditions. Appendix C.1 details how I used well-specific fixed effects to compute well-specific EURs from production data. Blue points in Figure 9

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



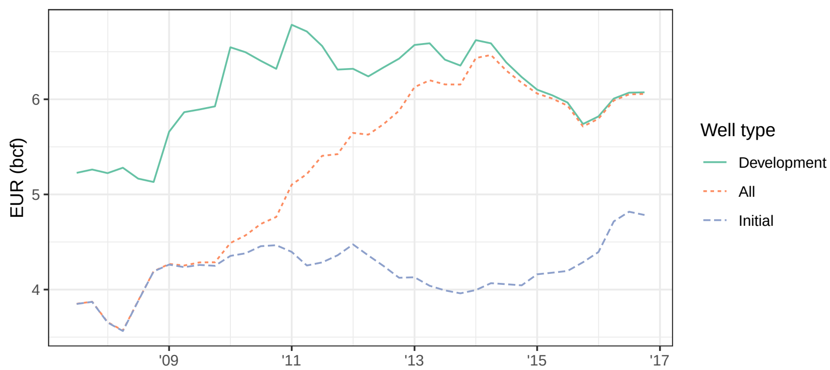
Red 'Model simulation' line represents mean EUR predicted by model given royalty rates and prices. Blue points represent EURs calculated based on estimated well fixed–effects from production data, and blue 'Well FE' line is an estimate of the mean of the points. Gray band is 95% confidence interval.

represent each well on the date it was drilled (spudded) and the well's estimated EUR. The blue line is a smoothed mean of these well-specific EURs. To simulate path of mean EURs over time, I take an expectation of EUR with respect to the distribution of ψ_i^1 conditional on royalty rates, prices, and lease terms: $\mathbb{E}\left[Q_{iw,240}\middle|x_{ri}, r_i, \{z_{it}\}_{t=1}^{T_i}, \{s_{ij0}, \Pr(j)\}_{j=1}^{J_i}\right]$. 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 EURs are 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. However, the model predicts that mean EURs for development wells are higher than what we see empirically.

6.2 Why mean EURs rose

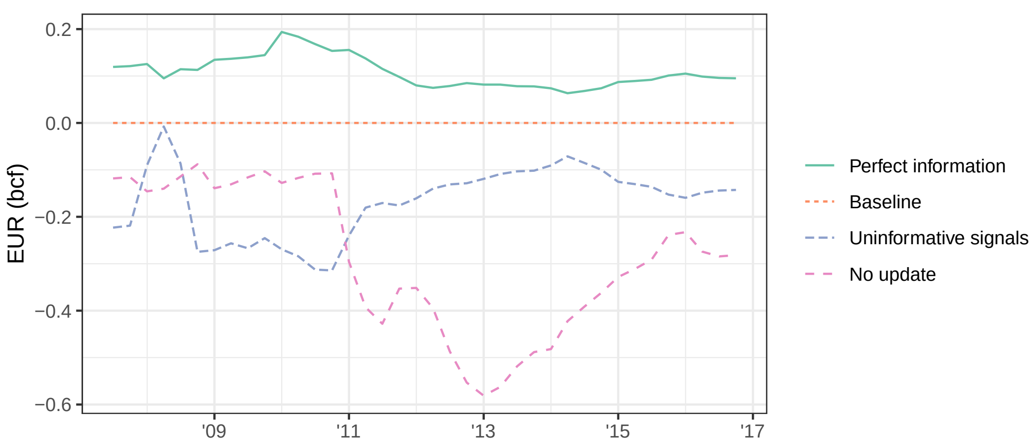
Both mineral lease expirations and firms' learning about geology imply that on average, initial wells produce less than subsequent development wells. This is purely a selection effect. The separate panes of Figure 9 illustrate that technological progress causes only a mild rise in output per well. However, a one-time shift from drilling initial to development wells implies a 1-5–2 bcf increase in overall mean EURs.

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.

Figure 10: Difference of counterfactual mean EUR from baseline simulations. The *Baseline* scenario corresponds to the *Overall* mean in Figure 9. Points above zero imply that counterfactual mean EURs are above *Baseline* estimates. Positive slopes imply that mean EURs are rising faster than *Baseline* estimates, and negative slopes, that they are falling faster.



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

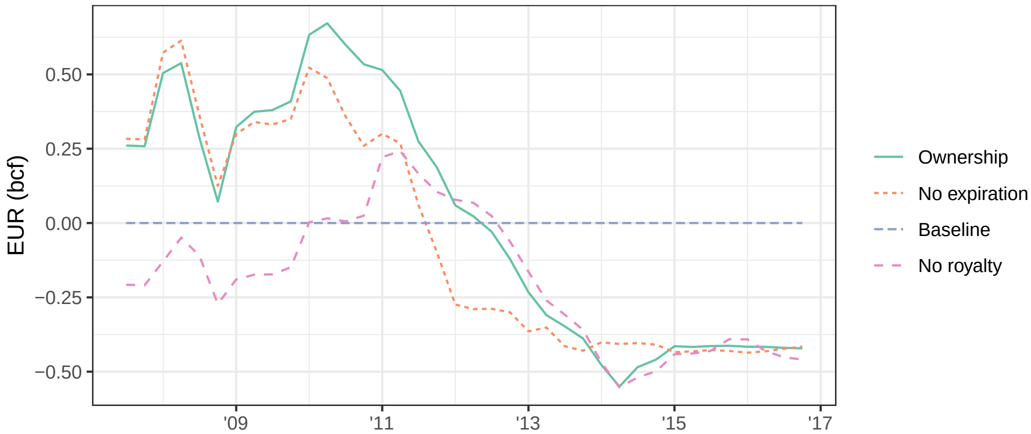
To further understand the way learning about geology impacts overall mean EURs, I simulate three counterfactual informational environments. In the first, firms have perfect information, so the correlation of signal and actual productivity is perfect: $\rho(\psi^0, \psi^1) = 1$. In the second, firms have totally uninformative signals 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 $\psi_{it} = \psi_i^0 \; \forall t$.

