ABSTRACT

Title of Dissertation:	THREE ESSAYS ON ENVIRONMENTAL ECONOMICS Lauren Beatty Doctor of Philosophy, 2023
Dissertation Directed by:	Professor Joshua Linn Department of Agricultural and Resource Economics

Oil and gas production is associated with numerous types of environmental damages and hazards. This dissertation is a collection of three essays which empirically examine how oil and gas production affects environmental outcomes, with a particular lens on avenues for policy to remediate these damages.

In chapter 1, I focus on the joint relationship between drilling, building pipelines, and emissions. Most oil wells co-produce natural gas. Producers can choose to burn this valuable co-product on site (known as flaring) if the cost of connecting to the existing natural gas pipeline network is sufficiently high. I construct and estimate a dynamic model of producer drilling and flaring decisions which depend on the current state of the pipeline network and expectations over its evolution. My model also allows producers to internalize spillover effects for their neighbors – any pipeline they build will extend the network and weakly decrease their neighbors' future pipeline connection costs. Using my model estimates, I simulate pipeline development and flaring outcomes under counterfactual policies. My counterfactual simulations show that flaring

abatement costs are higher than previous studies but suggest that a flaring tax could substantially reduce flaring. A $\$5/$Mcf tax reduces flaring by 39%.

In chapter 2, I focus on the non-point source pollution nature of methane emissions from oil and gas production. New scientific literature demonstrates that methane emissions from oil and gas production are a far larger problem than previously estimated. However, very little economics work has been conducted on this problem. Methane emissions are caused by leaked natural gas which escapes during the normal operation of equipment as well as leaks from faulty equipment. This implies that there are two avenues to abate emissions – operators can install more efficient pumps and pneumatic devices, or they can expend more effort finding and fixing leaks from faulty equipment. In this chapter, I seek to understand how operators respond to prices on each margin using output from an inverse atmospheric modelling tool which outputs a gridded inventory of emissions. If producers primarily abate emissions through capital upgrades, then an input-based scheme where the regulator observes capital choices, then asses a tax on imputed emissions will be fairly efficient. I find that both channels of abatement are important, and that a tax on imputed emissions would achieve significantly less emissions reductions than an equivalent Pigouvian tax.

Finally in chapter 3, I focus on policy options to deal with governmental liabilities from abandoned oil and gas wells. There are hundreds of thousands of aging oil and gas wells scattered throughout the United States that pose serious environmental and safety risks. These well sites will require billions of dollars of investment in remediation. When producers go bankrupt before remediation is complete, the responsibility to clean up the site often lands with either the state or federal government. These wells are known as orphan wells, and have received increasing attention in the scientific and policy literature. In this chapter, I estimate a model of well-level status transitions, then use my model to simulate how a policy requiring producers to either bring wells back into production or plug them after two years of inactivity would affect well orphan rates. I find that since many wells are left inactive for years at a time, this simple policy would be an effective way to decrease government plugging responsibilities and prevent environmental damage without dramatically reducing oil and gas production.

THREE ESSAYS ON ENVIRONMENTAL ECONOMICS

by

Lauren Beatty

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2023

Advisory Committee:
Professor Joshua Linn, Chair/Advisor
Professor James Archsmith
Professor Louis Preonas
Professor Andrew Sweeting
Professor Roberton Williams III

Dedication

For my mother, who gave me everything she possibly could have.

Acknowledgments

I owe a great deal of gratitude to many individuals who made this thesis possible. I cannot possibly pay it back, so I aspire to pay it forward. First, I'd like to thank my advisor, Joshua Linn. In addition to giving me thoughtful feedback and providing me with important insights, Josh was always incredibly generous with his time. I also owe a great deal to my other advisors, Prof. James Archsmith, Prof. Louis Preonas, and Prof. Rob Williams III. There were numerous times during my graduate studies when I felt hopelessly lost and my advisors were always there to listen, brainstorm, and encourage me. In classes, in meetings, and as supervisors they've been excellent role-models of scholarship. I'd also like to sincerely thank Prof. Andrew Sweeting for serving on my dissertation committee and providing an incredible introduction to empirical industrial organization, as well as Prof. Sarah West, my undergraduate advisor, for helping me discover economics and giving me the confidence to pursue graduate work.

Funding that made this dissertation possible was generously provided by the University of Maryland, College Park. In addition, the first two chapters were supported by funding from the National Bureau of Economic Research Pre-Doctoral Fellowship Program in Energy Economics, which allowed me to focus exclusively on research during a crucial time during my PhD.

My experience in the graduate program was greatly enhanced by the excellent faculty and staff. I'd like to thank Anna Alberini, Lars Olson, Robert Chambers, Erich Battistin, Jorge Holzer, and Maureen Cropper for preparing excellent coursework. I could not be the scholar I am today without the solid base in econometrics and applied microeconomics that I learned in their classes. Pin-Wuan Lin, Jeff Cunningham, John Jensenius, and Dany Burns kept the department running smoothly during my time at in the department. I'm grateful to them for helping me navigate the bureaucracy, providing me with amazing computing resources, and making sure I had money and health insurance.

I cannot possibly imagine what my six years at the University of Maryland would have been like without the companionship and support of my colleagues and friends Zed Higgs, Alejandro Lopez, Dan Kraynak, Youngho Kim, Alex Perrault and Sam Williamson. Thank you for sharing the highs and lows of being a graduate student with me. I look forward to many more years of friendship.

Finally, I'd like to thank my boyfriend, Alex Broholm and my mother, Marilynn Beatty. I am incredibly fortunate for their unyielding love and support.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

CCP	Conditional choice probability
CDF	Cumulative distribution function
DDC	Dynamic Discrete Choice
EIA	Energy Information Administration
EPA	Environmental Protection Agency
GHGRP	Greenhouse Gas Reporting Program
IMI	Integrated Methane Inversion
INGAA	Interstate Natural Gas Association of America
IRA	Inflation Reduction Act
Mcf	Thousand Cubic Feet
RRC	Railroad Commission

1.1 Introduction

Networks are frequently studied in economics, but typically in static contexts where space is either unimportant or abstracted away from such as communication networks or two-sided markets. When space is abstracted away from networks can be modeled using graph theoretic techniques or even more simple descriptors such as numbers of network participants. For example, Rochet and Tirole (2003), in their classic models of two-sided markets, simply specify buyer and seller gross surplus as a function the of number of participants on each side. Typically network problems are characterized by network externalities. Consider the early development of phone or internet networks where users are positively affected by the entrance of a new user.[1]

However, there are many examples of network problems where space is crucial and where dynamics are important such as electricity grids, electric vehicle charging networks, and pipeline networks. These structures are still characterized by traditional network externalities, but in addition dynamics and space are important. The effect of a new entrant into the network is dependent upon where that entrant is sited, and the future states of the market are dependent upon current and past states. For example, consider the siting and construction of new renewable energy sources. Many of the best places to site renewable energy projects are located in very rural areas with relatively little transmission infrastructure and high costs of connecting to the existing grid. When a new generation source, like a wind or solar farm, is proposed the developer is often on the hook for expensive upgrades to transmission infrastructure that later entrants can use without sharing in the costs (Department of Energy, 2022). Clearly, there are network externalities – one producer's decision to upgrade the grid will affect future potential entrants, but

[1] However, network externalities need not be positive or even monotonic.

in addition the decision to enter generates spatial and temporal dependencies. Moreover, in all of these examples network externalities interact with more traditional environmental externalities. The interaction of spatial and temporal dependencies have presented challenges in past work modelling these types of problems. This paper will focus on the growth of a natural gas pipeline network, its relationship with producer decisions to connect to it, and its implications for environmental outcomes.

In this paper, I set up and estimate a dynamic model of producer decisions to drill and connect to natural gas pipelines in the Permian Basin in Texas, accounting for the fact that the pipeline network grows over time and that producers might consider spillover effects for their neighbor – that when a driller connects, the pipeline network expands, weakly reducing costs for nearby potential wellsites. Oil wells often co-produce natural gas, and producers have the option to burn the gas at the wellhead (flare) if the cost of connecting to pipelines is too high. Flaring is socially costly, so this paper will be primarily focused on investigating the relationship between the pipeline network and producer drilling and flaring decisions. I answer three main questions. First, how do producers choose to drill and flare and how is that decision affected by the pipeline network? Second, how might potential policies aimed at reducing flaring affect flaring and production decisions, and the growth of the pipeline network over time? Lastly, do producers internalize spillovers for their neighbors, and if not, how much does this matter for flaring? I find that distance to pipelines is an important driver of firm drilling and flaring decisions. I also show that flaring abatement costs are relatively high, but that a flaring tax near the social cost of flaring could substantially reduce the practice. Lastly, I estimate that firms do not consider spillovers for their neighbors which leads to less drilling and higher emissions from flaring than in a counterfactual world where these spillovers were internalized.

These questions are important because flaring likely has large social costs. The U.S. EIA estimates that in 2020 about 420 million Mcf of natural gas was flared (or vented) in the U.S (U.S. Energy Information Administration, 2021b). Natural gas primarily consists of methane (CH4) a very potent greenhouse gas, but when it is burned it is mostly converted to CO2. Flaring is less socially costly than simply letting the methane escape into the atmosphere (venting).[2] However, there is likely a wedge between the social cost of flaring gas at the wellhead and the social cost of selling it for end use precisely because flares are oftentimes inefficient or unlit, allowing methane to escape into the atmosphere. A study by the Environmental Defense Fund found that 7% of gas flared is actually released directly as methane. Their estimates imply that flares in the Permian Basin release 300,000 metric tons of uncombusted natural gas per year–resulting in around a half billion dollars in external damages just from methane from flares (EDF, 2022). Flaring is mostly unpriced in the U.S. In fact, most leases do not contain royalty clauses on flared gas, and most states do not tax flared gas. Moreover, pipelines can be quite expensive. The Interstate Gas Association of America (INGAA) reported that in 2016 an 8-inch gathering pipeline costs $\$264,856$ per mile (ICF, 2018). Thus, flaring can be an attractive option for producers – especially when gas prices are low, when the expected level of gas production from a well is low, or when the well is very far away from existing pipelines.

