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

Mark Agerton

March 27, 2020

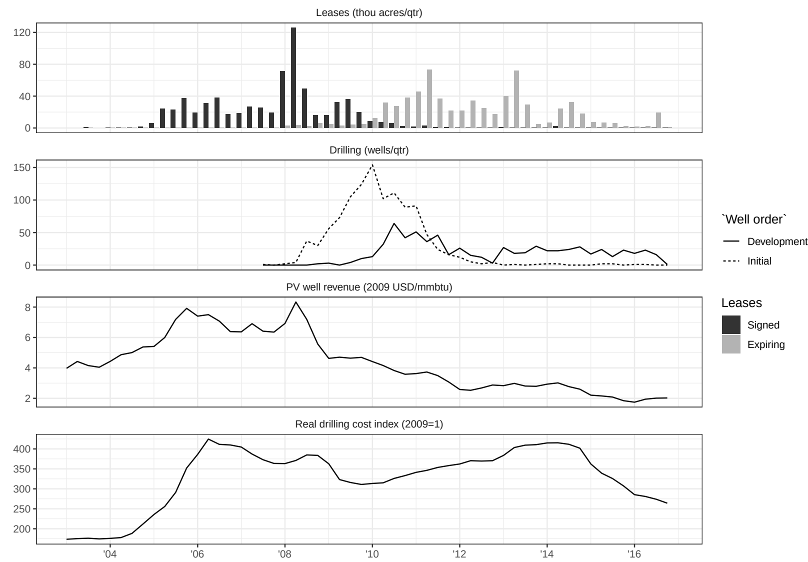
A Data construction

A.1 Merging data

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

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

Figure 1: Haynesville development over time



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

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

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

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

A.2 Sample Selection

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