I plot the deviation of the three counterfactual mean EUR paths from the baseline mean EUR path in Figure 10. The figure shows that changes to firms' information about geology also change the path of mean output per well. Providing firms perfect information raises mean EURs in every period compared to the baseline world. When firms' initial signals are noisy, they end up drilling bad locations in search of good ones, and they fail to drill some profitable locations. When firms get uninformative signals, they learn more about geology from drilling an initial well. This causes mean output per well to increase slightly faster over 2009–2014 compared to the baseline scenario. Finally, when firms can make no update to their initial signals, mean EURs rise more slowly starting in 2010 than in the baseline scenario, and they end up 0.3 bcf lower. 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 EURs over the 2008–2016 period.

Parameter estimates imply that the distortions induced by mineral lease contracts matter far more to average output per well than does learning about geology. I compare mean EURs under three counterfactual lease contract structures with baseline mean EURs that use actual mineral lease contracts. Figure 11 shows the deviation of counterfactual mean EURs from the baseline scenario. In the first counterfactual, firms have full ownership of the minerals. No royalty rates or lease expirations distort their incentives. In the second counterfactual, firms pay royalty rates but leases do not expire.[35] In both of

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

Figure 11: Mean EUR under alternative mineral lease contracts



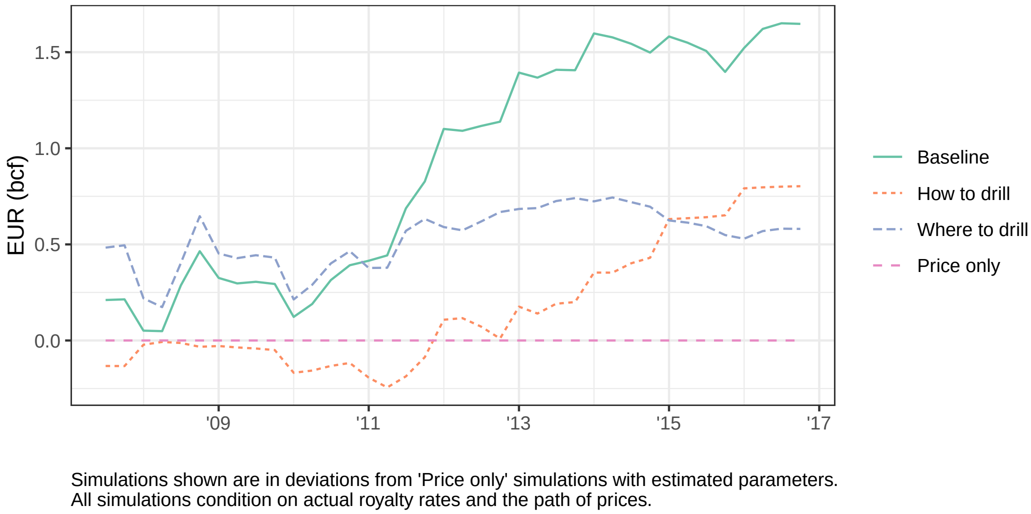
Simulations shown are in deviations from baseline simulations with estimated parameters.
All simulations condition on actual royalty rates and the path of prices.

these scenarios, mean EURs rise more slowly than in the baseline scenario: they start around 0.4 bcf higher compared to baseline, and they end 0.4 bcf lower. In the third counterfactual, we eliminate royalty rates, and the level of mean EURs generally decreases as firms are able to drill lower-quality locations.

Learning and mineral lease expirations both induced firms to systematically change *where* they drilled over time. Technological progress implied firms improved *how* they drilled. To summarize the relative importance of where and how firms drilled, I compare four paths for mean EURs. For a reference point, I simulate the path of mean EURs under a price only scenario that eliminates learning about geology, mineral lease expirations (but not royalty rates), and technological progress.[36] The price-only scenario serves as the reference zero axis in Figure 12, and I measure the three mean EURs relative to it. The baseline scenario produces the maximum relative increase in mean EURs by including learning, lease expirations, and technology. Together, the changes in where and how firms drilled raised mean EURs by 1.5 bcf relative to the price only world. The third path simulates a where-to-drill world in

[36]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 technology vs non-technology factors on mean EURs



which learning about geology and lease expirations affect firms' choices, but technology is fixed at 2007 levels ($\alpha_t = 0$). In this scenario, mean EURs initially increase rapidly along with the baseline scenario. In 2011, the increase slows and mean EURs peak at 0.5 bcf above the reference price only scenario. Finally, I simulate a how-to-drill world that allows for technological progress ($\alpha_t = \hat{\alpha}_t > 0$) but eliminates learning and lease expirations. In this fourth simulation, mean EURs end up 0.75 bcf higher than the price-only world, a little more than the increase produced by non-technology factors. It should be noted that in level terms (shown in the Appendix in Figure 26) some of the relative increases in mean EURs in both the how and where to drill scenarios simply serve to forestall the effects of depletion and prevent declining mean EURs.

6.3 Profit and rent implications

In addition to affecting the path of mean EURs, learning about geology, mineral lease contracts, and technological progress also affected firms' profits and realized resource rents. I compute profits and rents from the time leases are signed through the last quarter of 2016. I maintain firms' discount rate and

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.

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, $\mathbb{E}[\epsilon]$.[37] 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, social surplus, 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

[37] 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 (9).

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 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 have have dramatically larger 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 by 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 logged cumulative production: $\log Q_{iw\tau}$. The appropriate selection correction is the conditional expectation of ψ_i^1 given royalty rates and the history of drilling: $\mathbb{E}\left[\psi_i^1 \middle| \{d_{it}, z_t\}_{t=0}^T, \{s_{ij0}\}_{j=1}^{J_i}, x_{ir}, r_i\right]$. I return to the initial regression model, equation (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. It estimates that output per well grows at 7% per year. In the second column I include the selection correction term in and leave coefficients for OGIP and $\mathbb{E}[\psi_i^1]$ 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.

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 also allowed firms dramatically reduce drilling and completion costs. The focus on studying innovation in the shale extraction process plays into a broader narrative. Innovation offsets the physical limits of natural resources: technology vanquishes Malthus.

As I show in this paper, 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 and mineral lease contracts, as well as learning about the resource distribution. While mineral lease contracts distort firms' incentives and reduce resource rents, Herrnstadt, Kellogg, and Lewis (2018) show that the form of most private mineral leases is relatively close to optimal. 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 implicilty assume that we can, our forecasts may be overly optimistic. It is possible that Malthus might bite back.