In my model, drillers hold leases with time-limited options to develop parcels of land by drilling a well. As more pipelines are built over time, connection costs for new wells decline.

[2] Assuming a discount rate of 3% the social cost of carbon is around \$50 per ton while the social cost of methane is \$1500 per ton (Interagency Working Group on Social Cost of Greenhouse Gases, 2021b). An Mcf of gas weighs around 50lbs. When burned it is converted into about 121 pounds of CO2. Therefore if 100% is converted to CO2 the social cost of burning an Mcf of gas is around \$3. If instead, the same amount of gas is vented, this would result in whopping \$37.5 in damages. Thus, it's easy to see why the social cost of flaring is heavily dependent upon the efficiency of the flares – which is still hotly debated. EDF (2022) find that 7% of flared gas is released as methane and a recent study by Plant et al. (2022) estimate that 9% is. In contrast, the EPA assumes this number is 2% (40 CFR 98.233, 2010).

Producers have expectations over the evolution of prices and connection costs, and choose when to drill and whether to flare or connect to pipelines. The state space of my model is too large to compute a full solution method (through backward induction) so I leverage the insights of Hotz and Miller (1993) and Arcidiacono and Miller (2011) and use a two-step estimator with conditional choice probabilities (CCPs) to estimate the parameters of the model. Estimation of the model uses variation in expected revenues, which comes through changing prices and differences in lease-level geology, and variation in distance to pipelines to back out the implied fixed costs of drilling, the implied costs of building pipelines, and whether spillovers are internalized.

This paper makes a few important contributions to the literature. First, this paper is the first to explore the endogenous relationship between drilling and gas pipeline construction. Second, this paper is the first paper which explores how policies aimed at reducing flaring might affect the drilling margin. Third, this paper is the first which estimates pipeline costs directly instead of relying on reported costs. Lastly, this paper estimates whether firms internalize spillovers to their neighbors in a new setting – pipeline construction. Combined, these contributions make this paper an important contribution to the oil and gas literature and the best model of how potential policies to reduce flaring would affect the drilling and connecting decisions of firms.

The closest paper is perhaps Lade and Rudik (2020). This paper presents a static model of wellpad-level flaring decisions which rules out spillovers and assumes the drilling margin is fixed. They demonstrate that a policy instituted in North Dakota mandating a firm-level flaring standard is far more costly than a tax which generates an equivalent reduction in flaring due to heterogeneous costs across firms. They do this by constructing firm-level marginal abatement curves using estimates of pipeline costs from an industry report. In contrast my model considers dynamics, allows for producers to internalize spillovers, and allows policy to affect the drilling margin. I also estimate pipeline costs directly rather than relying on industry data. I find much higher abatement costs and demonstrate that these estimates do a better job of rationalizing the observed flaring outcomes.

This paper also relates more generally to the economics literature on oil and gas development. It's common in this literature to treat the decision to drill as a dynamic optimization problem (see Paddock, Siegel, and Smith (1988), Kellogg (2014)). This paper is the first that borrows the use of CCPs to estimate the parameters of a drilling function, demonstrating that credible estimates of the parameters can be estimated with this computationally inexpensive method. Next, the context of this paper is quite similar to that of Covert and Kellogg (2017) who study the choice between shipping crude via rail (which is environmentally costly and has high variable costs) and via pipeline (which has high investment costs, negligible variable costs, and significantly lower environmental damages). They find that pricing the externalities of rail shipping leads to a 12-29% increase in pipeline capacity. In my context the "dirty", low fixed-cost option is flaring, while the clean high fixed-cost option is constructing a pipeline. I find that pricing the externalities of flaring leads to a negligible increase in pipeline investment but a sizable reduction in flaring.

My model estimates indicate that industry reported estimates of pipeline costs underestimate the true costs of pipeline building and that firms do not consider spillovers to their neighbors. I run counterfactual simulations to explore the importance of spillovers and evaluate policy options to reduce flaring. My counterfactual simulations suggest that the benefits of spillover internalization are relatively small. Thus a sensible explanation for the lack of spillover internalization is simply that the costs of contracting or negotiating might simply be higher than the benefits. My policy counterfactuals suggest that a tax on flaring close to its social cost would substantially reduce flaring. A $\$5/\text{Mcf}$ tax reduces flaring by 39% while a $3\$/\text{Mcf}$ tax reduces flaring by 29%.

1.2 Background

Oil and gas are often co-produced in wells. Oftentimes, the driller is primarily interested in the oil and gas is treated as a by-product. This happens when the oil-to-gas ratio at a potential well site is high, when gas prices are very low, or when the well is far away from existing infrastructure. When the costs of collecting the gas are high, it is a common practice for drillers to flare the gas. Texas allows producers to flare for up to 10 days after a new well completion, then requires a permit to continue flaring. The Texas Railroad Commission (RRC) states that to receive a permit, the operator "must provide documentation that progress has been made toward establishing the necessary infrastructure to produce gas rather than flare it" (Texas Railroad Commission, b). However, the number of flaring authorizations has ballooned over the past decade with the increase in fracking. Figure 1 plots the number of exemptions over time, which I've obtained via a Public Information Act request. In practice, many permits are issued because the gas is "uneconomical to collect." Moreover the Texas RRC has a reputation for "rubber stamping thousands of flaring permits without requiring oil companies to come up with a plan to curb the practice" (Chapa, 2021). In fact, between 05/02/2021 and 05/02/2022 401 applications for permanent exemptions have been approved and zero have been denied.[3]

This paper studies flaring caused by producers choosing to not build pipelines. However, other causes of flaring do exist. Agerton and Upton (2020) show that substantial portions of flaring come from wells that have previously sold gas–implying that the wells are indeed connected

[3]Beginning on 05/02/2021, the Texas RRC implemented online applications for flaring permits, and has made this data publicly available at `https://webapps.rrc.state.tx.us/swr32/publicquery.xhtml`.

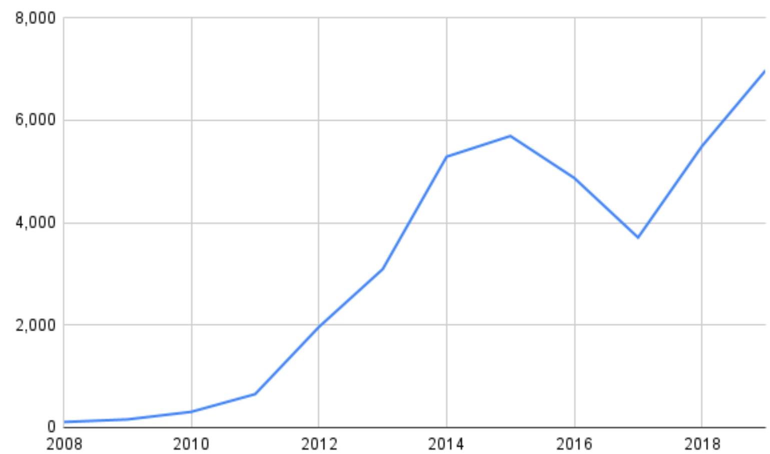


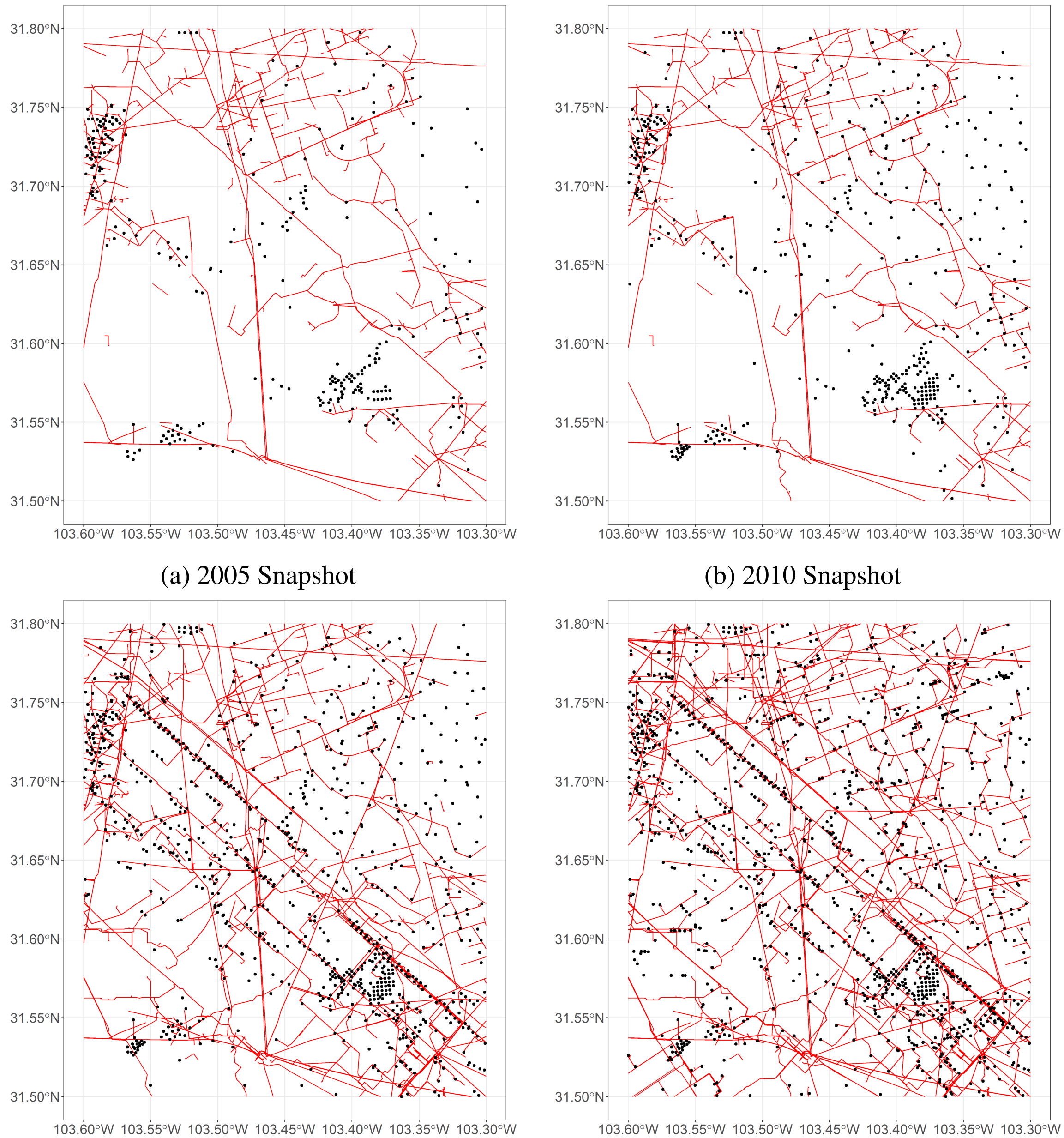
Figure 1: Number of Flaring Exceptions Issued by the Texas RRC by Year

but must flare due to pipeline capacity or processing constraints. In this paper, I abstract away from processing and capacity constraints since I cannot directly observe processing and pipeline capacity utilization rates.[4]