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

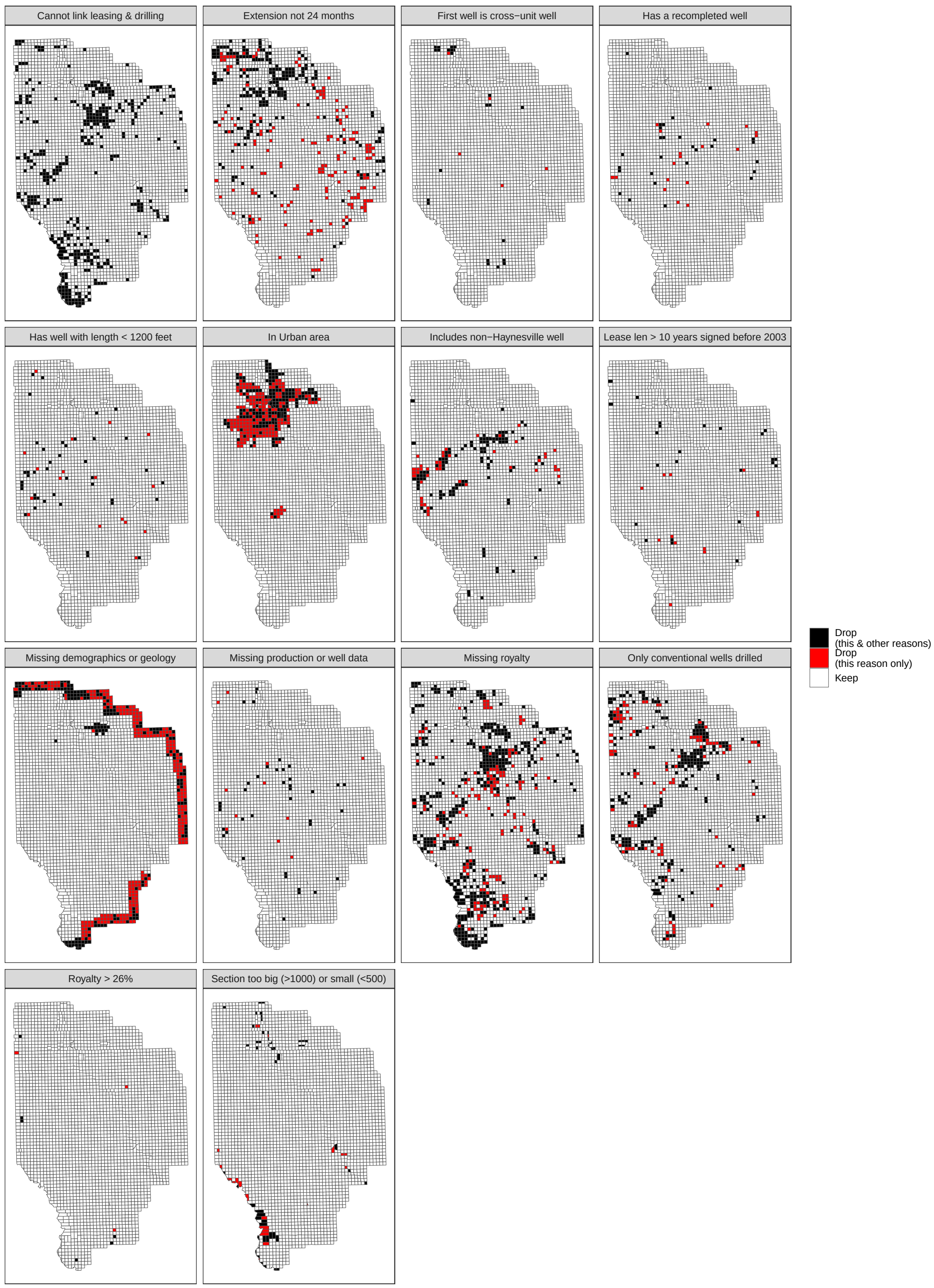


Figure 2: Sections dropped from final sample

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

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

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

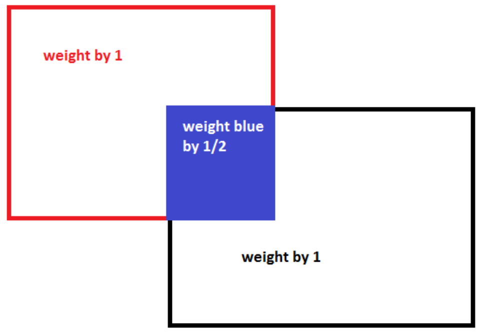


Figure 3: Lease weighting method

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

A.3 Overlapping leases

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

B Descriptive statistics

B.1 History of shale activity