References

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.

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.

Covert, Thomas R. (2015). *Experiential and Social Learning in Firms: The Case of Hydraulic Fracturing in the Bakken Shale.* SSRN Scholarly Paper ID 2481321. Rochester, NY: Social Science Research Network.

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.

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.

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

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.

Male, Frank, Akand W. Islam, Tad W. Patzek, Svetlana Ikonnikova, John Browning, and Michael P. Marder (2015). "Analysis of Gas Production

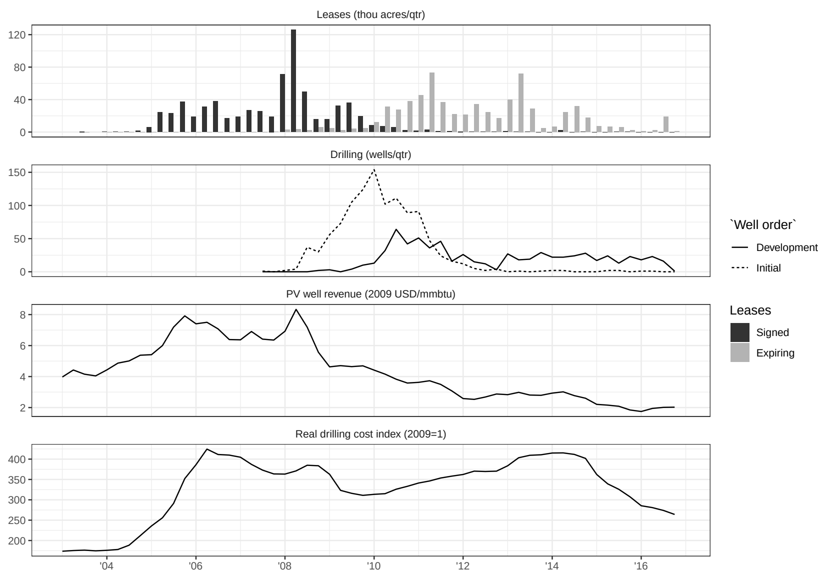
from Hydraulically Fractured Wells in the Haynesville Shale Using Scaling Methods". *Journal of Unconventional Oil and Gas Resources* 10, 11–17.
Meisner, J. and F. Demirmen (1981). "The Creaming Method: A Bayesian Procedure to Forecast Future Oil and Gas Discoveries in Mature Exploration Provinces". *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.
Quyen, N. V. (1991). "Exhaustible Resources: A Theory of Exploration". *The Review of Economic Studies* 58.4, 777–789.
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.
Seitlheko, Likeleli (2016). "Characterizing Production in the Barnett Shale Resource: Essays on Efficiency, Operator Effects and Well Decline". Thesis.
Smith, James L. (1980). "A Probabilistic Model of Oil Discovery". *The Review of Economics and Statistics* 62.4, 587–594.
— (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".
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.
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.

Vissing, Ashley (2016). “One-to-Many Matching with Complementary Preferences: An Empirical Study of Natural Gas Lease Quality and Market Power”.

A Data construction

A.1 Merging data

Figure 13: Haynesville development over time



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 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 Drillinginfo to each well based on the well's API number. While the DNR does report production data , it does so at varying levels of aggregation: the lease, unit, or well. Drillinginfo 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 Drillinginfo 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 5 tabulates the reasons I drop certain sections, and Figure 14 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. 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