Producers drilling in new fields often drill first, then wait to build infrastructure until they are sure the well will be productive, or until gathering and processing capacity becomes available in the area. This leads to any produced gas being flared during the interim. In figure 2 I show completed wells and permitted pipelines at four points in time. In the first two panels, there are many wells that are unconnected (and therefore flaring any associated gas), but will eventually go on to be connected by the last panel.

The pipeline networks which collect gas likely exhibit some network properties. Once a line is built, more connections can be added to that line at a later date (subject to capacity constraints). This means that the costs of connection at a given site likely decrease over time as new infrastructure creeps closer to that site. In figure 3 you can see that between 2005 and

[4]Moreover, the analysis of Agerton and Upton (2020) seems to complement the main idea of this paper. They show that the proportion of gas from unconnected wells in three out of four basins decreases over the 2008-2020 period, consistent with the idea that the practice of flaring without connecting to a pipeline will be more prevalent earlier in a basin's productive life and decrease as the network get built out and as infrastructure becomes available.



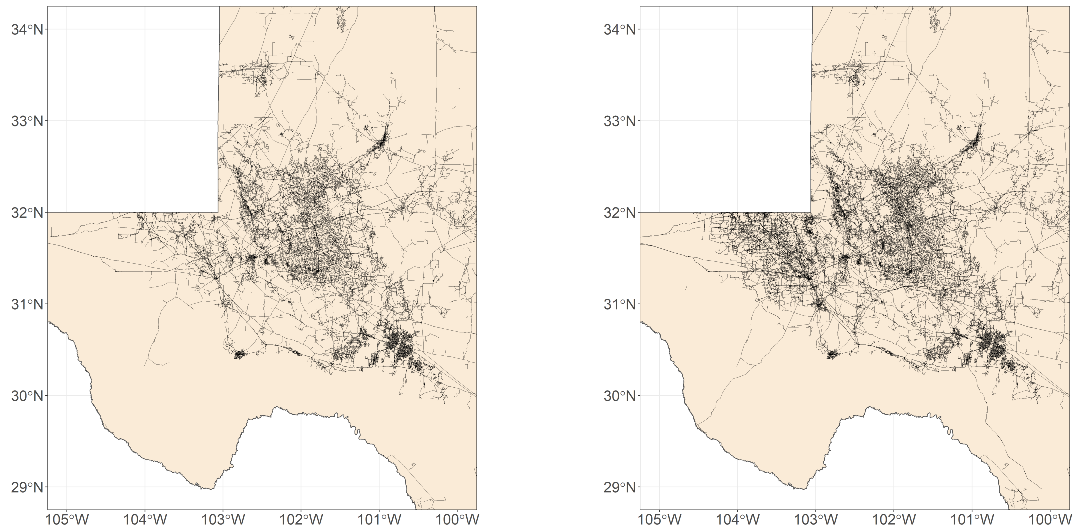
(a) 2005 Snapshot

(b) 2010 Snapshot

(c) 2015 Snapshot

(d) 2020 Snapshot

Figure 2: Snapshots of the development of a field from 2005-2020. Black dots are the locations of well surface holes, and red lines are natural gas pipelines permitted by that date.



(a) 2005 Pipeline Network (b) 2020 Pipeline Network

Figure 3: 15 Years of Network Expansion

2020, the pipeline network in the Permian Basin, especially the westernmost parts, fills out substantially. In theory, contracting would likely be able to internalize this network externality. However, little research has been done on whether this happens and to what degree.

Typically drillers will contract out with mid-stream gas gathering companies to have their wells connected. These gathering companies are one way that network externalities might be internalized. For example, a gathering company might be willing to charge below marginal cost to connect a well that would generate lots of spillovers (and hence contracts for the gathering company in the future). However, the industry is mixed between independent well operators and gatherers and larger producers which are vertically integrated and have the ability to both drill and build pipelines. In this paper, like many others in the literature, I abstract away from these firm-type heterogeneities.

1.3 Data

1.3.1 Lease data

I use lease data collected by Enverus (formerly DrillingInfo). This lease data provides the spatial polygon of the lease as well as observables such as the grantor, grantee, effective date, expiration of primary term, and royalty rate. I process the raw lease data through two layers of clustering, described more in the appendix. The first issue in the lease data is that there are often multiple rows identifying what is clearly the same lease. This seems to be primarily a result of undivided mineral interests.[5] To deal with this, I follow many of the lease processing choices of Herrnstadt, Kellogg, and Lewis (2020). Namely, I use a hierarchical clustering algorithm on the observables of leases to collapse these into single rows.

Next, I identify lease amendments by also using a hierarchical clustering algorithm, this time, only on the spatial data. For example, there might be two rows of data corresponding with the same polygon. One has a primary term from 12-01-2005 through 12-01-2008, the other has a primary term running from 12-01-2007 through 12-01-2010. I interpret this as a lease amendment whereby the primary term is extended mid-lease. More details on the processing of the lease data can be found in the Appendix.

In my analysis I only use leases from the Permian Basin.[6] The reason for this is primarily computational. Different basins have vastly different geological characteristics, and time-profiles of drilling. In my computational model, I would be unable to compute basin-time fixed effects, so I simply select the Permian Basin, the most active basin in Texas for my time sample (2010-

[5]For example, multiple members of a family might each show up as a distinct grantor with her own row in the data.

[6]I use U.S. EIA boundaries to define the Permian basin. These boundaries can be found at the EIA.

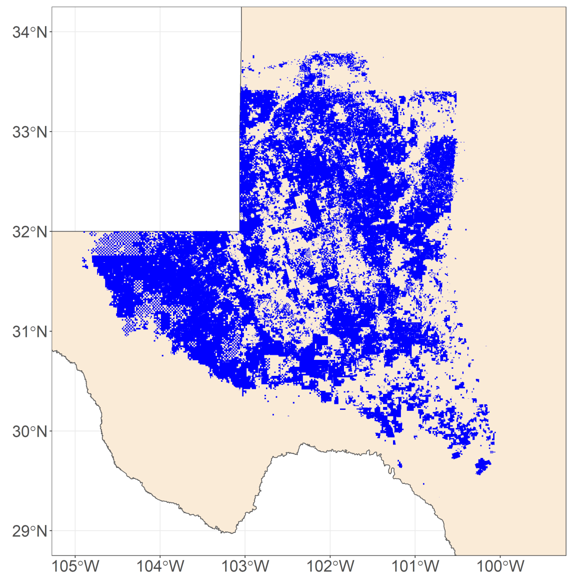


Figure 4: Leases in the Sample

2019). Figure 4 plots the leases in my estimation sample.

1.3.2 Production data

I use well-level monthly production data obtained from Enverus to predict expected production for leases (both drilled and never drilled). Enverus provides data on well-level monthly production of both oil and gas. It also contains information on the observables of the well such as the spud and completion dates, and the depth, lateral length, and location. To construct expected oil and gas production from lease i at t, which I call ξ_{it}^{o} and ξ_{it}^{g}, respectively, I calculate the inverse-distance weighted mean of expected production from all wells completed within the last 18 months of t within 20km of i. Section 6.4 provides more detail on how I construct expected production.

Unfortunately, the Enverus data does not provide information on flared versus sold volumes at the well-level so I gather this from the Texas Railroad Commission (RRC). The RRC provides monthly production data at the lease-level (not the same lease definitions as Enverus), with gas volumes broken down into flared, sold, and vented volumes. I infer a connection date for each lease as the earliest date when gas is sold. I merge that connection date into the Enverus production dataset using an RRC lease-Well API crosswalk.

1.3.3 Pipeline data

I obtain geospatial pipeline data from the Texas Railroad Commission (RRC). I also collect data on when individual pipelines were permitted from 2010 through 2019, which allows me to construct the evolution of the gathering network over those 9 years. Unfortunately, these permitting dates are not necessarily construction dates and construction typically happens after permitting. Moreover, I am unable to observe expansions to a line under the same permit, so these data contain errors where the direction of bias is ambiguous.

1.3.4 Data processing and panel construction

I begin by turning the lease data into a quarter-lease panel dataset with a start date of 2010-01-01 and an end-date of 2019-12-01. I spatially join the lease data with the Enverus well data. A lease becomes “drilled” if I observe a well spudded between the expiration and effective dates and intersect the lease polygon. I call a lease “unconnected” if either zero gas is sold for the first year of the lease or if there are at least 200 days between the first production dates and first connection date (inferred from Texas RRC data) and over 70% of the RRC lease’s first-year of produced gas is flared.

Before dropping any data, there are 64,948 leases and 617,846 lease-quarter observations.

I drop 18,989 lease-quarter observations (3%) due to missing distance or expected productivity. Another 37,319 are dropped because of missing neighbor observables (which are needed to calculate the spillover effects).