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

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

B.2 Descriptive figures

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

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

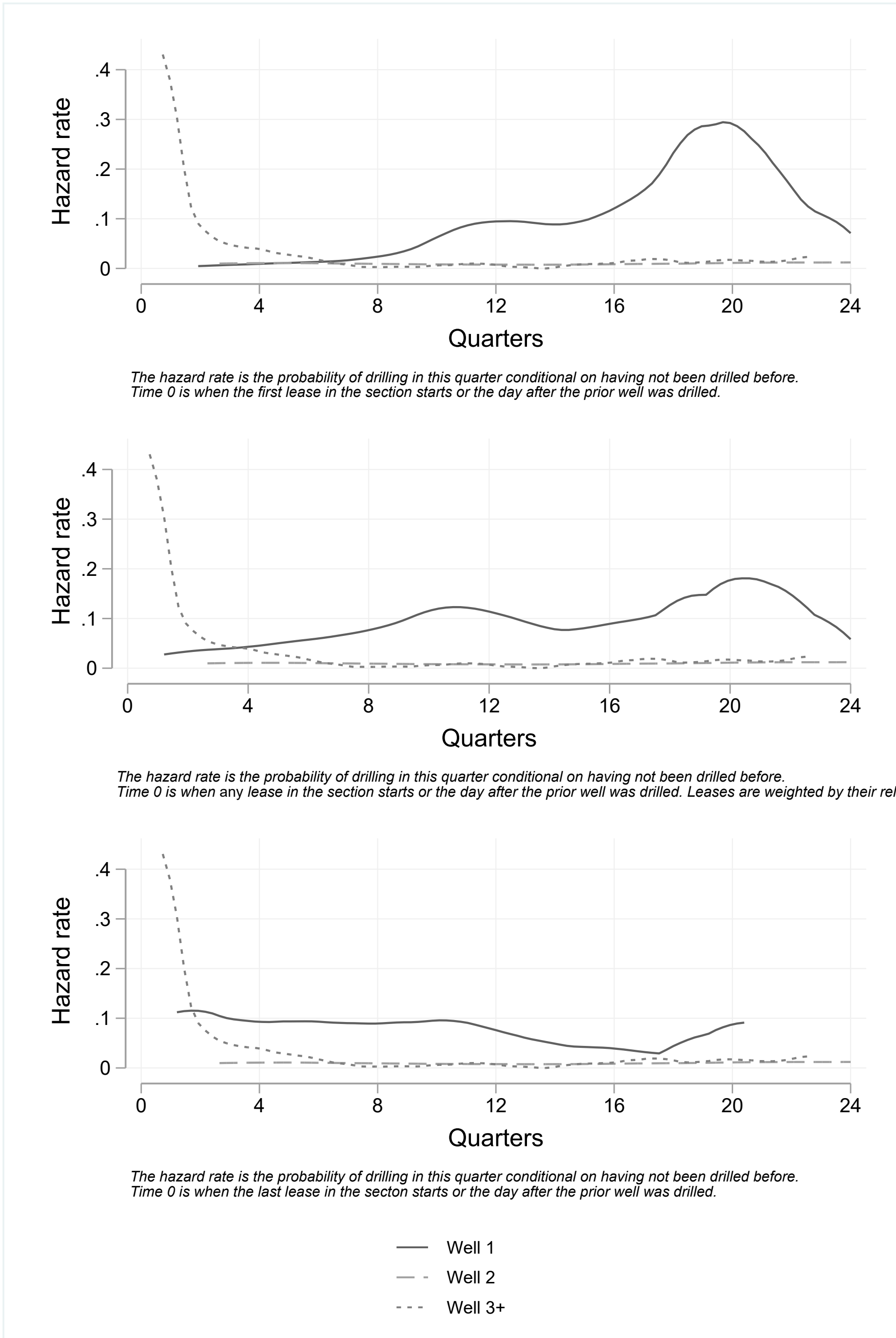
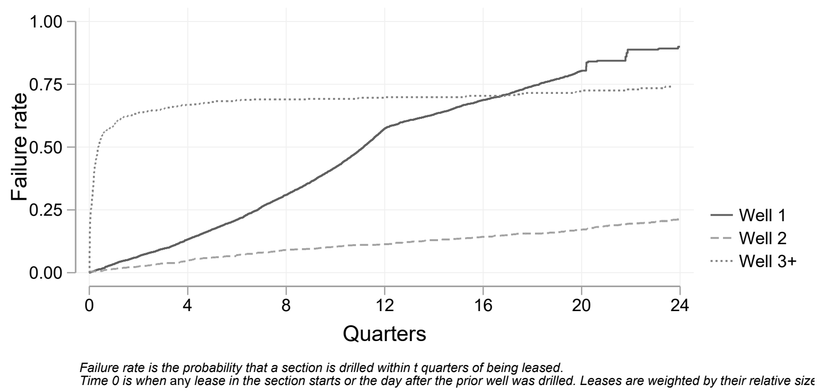


Figure 4: Drilling hazard rates when just the first lease is used, all leases are used, and just the last lease is used

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



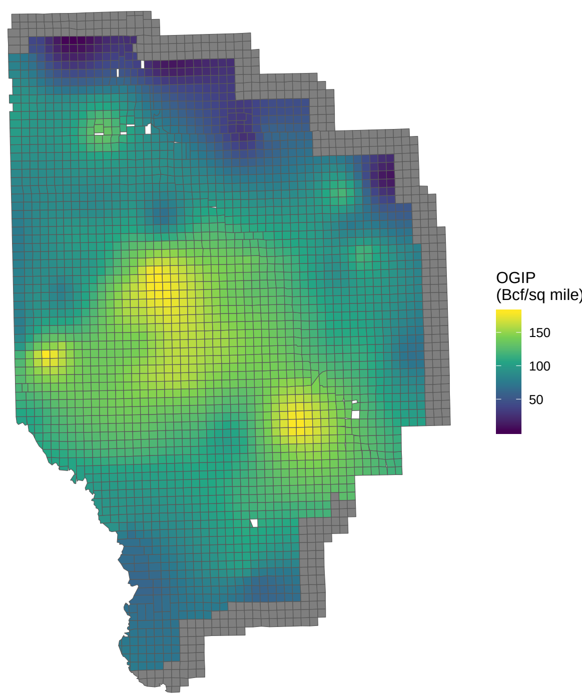


Figure 6: Original gas in place (Gülen et al. 2015)