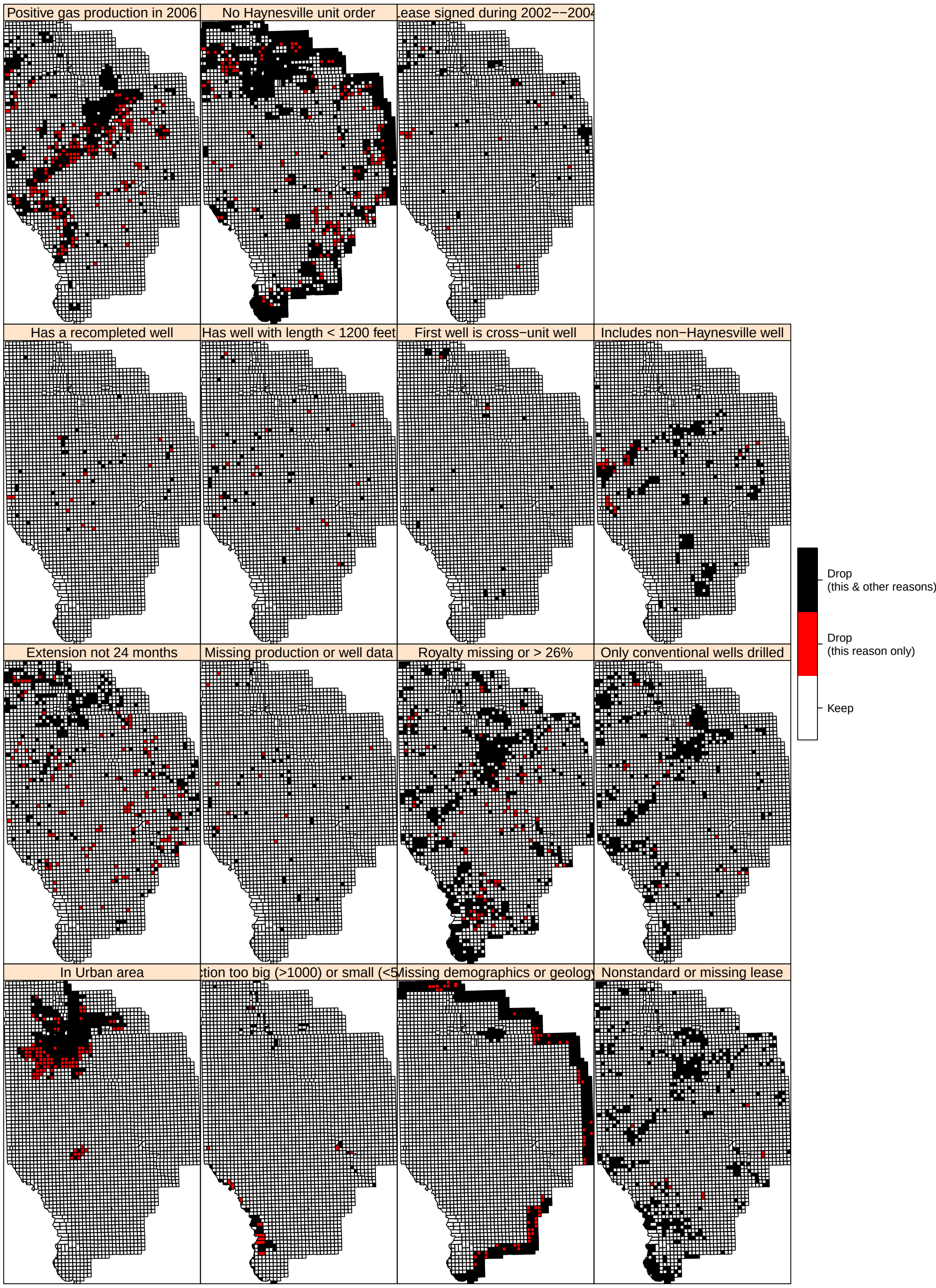


Figure 14: Sections dropped from final sample

first set of reasons I drop sections are that lease terms are nonstandard (or are missing). I drop 331 sections that have leases with 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 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.

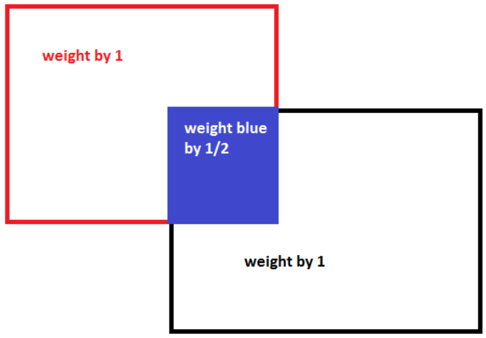


Figure 15: Lease weighting method

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 in 40 acre blocks. 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 15 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 13 plot the history of investment from 2003 to 2016. The top pane shows quarterly mineral leasing when leases expire.[38] The second pane breaks out the number of wells drilled per month by

[38] 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.

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.[39] 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

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

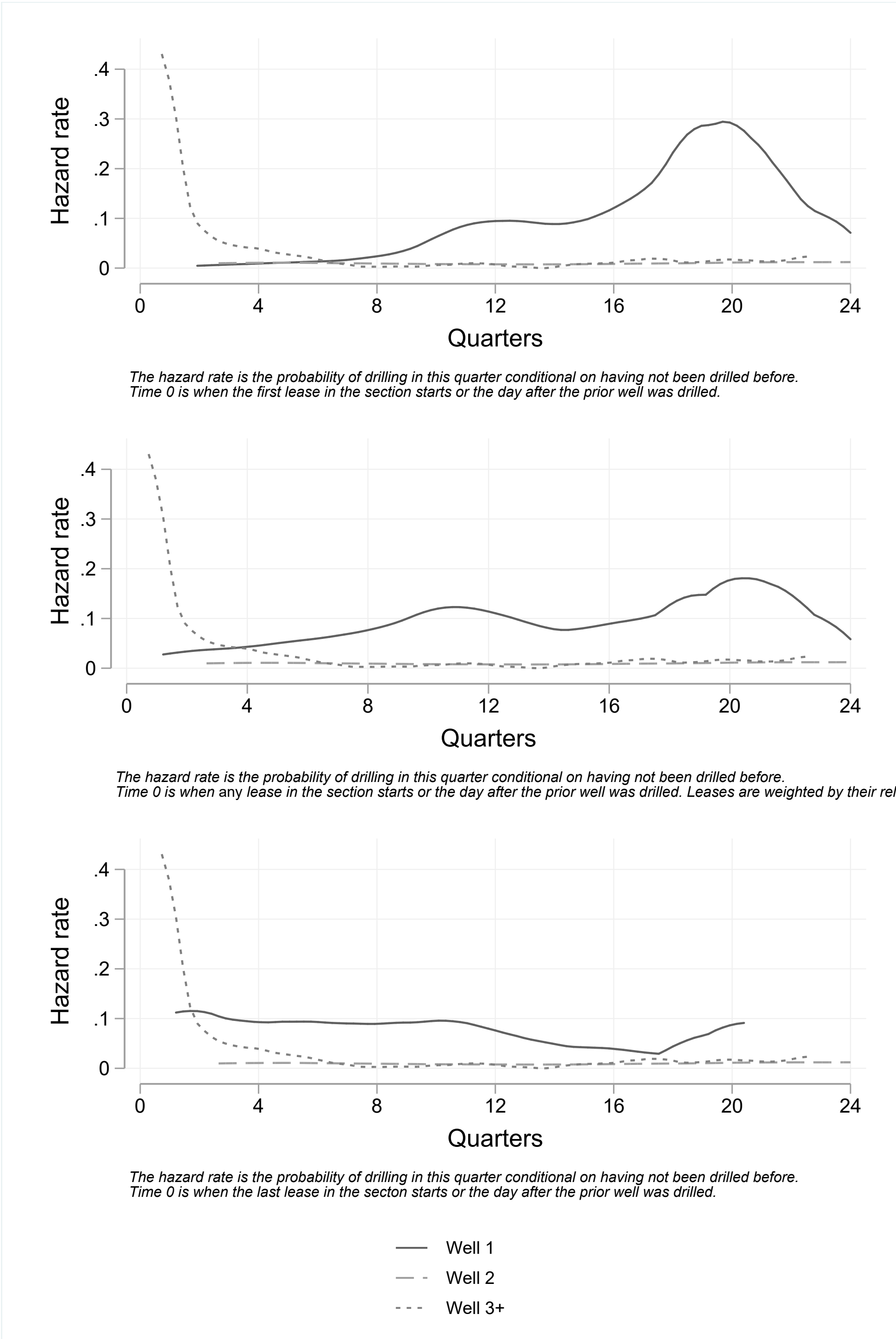
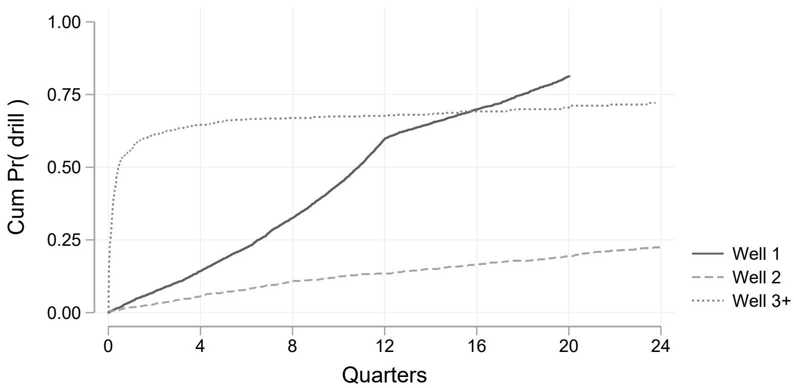