In addition I am missing observations on whether a drilled lease becomes connected. There are 24,159 drilled leases and I cannot determine how much 851 of those flared. I drop all lease-quarter observations for those wells. I also randomly drop a proportionate amount of never-drilled leases since my estimation strategy hinges on being able to estimate the probability of drilling and connecting. I am left with 57,776 leases and 528,883 lease-quarter observations. Summary statistics for the leases are presented in Table 1.

Table 1: Summary Statistics

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
Primary Term Length	57776	38.775	15.252	-2	36	36	360
EffectiveYear	57776	2013.026	3.405	1995	2011	2016	2020
ExpirationofPrimaryTerm	57776	2016.288	3.398	2010	2014	2019	2045
Unconnected	23575	0.099	0.299	0	0	0	1
Expected Discounted Oil	57776	0.005	0.005	0	0.001	0.008	0.024
Expected Discounted Gas	57776	0.02	0.033	0	0.002	0.024	0.432
Start Distance	57776	1.478	3.522	0	0	1.345	34.239
End Distance	57776	1.402	3.508	0	0	1.127	34.239
Royalty	33514	0.232	0.032	0.01	0.2	0.25	3.1

Most leases have a primary term length of three years. Both the 25th percentile and 75th percentile lease have primary term lengths of 36 months. About 40% of leases become drilled and of those about 10% of leases that are drilled are unconnected. The average lease is about a kilometer and a half away from the nearest pipeline, though many leases have a pipeline that intersects with the lease polygon.

1.4 Descriptive Results

To motivate the setup of my model and confirm that the variation in the observables has the predicted effect on drilling and flaring outcomes I estimate some descriptive regressions. The first set of regressions explores how flaring outcomes conditional on drilling vary across leases. The second set of regressions explores how the lease observables affect the timing of drilling – regardless of flaring outcomes.

First, I divide the spatial extent of the Permian basin into 0.05 longitude $\times$ 0.05 latitude grids.[7] I use this grid to include grid-square fixed-effects in all regressions. The underlying logic is that geology should be similar within a given grid-square. This approach is similar to Covert and Sweeney (2019) who use a 10 mile by 10 mile grid to control for underlying geology to study revenues received by landowners under auctions versus informal negotiations.

Specifically, the equations I estimate are:

$$y_{it} = \beta_1 d_{it} + \beta_2 d_{it}^2 + \beta_3 \xi_{it}^o + \beta_4 \xi_{it}^g + \beta_5 p_t^o + \beta_6 p_t^g + \beta_7 N_{it}^{\text{benefit}} + \gamma_{g(i)} + \epsilon_{it} \qquad (1)$$

where y_{it} is the outcome of interest, d_{it} is the distance from lease i at time t to the nearest pipeline, ξ^g, and ξ^o are expected discounted gas and oil production, p_t are oil and gas spot prices during quarter t, N^{benefit} measures how many leases would have their distance reduced by i connecting, and $\gamma_{g(i)}$ is a grid fixed-effect. The outcomes of interest are quantity flared, quantity of gas sold, the percentage of gas flared, and a dummy for whether the lease is unconnected during

[7] In the Permian, these grids are approximate 10 square miles – smaller than Covert and Sweeney (2019) use in their paper with a similar research design.

its first year. I expect distance to be correlated with increased flaring, high natural gas prices to be correlated with decreased flaring, and high oil prices to be correlated with increased flaring. The coefficients on N^{benefit} will be negative if leases which could generate higher spillovers for their neighbors are more likely to connect.

Results from these regressions are presented in table 2. All dependent variables are standardized to have a mean of zero and a standard deviation of one. The first three columns present OLS results with standardized dependent variables, the last column presents the coefficients of a logistic regression with the binary dependent variable, 1(Unconnected). The last column is the primary column of interest since it maps to the outcomes in the full dynamic discrete choice model. Focusing on column (4), the effect of distance is quite large. A one standard deviation increase in distance increases the odds of a well being unconnected by 68% ($\exp(0.576 - 0.057)$). More productive wells tend to be far less likely to remain unconnected. However, looking at columns (1), (2) and (3) suggest that wells with higher expected oil productivity tend to both flare more and sell more gas, while wells with higher expected gas productivity do not flare more gas than their less productive counterparts. The effect of prices on all of the dependent variables appears relatively small, though the coefficients have the expected sign and many are significant. Higher oil prices are associated with higher percentages of gas flared, while high gas prices are associated with lower percentages of gas flared. Lastly, none of these regressions present any evidence that leases are more likely to connect if they have neighbors that might benefit from that choice.

Next I show how the observables are associated with drilling probabilities using a similar

Table 2: Effect of Observables on Flaring Outcomes Conditional on Drilling

	Percent Flared (1)	Q Flared (2)	Q Sold (3)	Unconnected (4)
Distance	0.238***	0.006	-0.013	0.576*
	(0.073)	(0.048)	(0.054)	(0.350)
Distance^2	-0.025*	-0.003	0.008*	-0.057
	(0.014)	(0.008)	(0.004)	(0.050)
Expected Oil Productivity (ξ^o)	0.005	0.141***	0.116***	-1.24***
	(0.023)	(0.021)	(0.026)	(0.245)
Expected Gas Productivity (ξ^g)	-0.096***	-0.019	0.234***	-0.452**
	(0.035)	(0.034)	(0.043)	(0.205)
Oil Price (p_t^o)	0.056***	0.002	-0.051***	0.082
	(0.015)	(0.015)	(0.012)	(0.107)
Gas Price (p_t^g)	-0.019*	-0.007	0.011	-0.005
	(0.011)	(0.014)	(0.010)	(0.083
N Benefiting Leases	0.008	-0.007	-0.013	0.121
	(0.014)	(0.011)	(0.019)	(0.153)
Family	OLS	OLS	OLS	Logit
Observations	18,846	19,737	19,737	5,135
R^2	0.627	0.503	0.698	
Within R^2	0.015	0.012	0.088	
Squared Correlation				0.288
Pseudo R^2				0.268
BIC				8,562.2
Grid fixed effects	✓	✓	✓	✓

Notes: Each column displays estimates of equation 1. Standard-errors are clustered at the grid-level. All independent variables are standardized.

grid fixed-effect strategy. I run a logit of:

$$\mathbb{1}(i \text{ drilled during } t) = \beta_1 d_{it} + \beta_2 d_{it}^2 + \beta_3 \xi_{it}^o + \beta_4 \xi_{it}^g + \beta_5 p_t^o + \beta_6 p_t^g + \beta_7 N_{it}^{\text{benefit}} + \gamma_{g(i)} + \epsilon_{it} \quad (2)$$

Results are presented table 3. Again, distance has a strong negative effect on the likelihood of drilling while expected oil production has a strong positive effect. Here, the characterization of drillers being primarily interested in oil seems true. Higher expected oil productivity and higher oil prices are both strongly associated with increases in the propensity to drill, while neither expected gas productivity or gas prices have significant coefficients. A one standard deviation increase in ξ^o is associated with a 127% increase in the likelihood of drilling while a one standard deviation increase in the price of oil is associated with a 38% increase in the likelihood of drilling.

The coefficient on the number of benefitting leases is negative. If firms internalized benefits for their neighbors, we would expect that coefficient be positive. However, N benefitting leases is likely correlated with a lease's number of undrilled adjacent leases and firms might prefer to wait to see if adjacent tracts are productive before drilling or might be interested in the productive inputs used by neighboring firms.[8] These issues have been previously explored by Décaire, Gilje, and Taillard (2019), Covert (2015), and Covert and Sweeney (2022).

1.5 A Dynamic Model of Drilling and Flaring

I formulate the drillers' problem as a discrete-time optimal stopping problem. In each quarter up until and including the time of the leases expiration, the lessee observes a vector of

[8]When a well is drilled, the depth is usually observable by neighbors. The inputs in the fracking job are also required to be disclosed on fracfocus.org, so firms can learn about the best inputs for production by learning from their neighbors.

Table 3: Fixed-Effect Logit of Observables on Drilling Outcomes

	Drill	
	(1)	(2)
Distance	-0.617***	-0.763***
	(0.187)	(0.134)
Distance^2	-0.001	0.025*
	(0.022)	(0.014)
Expected Oil Productivity (ξ_o)	0.417	0.819***
	(0.394)	(0.104)
Expected Gas Productivity (ξ_g)	-0.216	-0.093
	(0.510)	(0.085)
N Benefiting Leases	-0.433***	-0.199***
	(0.068)	(0.065)
Gas Price (p^g)		0.039
		(0.045)
Oil Price (p^o)		0.325***
		(0.059)
Standard-Errors	Grid-date	Grid
Observations	30,087	200,341
Squared Correlation	0.315	0.091
Pseudo R^2	0.338	0.212
BIC	44,158.5	56,027.4
Grid-date fixed effects	✓	
monthstilexp fixed effects	✓	✓
Grantee fixed effects	✓	✓
Grid fixed effects		✓

Note: Each column displays estimates of equation 2. Standard errors are clustered at the grid level. All independent variables are standardized.

the state variables, $\vec{x}_{it}$ and makes the decision to either wait, drill and flare, or drill and gather. I denote this set $j \in \{W, F, G\}$. Once the lessee decides to drill, either with $j = G, F$, that period becomes the terminal period. The operator then receives the expected sum of discounted revenue from production and pays the drilling and connecting costs.[9] This means the drillers' problem is a finite-horizon optimal stopping problem with two potential actions for stopping, and can be formulated as a dynamic programming problem with a Bellman equation.