Table 1: Reasons sections are dropped

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

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

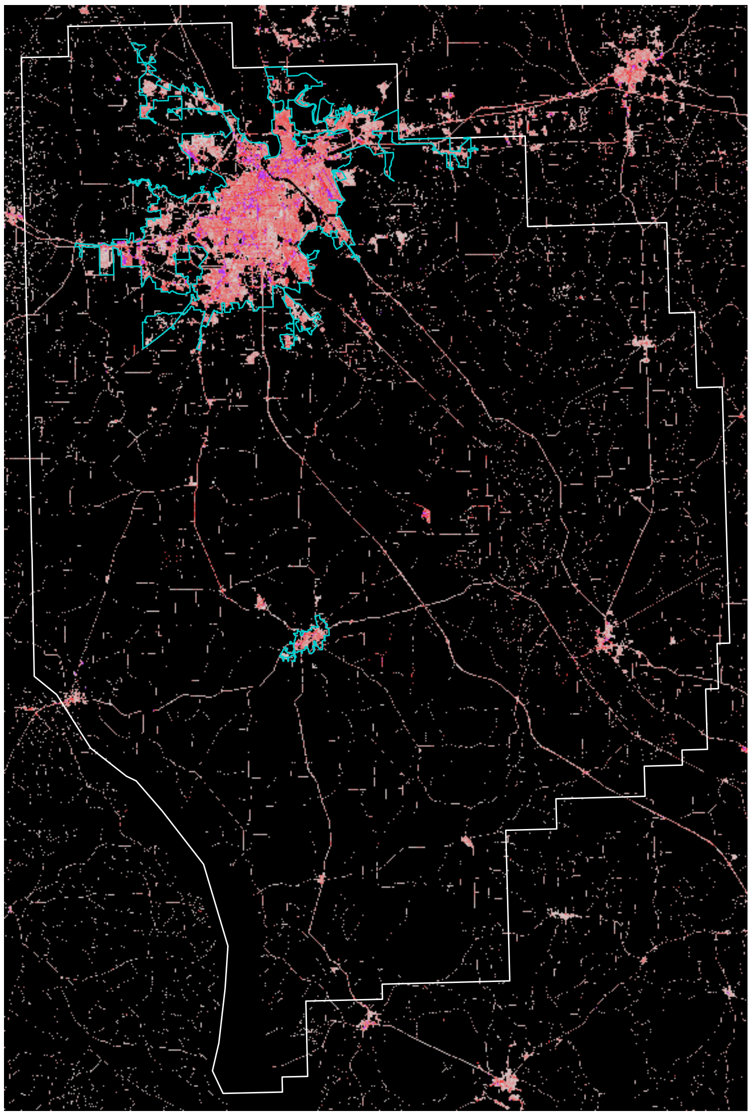


Figure 7: Imperviousness (pink) and urban areas (blue outline)

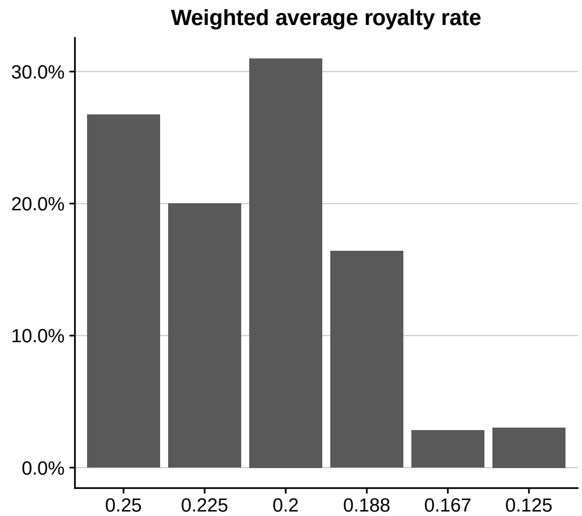


Figure 8: Distribution of discretized, averaged royalty rates r_i (unit-level)