Figure 16: Drilling hazard rates when just the first lease is used, all leases are used, and just the last lease is used

Figure 17: Cumulative weekly failure rate by well-order for 36-month leases



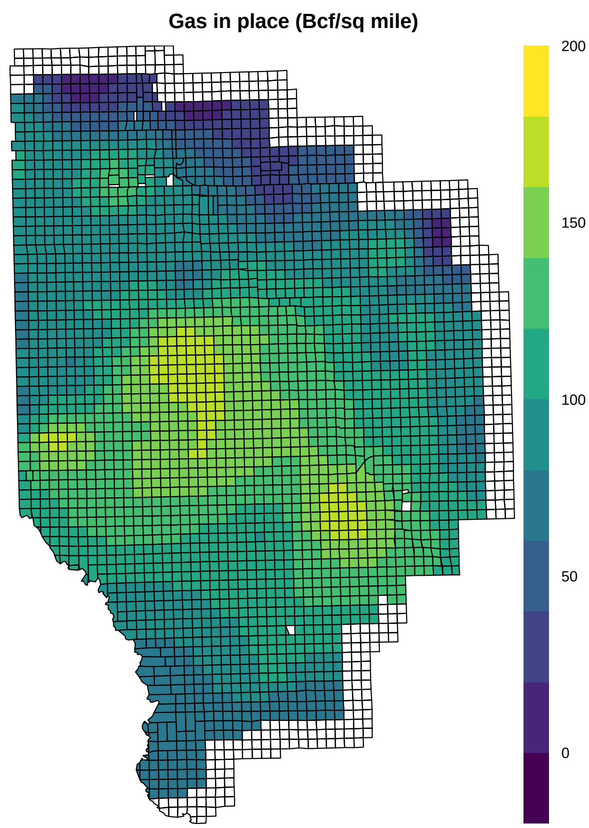


Figure 18: Original gas in place (Gülen et al. 2015)

Table 5: 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.

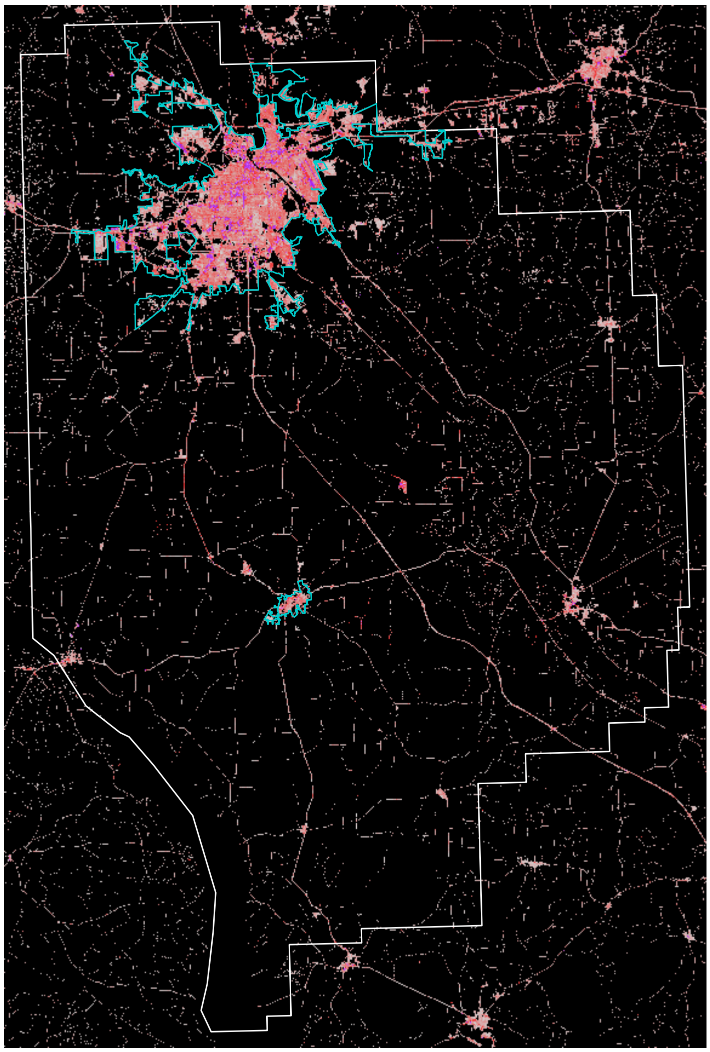


Figure 19: Imperviousness (pink) and urban areas (blue outline)

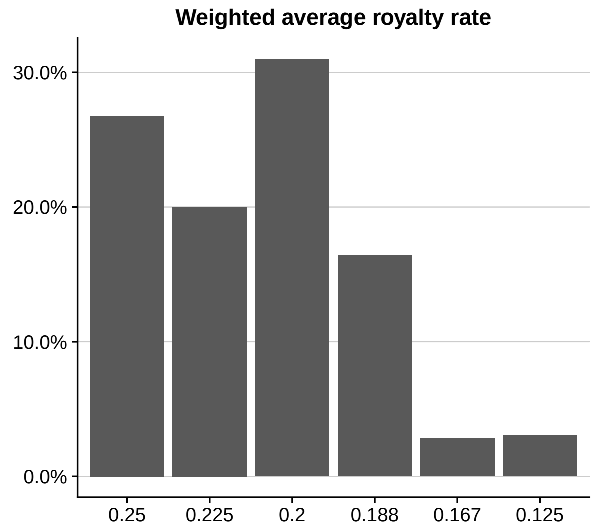


Figure 20: Distribution of discretized, averaged royalty rates r_i (unit-level)