Thus, the firm's problem is

$$V(\vec{x}_{it}, \epsilon_{it}) \equiv \max_j E\left[\sum_{t=1}^{T} \beta^{t-1}(\pi(j, \vec{x}_{it}) + \epsilon_{itj})\right] \tag{3}$$

where T is expiration quarter of the lease, $\pi(j, \vec{x}_{it})$ is the per-period profit of taking action j, with observable lease state variables, $\vec{x}_{it}$. ϵ_{itj} is an unobserved shock assumed to be i.i.d. Type-I Extreme Value with scale parameter σ. Unlike discrete choice estimation of consumer preferences where the outcome is scale-less indirect utility, σ is identified since expected revenues are being measured directly in dollars. The set of state variables, $\vec{x}_{it}$, include i's distance to the nearest gathering line at t, d_{it}, the price of oil and gas, p^o_t, p^g_t, the expected sum of discounted production, ξ^o_{it}, ξ^g_{it}, the effect of building a pipeline on i's neighbors $\Delta\tilde{V}_{it}$, the time t (and hence the time until i's expiration date), and the royalty rate of the lease, r_i.[10]

[9] An underlying assumption here is that each lease can only be drilled once. This is clearly not true - many parcels could contain multiple wellsites. However, this is a simplifying assumption also made by Herrnstadt, Kellogg, and Lewis (2020).

[10] I only consider pipeline construction spillovers. I do not, for example, think about common pool problems that can arise in some natural resource extraction problems (see **?** for an exploration of common pool problems and lease option exercise). The common pool problem is not a major issue in shale formations, since hydrocarbons are only released after fracturing the rocks under high pressures, causing little or no impact on neighboring parcels (Décaire, Gilje, and Taillard, 2019).

The firm's maximization problem can be expressed in Bellman form as:

$$V(\vec{x}_{it}, \epsilon_{it}) = \max_j \{\pi(j, \vec{x}_{it}) + \epsilon_{itj} + E[V(\vec{x}_{it+1}|\vec{x}_{it})]\} \tag{4}$$

The deterministic part of per period payouts are given by:

$$\pi_i(j, \xi^o_{it}, \xi^g_{it}, p_t, d_{it}, \Delta\tilde{V}) = \begin{cases} R^o_i(p_t, \xi^o_{it}) - C^F_t & j = F \\ R^o_i(p_t, \xi^o_{it}) + R^g_i(p_t, \xi^g_{it}) - C^G_t(d_{it}, \Delta\tilde{V}) & j = G \\ 0 & j = W \end{cases}$$

where R^o and R^g are the expected revenue from oil an gas, respectively, and C^F, C^G are the the costs of drilling and flaring and drilling and gathering. Revenues are calculated as expected discounted production (ξ) multiplied by the spot price less royalties and state severance taxes (4.6% for oil and 7.5% for gas). Royalties often depend upon the lessor and for private land, are often negotiated along with the up-front price of the lease.

$$R^o_i(p_t, \xi^o_{it}) = \xi^o_{it} p^o_t (1 - r_i - 0.046)$$

$$R^g_i(p_t, \xi^g_{it}) = \xi^g_{it} p^g_t (1 - r_i - 0.075)$$

I assume costs can be decomposed into:

$$C^F_t = \delta^F + \delta_{y(t)}$$

$$C_t^G = \delta^G + \delta_{y(t)} + \alpha_d d_{it} - \alpha_{\tilde{V}} \sum_k \Delta\tilde{V}_{ik}$$

δ^F and δ^G are the fixed costs associated with flaring and gathering, respectively. $\delta_{y(t)}$ are year fixed effects that allow the price of drilling to move over time. $\delta_{y(t)}$ capture changes in the rental rate of drilling rigs as well as changes over time in average well depth or complexity. $\Delta\tilde{V}_{ik}$ is a measurement (in the same units $\tilde{V}$ is denoted in) of how much building a pipeline to i increases the value of a neighboring lease, k, and will be discussed in greater detail in the next section. The logic behind this cost function is that pipeline connection costs will be a function of overall distance, but that there might be a discount to the drilling firm if the pipeline is expected to be valuable for the gathering firm in the future. $\alpha_{\tilde{V}}$ will measure the proportion of spillover benefits internalized by firm i.

Define the ex-ante value function as the expected value of V_{it} before the revelation of ϵ_{it}:

$$\bar{V}_{it}(\vec{x}_{it}) \equiv \int V_{it}(\vec{x}_{it}, \epsilon) f(\epsilon) d\epsilon \tag{5}$$

where $f(\cdot)$ is the distribution of the error term, ϵ.

Finally, let $\Omega(\cdot)$ be the transition probabilities of the state variables and define the conditional value function as the value of making a choice j in period t (less the error term), then behaving optimally from $t+1$ onward:

$$v_{it}(j, \vec{x}_{it}) \equiv \pi_i(j, \vec{x}_{it}) + \beta \int \bar{V}_{it+1}(\vec{x}_{it+1}) \Omega(\vec{x}_{it+1} | \vec{x}_{it}) d\vec{x}_{it+1} \tag{6}$$

Since the error terms are assumed to be type-I Extreme Value, the choice probabilities are given by the familiar logit form:

$$Pr(i \text{ chooses } j|\vec{x}_{it}) = \frac{\exp(v_{it}(j, \vec{x}_{it})/\sigma)}{\sum_k \exp(v_{it}(k, \vec{x}_{it})/\sigma)} \tag{7}$$

1.6 Estimation

1.6.1 Dynamic Discrete Choice Estimation

Because of the assumption on the error term, ϵ, Hotz and Miller (1993) and Arcidiacono and Miller (2011) show that the ex-ante value function admits the closed form:

$$\bar{V}_{it}(\vec{x}_{it}) = -\sigma \ln(Pr(j = \tilde{j}|\vec{x}_{it})) + v_{it}(\tilde{j}, \vec{x}_{it}) + \sigma\gamma \quad \forall \tilde{j} \tag{8}$$

where γ is Euler's constant. The intuition here is that the ex-ante value function can be found by evaluating the conditional value function, v, at any $\tilde{j}$, then applying a correction which is larger when the probability of choosing $\tilde{j}$ is smaller. The constant term at the end, $\sigma\gamma$ is simply the expected value of ϵ. I omit this term for the remainder of the paper since choice probabilities are invariant to an additive constant in each v_j.

Since $\bar{V}(\vec{x}_{it})$ can be evaluating using any choice $\tilde{j}$, I can select a terminating action $\tilde{j} = G$, then

$$\bar{V}_{it}(\vec{x}_{it}) = -\sigma \ln(Pr(j = G)|\vec{x}_{it}) + \pi(G, \vec{x}_{it}) \tag{9}$$

Equation 9 can be substituted into the conditional value function, equation 6 to yield:

$$v_{it}(j, \vec{x}_{it}) = \pi(j, \vec{x}_{it}) + \beta \int [-\sigma \ln(Pr(j_{t+1} = G|\vec{x}_{it+1})) + \pi(G, \vec{x}_{it+1})]\Omega(\vec{x}_{it+1}|\vec{x}_{it})d\vec{x}_{it+1} \tag{10}$$

All of the pieces of equation 10 can be constructed from the data. $\pi(\cdot)$ is simply a linear function of the parameters and observables. $Pr(j_{t+1=G}|\vec{x}_{it+1})$ and $\Omega(\vec{x}_{it+1}|\vec{x}_{it})$ can be directly estimated from the data. Since π is linear in the stochastic state variables, it can be brought outside the integral and evaluated at $E[\vec{x}_{it+1}|\vec{x}_{it}]$. To construct $\int -\sigma \ln(Pr(j_{t+1} = G|\vec{x}_{it+1})\Omega(\vec{x}_{it+1}|\vec{x}_{it})d\vec{x}_{it+1}$, I estimate the transition function for $\vec{x}_{it}$, $\Omega(\vec{x}_{it}|\vec{x}_{it+1})$, then take simulation draws of $\vec{x}_{it+1}$, calculate the log probabilities associated with the simulation draws, then average.

For each j, you can express the v as:

$$v(j, \vec{x}_{it}) = \begin{cases} \xi_{it}^{o} p_t^{o} - (\delta^F + \delta_t) & j = F \\ \xi_{it}^{o} p_t^{p} + \xi_{it}^{g} p_t^{g} - (\delta^G + \delta_t + \alpha_d d_{it} - \alpha_V \sum_k \Delta\tilde{V}_{ik}) & j = G \\ \beta[\xi_{it}^{o} E[p_{t+1}^{o}] + \xi_{it}^{g} E[p_{t+1}^{g}] - (\delta^G + E[\delta_{t+1}] + \alpha_d E[d_{it+1} - \alpha_V \sum_k \Delta\tilde{V}_{ik})+ & j = W \\ \sigma \int - \ln(Pr(j_{t+1} = G|\vec{x}_{it+1}))\Omega(\vec{x}_{it+1}|\vec{x}_{it})d\vec{x}_{it+1} & \end{cases} \tag{11}$$

The parameters of to be estimated are $\alpha_d, \alpha_{\tilde{V}}, \delta_C, \delta_G, \delta_t$, and σ. In theory, β is identified also, but its common practice in the dynamic discrete choice literature to simply pick a reasonable value for β. I select 0.98, implying a yearly rate of about 0.92. δ_t and $E[\delta_t]$ are not all identified without some restrictions. The same issue appears in Gowrisankaran et al. (2012) and Murphy (2018). Both papers fit δ_{t+1} and δ_t to an AR(1), though their exact estimation process varies. I choose to estimate

$$\delta_{t+1} = \omega_0 + \omega_1\delta_t + \eta_t$$

within the likelihood function. Then replace $E[\delta_{t+1}]$ with $\hat{\delta}_{t+1}$.

Finally, with all of the v's in hand, I can estimate the parameters of of the model via maximum likelihood by constructing a likelihood function from the choice probabilities given in equation 7.

1.6.2 Transition Probabilities - $\Omega(\vec{x}_{it+1}|\vec{x}_{it})$

Some of the state variables, such as the time fixed effects, transition deterministically. Other state variables, namely $\xi^o_{it}, \xi^g_{it}, \Delta\tilde{V}_{it}$, transition over time. However, I do not want to think of the firm as "waiting for good draws" of these state variables so I assume that their expected value in $t+1$ is equal to their value in t. Kellogg (2014) essentially makes the same modelling assumption.[11]

That leaves transitions in d_{it}, p^o_t, and p^g_t. To simulate changes in d_{it} I simply take draws from the observed distribution in Δd_{it} for different bins of d. Figure 5 plots these distribution of these distance changes.