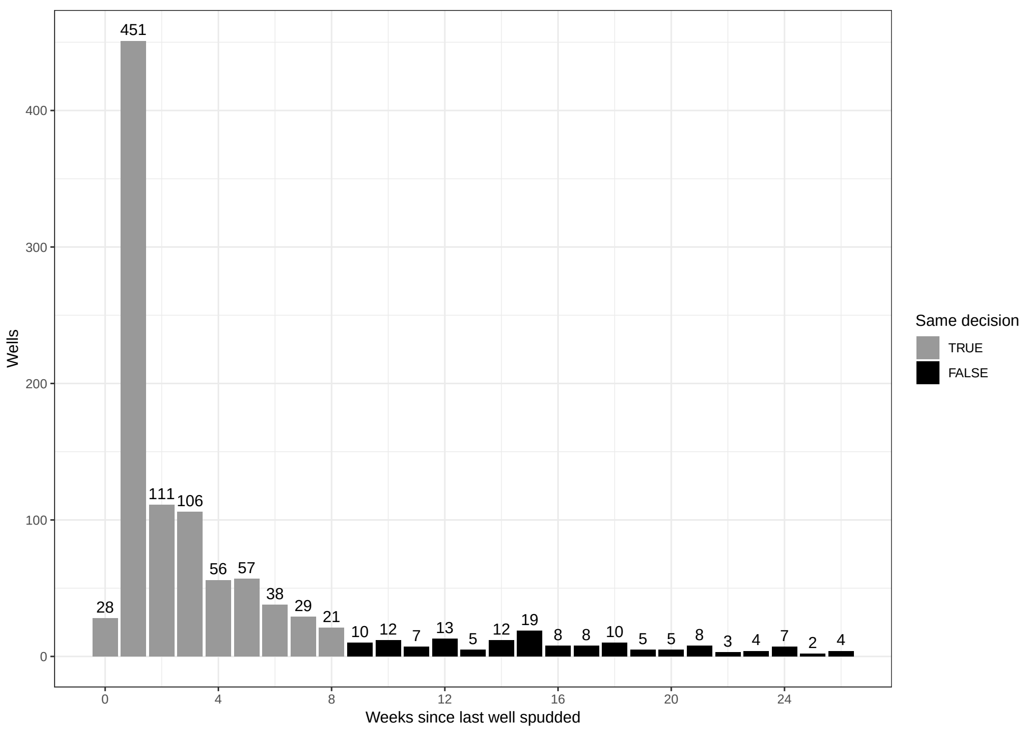
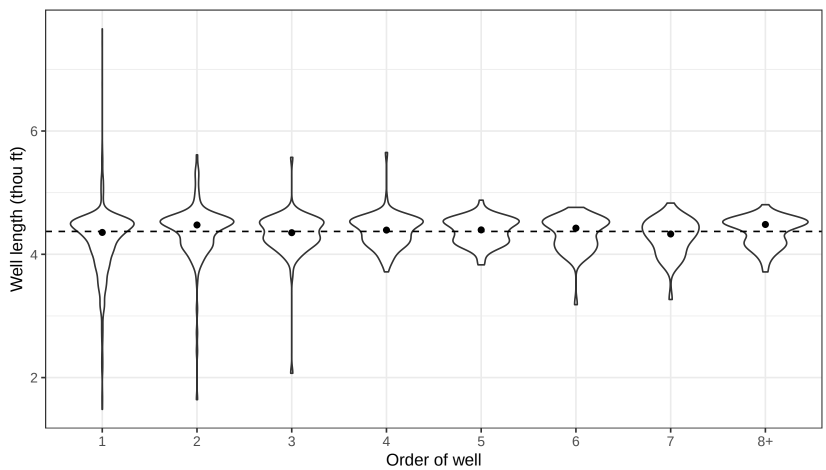