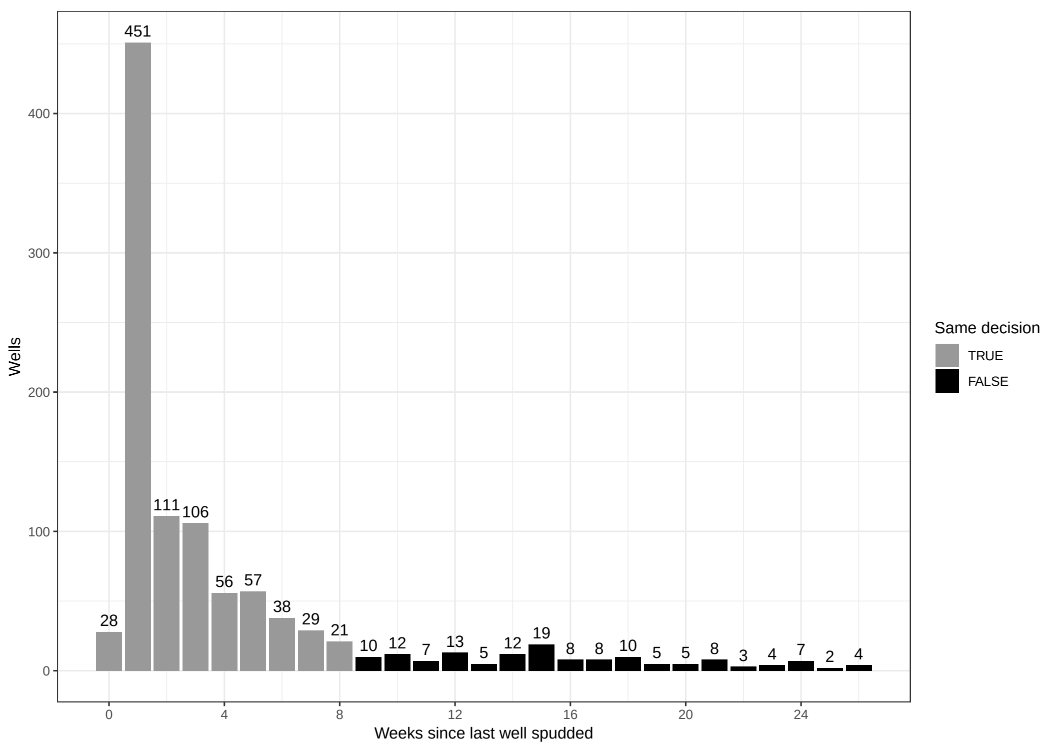
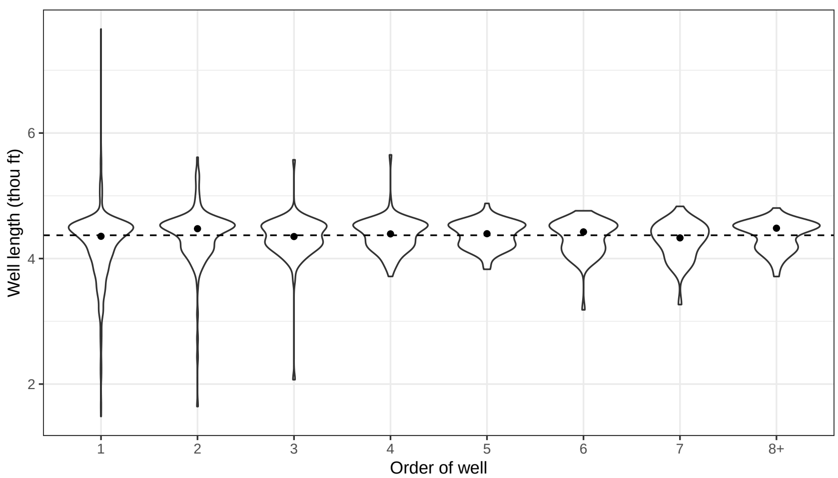


Figure 21: Weeks since previous well drilled

Figure 22: Distribution of well-length



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

Table 6: 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 7: 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 8: Summary: Wells per operator

Original operator	All
Aethon	24
BHP	201
Chesapeake	567
Comstock	71
Covey Park	47
Encana	3
Enduro	5
EOG	1
Exco	278
Fortune	1
Franks	2
GEP	224
Goodrich	4
Indigo	38
Matador	1
QEP	66
Sabine	12
Samson	22
SND	0
Swepi	7
Trinity	0
Vine	205
XTO	20
Other	0
All	1799

	N	Mean	SD	Min	Q1	Median	Q3	Max
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

Table 9: Summary: Periods

	N	Mean	SD	Min	Q1	Median	Q3	Max
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 10: 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 11: 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 monthly production decline, cumulative production, and well-specific estimates for EUR. I start by estimating a common production decline curve using all months of well production data as

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

The regression 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 use this decline curve to compute the present value of well revenues in the next section. Because I am most interested in EUR, which is related to cumulative production, I also estimate

$$\log Q_{iw\tau} = \gamma_{\min\{\tau,72\}} + u_{iw} + \eta_{iw\tau}.$$

This gives me well-specific fixed effects $\hat{u}_{iw}$ and a profile for cumulative production based on $\hat{\gamma}_{\min\{\tau,72\}}$ for months 4–72. I then extend this decline curve out to month 240 using the monthly production decline curve. EUR for well w in section i is then $\mathbb{E}\left[Q_{iw,240} \middle| \{Q_{iw,\tau}\}_{\tau=1}^{T_{iw}}\right] = \exp\{f(240; \hat{\gamma}^q_\tau, \hat{\gamma}^q_{\min\{\tau,72\}}, \hat{\gamma}_{\min\{\tau,72\}}) + \hat{u}_{iw} + \hat{\sigma}^2_\eta/2\}$ where $f()$ is the estimated profile of cumulative production.