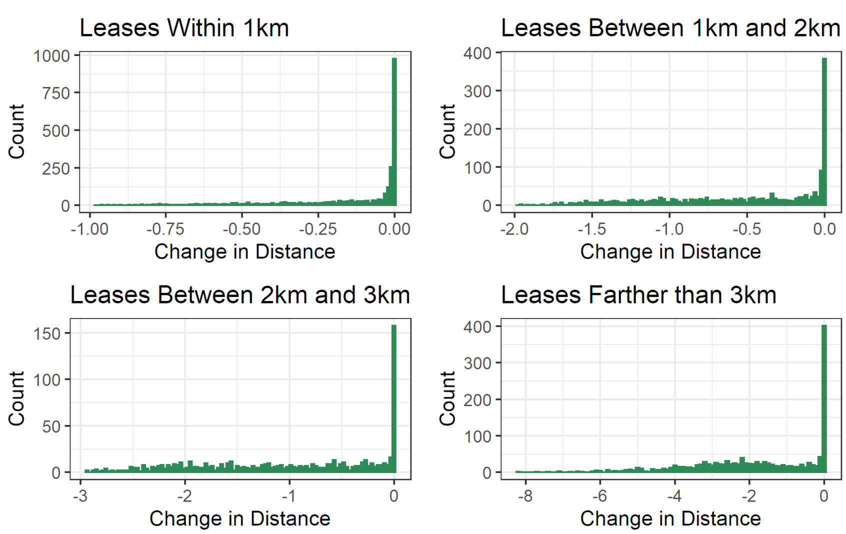


Figure 5: Histogram of Positive Distance Changes (m)

[11] Kellogg(2014) does not so explicitly try to construct expected production. Instead, he assumes production rationalizes costs (whereas I essentially do the opposite – construct expected revenue then estimate costs to rationalize production decisions.). However, in his dynamic model he makes the assumption that producers do not integrate over their expectation of new realizations in expected production to calculate their continuation value of the lease.

To follow estimate the transition of p_t^o and p_t^g I follow Hernnstadt et al (2020) and Kellogg (2014) and estimate the following Markov process for both oil and gas:

$$\ln(p_{t+1}) = \ln(p_t) + \kappa_o + \kappa_1 p_t + \sigma\eta_{t+1} \tag{12}$$

In this price process κ_0 and κ_1 are drift and mean reversion parameters, respectively. Coefficients from the estimation of equation 12 are found table 4. I've also plotted the time series of prices along with some simulations based off of these parameters in figure 6. The estimated coefficients indicate that both price series are process with positive expected drift and mean reversion, but that they are predominantly driven by the error term. You can see that for both processes, the residual standard error is much larger than the constant (drift term).

Table 4: Estimated Markov Process Parameters

	Dependent variable:	
	Δ log oil price	Δ log gas price
	(1)	(2)
Price	-0.002^{**}	-0.022^{**}
	(0.001)	(0.010)
Constant	0.107^{**}	0.096^{*}
	(0.048)	(0.049)
Observations	83	83
R^2	0.059	0.055
Adjusted R^2	0.047	0.043
Residual Std. Error (df = 81)	0.168	0.193
F Statistic (df = 1; 81)	5.047^{**}	4.697^{**}
Note:	*p<0.1; **p<0.05; ***p<0.01	

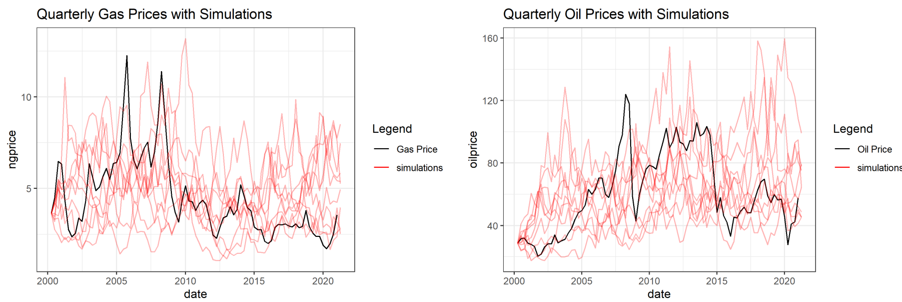


Figure 6: EIA Spot Prices and Examples of Simulated Price Paths

1.6.3 Construction of the spillover term - $\tilde{V}$

Consider the example in figure 7. There are nearby leases, A and B. A is considering building and connecting a well. If A chooses to connect, the future connection costs for B will decrease in the next period by $\alpha_d(|AB| - |B|)$. B should be willing to compensate A up to the difference in B's ex-ante value: $\overline{V_{Bt}}(p_t, d_{Bt} = |AB|) - \overline{V_{Bt}}(p_t, d_{Bt} = |B|)$ to connect.[12] In practice, the construction of pipes and wells are typically completed by separate companies. However, the general principle still applies. A gathering company might be willing to offer a discount to A (to entice it to connect), with the hopes of decreasing its future costs and increasing its expected number of future customers. However, the gathering company is unlikely to be able to capitalize the entire change in $\overline{V}_B$, and therefore will be unable (and potentially unwilling) to pass that entire change on to A. Moreover, there might be other price distortions such as market power that would lead the gathering companies to not pass along a discount to A. My estimation strategy will estimate the proportion of this change in value captured by driller when

[12] Some readers might notice that if the goal is to connect both A and B at minimum cost, the best way to do it would be to construct $|B|$, then draw a line orthogonal to $|B|$ to A. I believe that this is a *planning* issue as opposed to a spillover issue. For example, you would expect if A and B were owned by the same firm they would be more likely to select this minimum distance solution. Méndez-Ruiz (2019) studies whether contracting costs lead to more flaring by estimating a structural model of flaring in the Bakken, and addresses this planning issue more directly. He assumes that contracting costs are minimized within a firm, and estimates the degree to which a group of firms can transact as a single firm by exploiting variation in firm concentration across markets.

the he chooses to connect (which should be between zero and one).

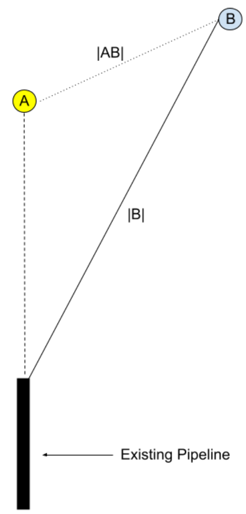
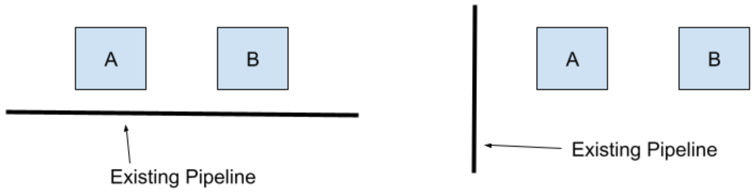


Figure 7: Hypothetical example of spillover effects. This figure shows that lease B's connection distance depends upon the choice of lease A.

I can use the Hotz-Miller identity, equation 9, to simplify this difference to become the sum of the difference in expected log probabilities and the difference in pipeline costs. The parts of equation 13 can be constructed by calculating the counterfactual distance for B if A connects, then constructing the predicted probabilities (as discussed later in section 6.5).

$$\overline{V}_{Bt}(p_t, |AB|) - \overline{V}_{Bt}(p_t, |B|) = \ln(pr(j_t = G||B|, p_t)) - \ln(pr(j_t = G||AB|, p_t) + \alpha_d(|AB| - |B|) \tag{13}$$

To construct $\tilde{V}$, for each lease-quarter I draw the shortest line from that lease to already existing pipes. For all leases $k \in K$ within 2km of that hypothetical line, I save their actual and counterfactual distances. Then $\tilde{V}_{it} = \sum_{k \in K} \Delta\overline{V}_{ik}$ is calculated during each iteration of the likelihood function. Variation in $\tilde{V}$ comes from differences in the likelihood of neighbor lease exercise (which is affected by, among other things, lease productivity, lease distance, and time



(a) The spillover for B of A connecting is zero

(b) The spillover for B of A connecting is positive

Figure 8: Sources of variation in $\tilde{V}$. This figure shows that variation in $\tilde{V}$ can result from geometrical variation in the layout of plots and pipelines.

until lease expiration) as well as geometrical variation in how much A choosing to connect affects B. Consider the simple example below in figure 8. In the first panel, A connecting has no effect on the distance B must traverse to connect. In the second panel, if A connects, that generates a spillover for B.

1.6.4 Construction of expected revenue

To measure expected productivity at each lease at time t, I take an inverse-distance weighted mean of expected production, ξ_i, from all wells completed within the past 18 months of t within 20km of the lease. In the dynamic model ξ_{it} enters as a state variable. Since the time window used to compute ξ moves with t, this means that expected productivity will change over time, despite the fact that the actual geology of the lease is time-invariant. This seems reasonable given that the outcome of neighboring wells likely leads to firms updating their beliefs.

Expected production from each well is an estimated term. To construct estimated production, I make a common assumption in the literature – to follow the Arps Model and assume production decays exponentially from a well (Lade and Rudik (2020), Méndez-Ruiz (2019)).[13] Oil and gas

[13] An assumption used throughout the construction of expected production is that producers take expected production from their leases as a given. This is both technically false and a good simplifying assumption. Producers can affect their production by varying well depth, horizontal length, and amount of fluid used in fracturing. However

production in each time period, t, is given by:

$$o_{jt} = O_{j0} t^{\beta o} \exp(\nu^o_{jt})$$

$$g_{jt} = G_{j0} t^{\beta g} \exp(\nu^g_{jt})$$

where O_{j0}, G_{j0} are the initial rates of production, β are the decline rates, and ν are the model errors.