Figure 9: Weeks since previous well drilled

Figure 10: Distribution of well-length



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

Table 2: Summary: Sections

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

Table 3: Summary: Wells

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

Table 4: Summary: Periods

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

Table 5: Summary: Leases

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

Table 6: Summary: Total drilling by geology and royalty

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

C Computation

C.1 Production-based EUR calculations

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

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

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

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

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

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

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

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

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

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

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

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

and EUR for well w in section i is simply

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

C.2 Constructing prices

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

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

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

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

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

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

C.3 Transitions for prices

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

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

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

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

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

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

C.4 Nested fixed point routine

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

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

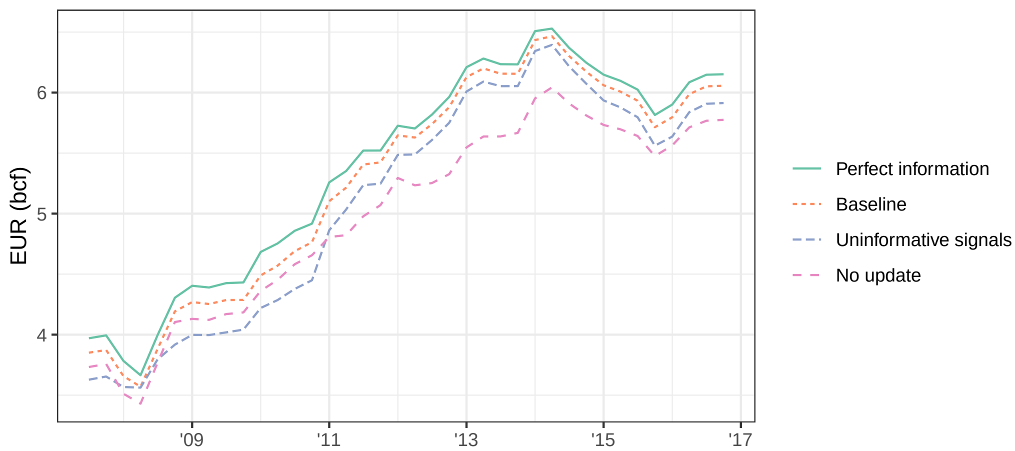
[6] This is the same as in Kellogg (2014).

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

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

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

Figure 11: Counterfactual mean EUR under alternate informational environments



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

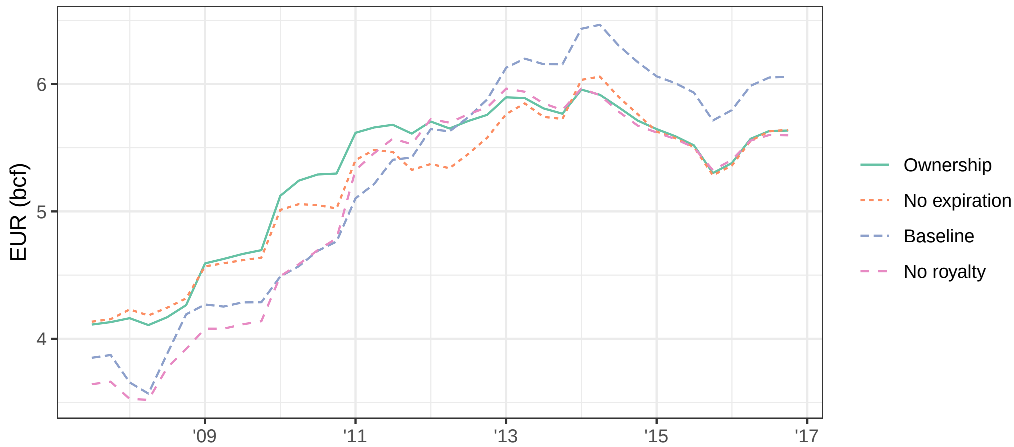
variables u and v. After discarding the first 5000 observations, for each unit i, I draw 2000 pairs of shocks. Results do not change meaningfully if I increase (or decrease) the number of simulated draws.

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

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

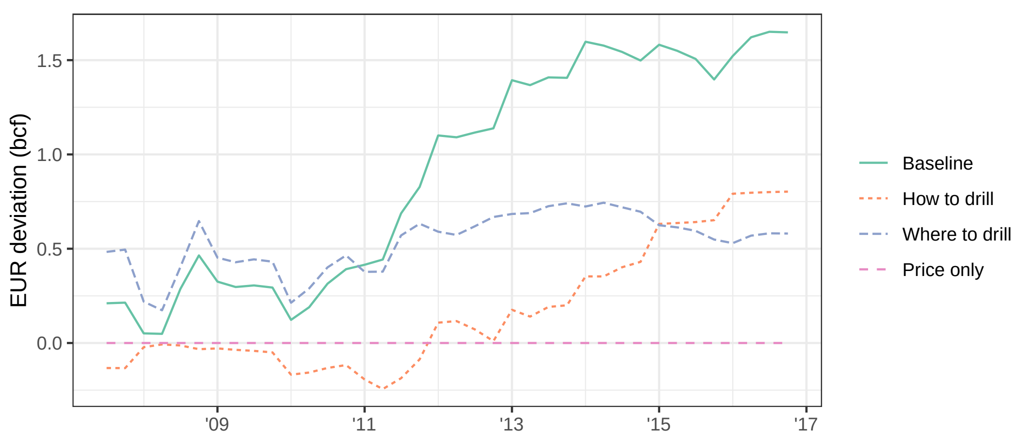
D Simulations: additional figures

Figure 12: Counterfactual mean EUR under alternate mineral lease contracts



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

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



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