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.[40] 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 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

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

and discounted average of futures prices less costs for gathering, treatment, and compression \$0.49[41] respectively:

$$p_t = \sum_{s=5}^{245} \left\{ \frac{\exp\{f_q(s-5)\}}{\sum_{k=1}^{240} \exp\{f_q(k)\}} \tilde{\beta}^{s/12} \left[F(t, t+s) - 0.49 \right] \right\} \tag{17}$$

where $\tilde{\beta}$ is the nominal discount factor, $f_q(k; \hat{\boldsymbol{\gamma}}^{\boldsymbol{q}}, \hat{\delta})$ is expected production decline curve with parameter estimates taken from a regression of log monthly production on a vector of well-specific and production-month fixed effects.

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 $\tilde{\beta} = 1/1.125 \approx 0.89$ 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 an expectation for the z_{1it}. I fail to reject unit roots in the logged weighted average price of natural gas, $\log p_t$ computed using (17), and logged drilling dayrate, $\log c_t$. I therefore assume they follow random walks[42] 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 $\hat{\rho}_{pc} = 0.3099$. For the baseline model I discretize prices on an evenly spaced grid of 51 points that goes from one-third

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

[42] 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.

of the lowest price in my dataset to thrice the highest price. When I include dayrates, $\log c_t$, the state space increases exponentially. Memory limits require me to scale back my grid to 17 points in each dimension, and I allow the grid to extend $\pm log(2.5)$ beyond the minimum and maximum prices observed.

When computing the value function, I must integrate over both next period's prices as well as next period's information. To do this, I discretize the variables, compute a transition matrix, and use the transition matrix to integrate over next period's values. Then I interpolate over continuous variables using quadratic B-splines. I choose quadratic B-splines over a spectral method like Cheybshev polynomials because the value function will have kinks.

I choose the probabilities in my 625 by 625 element transition matrix by integrating the CDF of the bivariate normal distribution as Kellogg (2014) does. I set the limits of integration to be the midpoints of the corresponding grid points. Because many of the probabilities are small, I zero out ones smaller than 10^{-5} and use sparse matrices. I discretize ψ_t over grids of 51 points if only prices are used, or 19 if prices and dayrates are. I use the Tauchen (1986) procedure to compute a dense transition matrix for ψ because it has a straightforward Frechet-derivative I can calculate easily. The Tauchen (1986) procedure sets the elements of the transition matrix Π_ψ to be

$$\pi_{ij} = \begin{cases} \Phi\left(x_+\right) & \text{if } j = 1 \\ 1 - \Phi\left(x_-\right) & \text{if } j = 2M_\psi/\Delta_\psi + 1 \\ \Phi\left(x_+\right) - \Phi\left(x_-\right) & \text{otherwise} \end{cases}$$

where

$$x_+ = \frac{\psi_j - \rho^2\psi_i + 0.5\Delta_\psi}{\sqrt{1-\rho}} \qquad\qquad x_- = \frac{\psi_j - \rho^2\psi_i - 0.5\Delta_\psi}{\sqrt{1-\rho}}$$

As with prices, when evaluating the integrated value function, $\mathbb{E}\mathrm{V}$, I interpolate between grid-points using quadratic B-splines. This has the added advantage of providing $\partial\,\mathbb{E}\mathrm{V}\,/\partial\psi$ for minimal additional computational cost.

In the inner nested fixed point (NFXP) loop, I solve the integrated value function by backwards induction one leasing-drilling state at a time. The leasing-drilling state s_{it} is a tuple $s_{it} = (\tau_{0it}, \tau_{1it}, d_{-1,it}, D_{it})$ where τ captures time-to expiration; d_{-1}, whether a well was drilled in the prior period; and D, the cumulative drilling to date. These are sorted lexicographically by $-\tau_1$, $-\tau_0$, $-d_{-1}$, 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: $\mathbb{E}V(s_{|S|}, z, \psi) = 0 \quad \forall z, \psi$. Computing $\mathbb{E}V$ at all s involves computing $\mathbb{E}V$ at $s_{|S|-1}$, then computing $\mathbb{E}V$ at $s_{|S|-2}$ using $\mathbb{E}V$ at $s_{|S|-1}$, and so on.

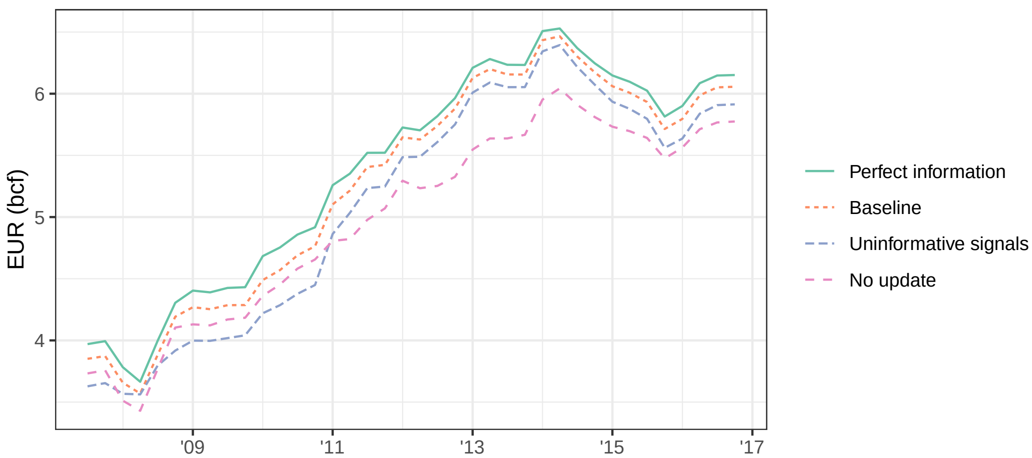
At all leasing-drilling states s_i with $i < |S|$, the firm's problem is finite horizon if the firm cannot remain at s_i by not drilling. Conversely, it is an infinite-horizon problem if the firm can. I solve finite-horizon problems by value function iteration, and infinite horizon problems by a hybrid 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. Because prices, volatility, unobserved information (ψ) is continuous state variables, I use quadratic B-splines to interpolate over the value function in these dimensions. I use Monte Carlo integration with two Halton (1960) sequences of bases two and three to integrate out the independent standard normal 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.