I estimate the decline rates, β^o and β^g, separately for wells drilled by year by estimating the well fixed-effect equations:

$$\log(g_{jt}) = \beta^g \log(t) + \delta_j + \nu^g_{jt}$$

$$\log(o_{jt}) = \beta^o \log(t) + \delta_j + \nu^o_{jt}$$

Figure 9 plots the estimated $\hat{\beta}^g$ and $\hat{\beta}^o$. For context, a decline parameter of -0.5 implies that 46% of discounted production over the first five years of a well's life will occur during the first year. With the estimated decline parameters in hand I can calculate expected monthly production by observing the well's initial production rate, which I obtain from Enverus. I use second-month production instead of first-month, since the first-month values are often truncated by not having a full month of production. Finally, I obtain expected well-level lifetime production by summing and discounting expected monthly production for each well over a five year period.

by constructing expected production using nearby wells, I am essentially assuming that well-design questions are already enveloped-in and allowing expected production to reflect changes in technology over time and differences in optimal well construction across space.

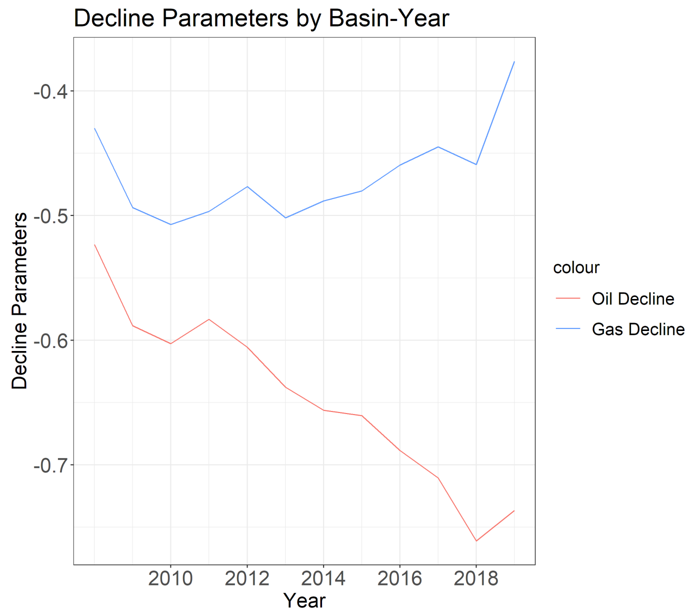


Figure 9: Estimated Permian basin decline parameters by drilling year. Oil decline parameters are plotted in red while gas decline parameters are plotted in blue.

1.6.5 Estimation of conditional choice crobabilities

Ideally, I would be able to construct conditional choice probabilities non-parametrically. However, the relatively large state space makes this impossible. Following other papers in the literature such as Murphy (2018), I estimate conditional choice probabilities by estimating a logit with b-spline expansions. Specifically I estimate:

$$\hat{Prob}\Big(j = G|\vec{x}_{it}\Big) = \Lambda\Bigg(\sum_{s=1}^{5} \phi_s^{\xi^o} S_s^{\xi^o}(\xi_{it}^o) + \sum_{s=1}^{5} \phi_s^{\xi^g} S_s^{\xi^g}(\xi_{it}^g) + \sum_{s=1}^{5} \phi_s^d S_s^d(d_{it}) + \sum_{y=2011}^{2019} \phi_y + \sum_{\tau=0}^{8} \psi_\tau\Bigg) \tag{14}$$

where Λ denotes the logistic CDF, $S(\cdot)$ are the spline basis functions of the argument in the superscript, ϕ_y are year fixed-effects, and ψ_τ are quarter-until-expiration fixed-effects.

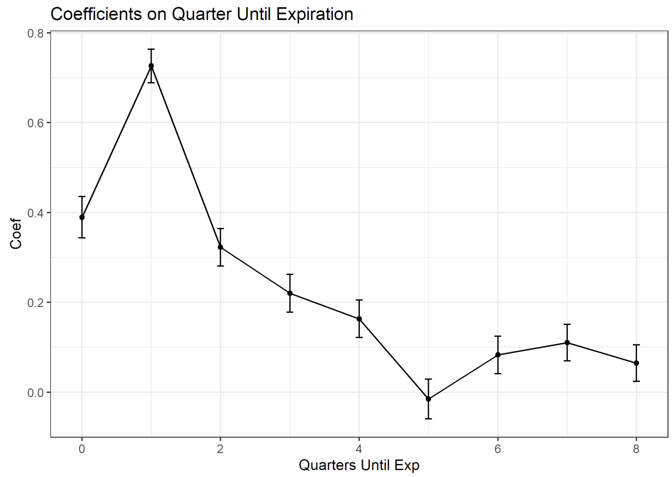


Figure 10: Coefficients on Quarter Until Expiration

1.7 Results

1.7.1 Conditional choice probabilities

The coefficients from the logit match the expected intuition from the underlying dynamic model. Figure 10 plots the coefficients on quarter until expiration from the choice probability logit. In the dynamic model the probability of drilling and gathering should increase exponentially until the time of expiration. The shape of the estimated fixed effects is mostly consistent with this, with the main exception being that the coefficient at the quarter of expiration is lower than the coefficient for one quarter away from the time of expiration. These coefficients are an important source of variation in the expected probability of drilling since they do not directly affect the payout associated with drilling.

1.7.2 Dynamic discrete choice cesults

Results for the dynamic discrete choice can be found in table 5 and the estimated year fixed effects are plotted in figure 12. Standard errors are calculated with 50 bootstrap replications where I sample leases with replacement. Expected revenue was entered into the likelihood as $\$10,000,000$ in revenue, so each parameter can be interpreted as tens of millions of dollars. The scale parameter of the error term is estimated to be $\$7$ million. This implies a standard deviation of the error term of $\$7$ million $\cdot \pi/\sqrt{6} = \$9$ million.

The estimated fixed costs are $\$38,000,000$ and $\$21,000,000$ for F and G, respectively, which are quite high. However, these are the implied costs for a randomly selected lease – most of which are not observed drilling. The implied cost for drilled wells is the estimated fixed cost plus the error term (which is structurally a cost shock). Conditional on drilling the error term is expected to be positive and relatively large. Thus, the expected cost for wells actual observed drilling is far lower. The same issue appeared in Murphy (2018) in his study of parcel development in San Francisco using very similar methods. His estimated cost of development were very high, but the fixed costs of development for parcels which actually developed were far more believable. I find that when the expectation of the error term for drilled wells is included with the fixed costs, most of the implied fixed costs are positive, suggesting that my estimates of revenue are biased downward. Figure 11 plots these implied fixed costs. You can also see that after including the expectation of the conditional error, the fixed costs for gathered wells are greater for gathered wells than flared wells.

The estimated cost of distance is also quite high. The estimated cost per kilometer of pipeline for less than 1km is almost $\$10,000,000$. I estimate that the cost per km is decreasing

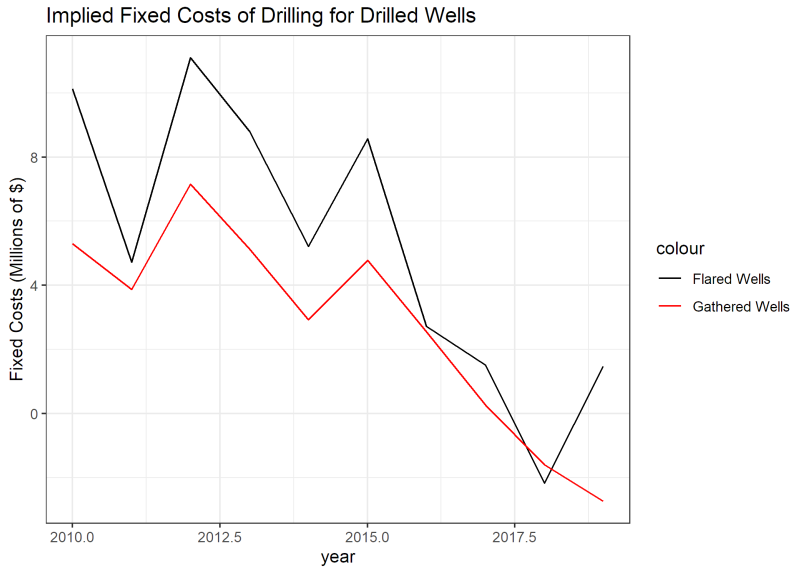


Figure 11: Expected value of fixed costs for drilled wells. For flared wells this is calculated as $C_f + \delta_y(t) + E[\epsilon_{iFt}|i \text{ chose } F \text{ during } t]$. It is calculated analogously for gathered wells.

in overall distance. However, these cost estimates are biased upwards since my counterfactual simulations show that predicted pipeline distance is less than actual distance built – eg if a lease is 1km from the nearest pipeline, in expectation more than 1km will be built. This gap can be explained by two factors. First, pipelines often do not traverse the shortest distance between two points and must go around obstacles or land where a right-of-way could not be secured. Second, I observe many pipelines being constructed in parallel with each other. This implies that some pipelines are at capacity and that the nearest connection for a new well is not necessarily available. Both of these reasons bias my cost estimates upward and will be discussed in greater detail in the following section.

Lastly, the parameter on the spillover term is 0.03, implying that producers do not internalize the vast majority of spillovers for their neighbors. The estimate has a very tight confidence interval that does not overlap with zero. However, the magnitude of the parameter is very close to zero. In order to assess the empirical magnitudes of the cost parameters and of the spillover

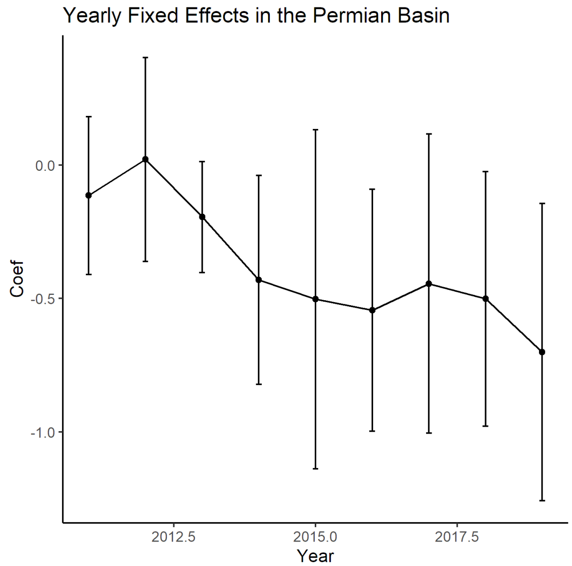


Figure 12: Year Fixed-Effects

internalization parameter, I use a forward simulation procedure described in the next section.