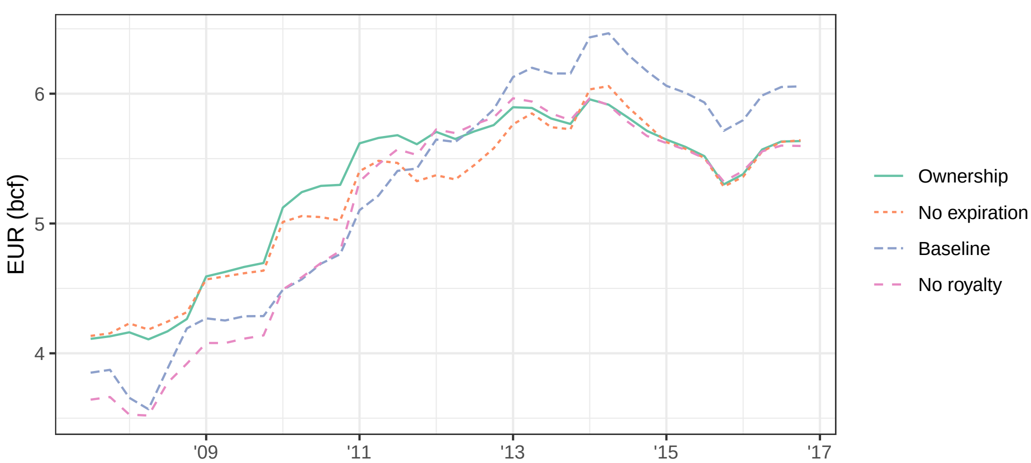
D Simulations: additional figures

Figure 23: Difference of counterfactual mean EUR from baseline simulations



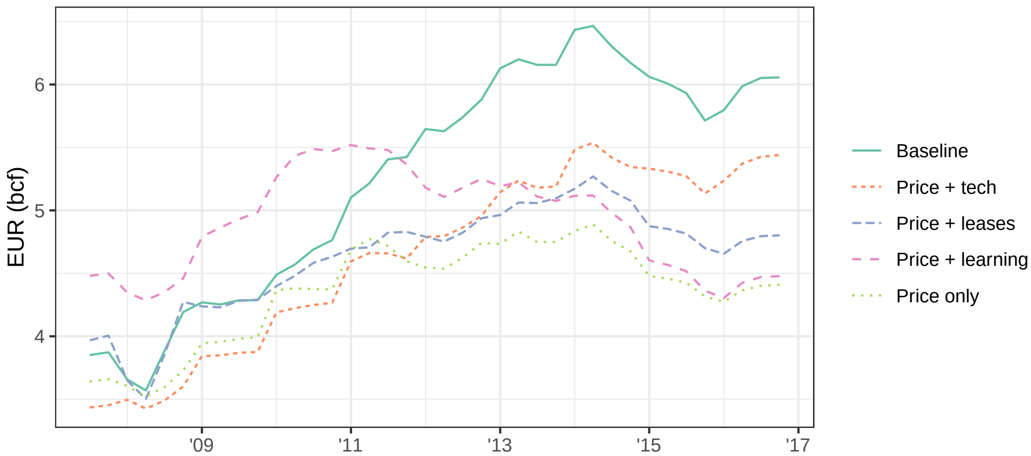
Simulations are based on estimated parameters. They condition on actual royalty rates and the path of prices.

Figure 24: Mean EUR under alternative information and contract structures



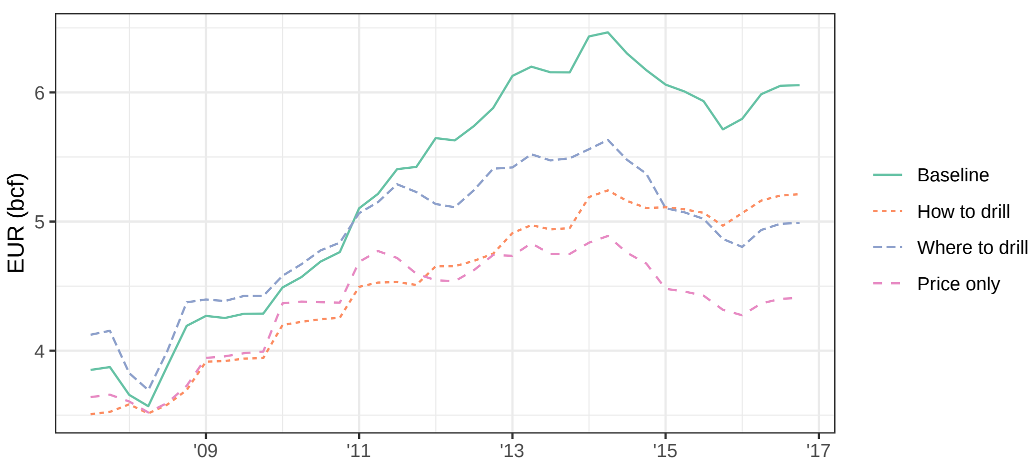
Simulations are based on estimated parameters. They condition on actual royalty rates and the path of prices.

Figure 25: Mean EUR with fewer drivers of increasing output



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

Figure 26: Mean EUR: technological vs non-technological drivers of output per well



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

Appendix References

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.

Covert, Thomas R. (2015). *Experiential and Social Learning in Firms: The Case of Hydraulic Fracturing in the Bakken Shale*. SSRN Scholarly Paper ID 2481321. Rochester, NY: Social Science Research Network.

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.

Halton, J. H. (1960). "On the Efficiency of Certain Quasi-Random Sequences of Points in Evaluating Multi-Dimensional Integrals". *Numerische Mathematik* 2.1, 84–90.

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

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

Patzek, Tad W., Frank Male, and Michael Marder (2013). "Gas Production in the Barnett Shale Obeys a Simple Scaling Theory". *Proceedings of the National Academy of Sciences* 110.49, 19731–19736.

Rust, John (1994). "Structural Estimation of Markov Decision Processes". In: *Handbook of Econometrics*. Ed. by Robert F. Engle and Daniel L. McFadden. Vol. 4. Elsevier, 3081–3143.

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