	Estimate	sd	Interval95
Error Scale Parameter (σ)	0.70	0.13	(0.442, 0.962)
Fixed Costs of Flaring (C_f)	-3.87	0.86	(-5.55, -2.183)
Fixed Costs of Gathering (C_g)	-2.13	0.54	(-3.179, -1.083)
Spillovers $\Delta\tilde{V}$	0.03	0.00	(0.02, 0.035)
Distance (d)	-0.96	0.19	(-1.336, -0.582)
$d \times \mathbb{1}(1 \leq d < 2)$	0.01	0.09	(-0.168, 0.194)
$d \times \mathbb{1}(2 \leq d < 3)$	0.36	0.11	(0.146, 0.579)
$d \times \mathbb{1}(3 \leq d)$	0.81	0.17	(0.49, 1.139)

Table 5: DDC Results

1.8 Simulations and Policy Counterfactuals

Having estimated my model, I now proceed to using my model estimates for simulations. The first set of simulations I discuss involve constructing a close analogue to the static marginal abatement curves constructed by Lade and Rudik (2020). Then, I proceed to a full dynamic forward simulation which will allow me to asses the performance of my model estimates, examine the estimated magnitudes of the parameters, and make predictions about drilling and flaring outcomes under varying policy counterfactuals.

1.8.1 Estimated costs perform better than reported costs

To compare my results with previous work I follow the methodology of Lade and Rudik (2020) and construct static marginal flaring abatement cost curves on my own sample using first, the reported cost estimates used in that paper, then my own cost estimates.

Lade and Rudik (2020) calculate the marginal abatement cost for wellpad i as:

$$MAC_{it} = \frac{FC + \alpha_d d_i}{G_{it}}$$

where G_i is the sum of discounted expected production from well i from t onwards. The industry-wide abatement curve is calculated by calculating MAC_{it} for each well, i, then ordering them by cost. Since my unit of analysis is the lease-level I use leases as opposed to wellpads. Lade and Rudik (2020) use a fixed cost estimate of $\$202,000$ lifted from a survey result. They predict α_d by using an ordered probit to predict necessary pipeline width, and pipeline cost estimates lifted from the same survey results.

I construct a simplified version of their cost curves. First I use the same costs they did, lifted from the report from INGAA. However, I do not do an ordered probit and instead make the conservative calculation of assuming all pipelines are 8 inches.[14] Next, I use my own distance and cost measures lifted from my model. The fixed cost estimates are taken as the difference in fixed costs between flared and gathered wells. I use the estimates from figure 11, since they

[14]This is conservative because 8-inch pipelines are the most expensive in their model, and I find that generally these engineering estimates over predict abatement.

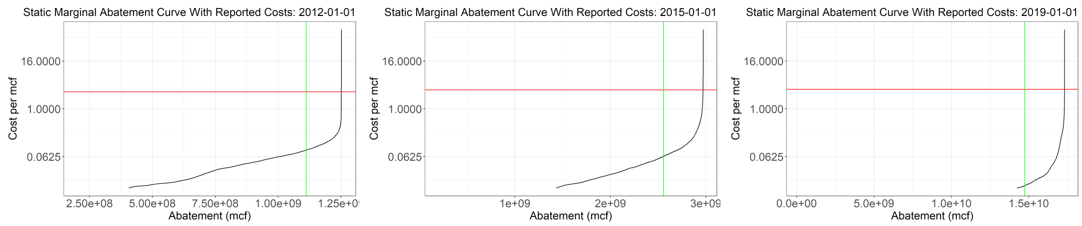


Figure 13: Static Marginal Abatement Curve Calculated with Engineering Estimates of Cost. The red line gives the observed spot price of gas at the time, while the green line gives the observed level of abatement.

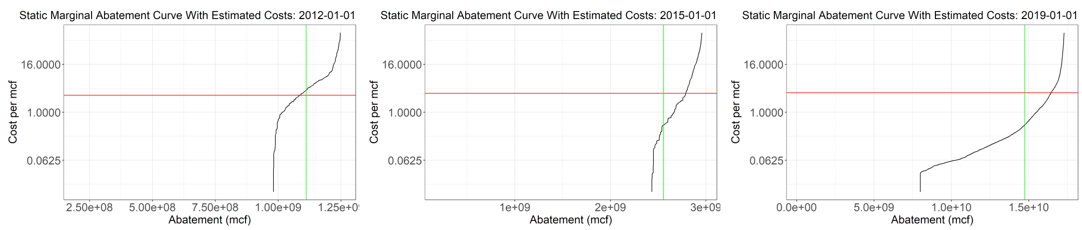


Figure 14: Static Marginal Abatement Curve Calculated with Estimated Costs. The red line gives the observed spot price of gas at the time, while the green line gives the observed level of abatement.

reflect differences in fixed costs for drilled wells as opposed to values from table 5, which reflect estimated fixed costs for randomly selected leases.

Figure 13 plots the marginal abatement curves following Lade and Rudik (2020) and figure 14 plots the marginal abatement curves with my estimated costs. I've also plotted the spot price of gas in red, and the observed level of abatement in green. When using engineering estimates, the observed level of abatement is too low – implying that the cost curves are also too low. When using my own cost estimates, the abatement cost curve shifts upwards, decreasing the wedge between expected and observed abatement.

1.8.2 Full dynamic simulations: computational procedure

To calculate a full dynamic forward simulation, I can no longer rely on having $\log(pr(j = G))$ in hand to calculate $E[V_{t+1}]$. Therefore, I take simulation draws of prices and distance changes, then use backwards induction to calculate the expectation of V. I take draws of ϵ_{it} to determine the action of each lease in each period. For each period, after the gathered leases are determined, I draw a line from the existing network to each lease. Then distances for remaining leases are re-computed for $t + 1$.

Since this process relies on taking draws of distance changes and the distribution of distance changes varies with the policy counterfactual, this routine iterates until the expectation of predicted distance changes is close to the expectation of the distribution of distance changes used to compute the counterfactual. Specifically, I specify that the mean predicted distance change must be within 10m of the mean of the distance changes used to compute the counterfactual. In practice, most counterfactual runs only iterate once. All converge within two iterations. This implies two things – first that the counterfactuals do a very good job of predicting observed distance changes. Second that the network is not particularly sensitive to changes in policy. This will be demonstrated in the following subsections.

Due to computational constraints, I can only compute counterfactuals on samples of the Permian Basin. Figure 15 plots all of the leases in my sample. These boxes were selected since they are from disparate regions of the Permian basin.

1.8.3 Performance of counterfactuals against observed activity

To assess the fit of the simulations I've plotted the predicted versus actual lease outcomes below in figure 16 for each box. Expected productivity is spatially correlated, so when counterfactuals are computed on only a small part of the basin it's unsurprising that predicted flaring or gathering might be systematically above or below the observed level. For example, actual flaring is higher

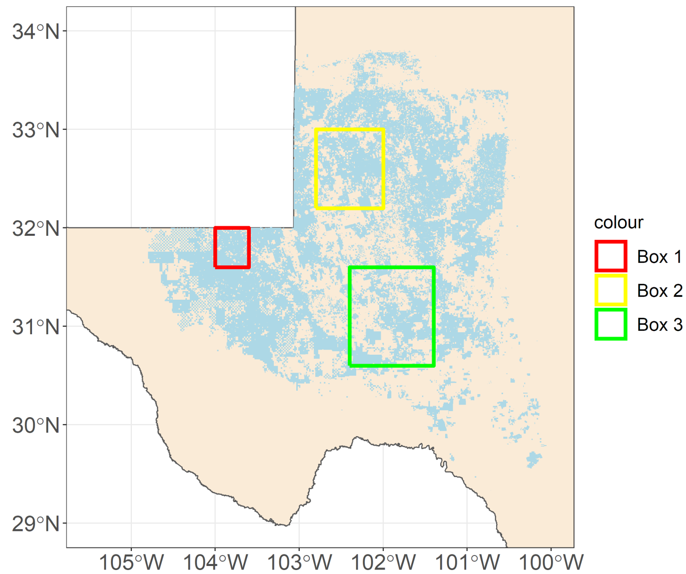


Figure 15: Plot of counterfactual boxes. Leases are plotted behind in light blue.

than the baseline counterfactual in Box 2 and lower in Box 1. However, the counterfactual simulation on Box 3 looks quite good. Overall, you would expect that if I could run counterfactuals on the basin as a whole, the errors across smaller areas would average out and the fit for the overall basin would be good.

As noted in the discussion of the computation of my counterfactuals, the counterfactuals match predicted distance changes with observed distance changes quite well. However, the overall prediction of pipelines built tends to be far too low. A comparison of predicted versus observed pipeline construction over my sample period (01/01/2010-12/01/2019) is in table 6. The discrepancies are likely due to the fact that my model ignores capacity and processing constraints. In practice, I observe many parallel pipelines, which could not be rationalized in my model. However, it does resolve why the estimated per-mile costs of pipelines in my model are so high. Recall that my estimate of pipeline costs for pipeline costs per kilometer was $\$9.6$ million for

Figure 16: Actual Versus Simulated Outcomes

Box	Actual Distance Constructed	Simulation Distance
Box 1	1129.875	93.63384
Box 2	3133.516	116.0214
Box 3	814.2046	268.3943

Table 6: Actual and Predicted Pipeline Distances (km)

distances less than 1km, $\$6$ million for distances between 2km and 3km, and $\$1.5$ million for distances greater than 3km. In Box 2, predicted pipeline construction was off by a factor of 27 – implying that for every well connected that was 1 km away from pipelines, 27km of pipelines was actually built. Adjusting my pipeline costs by this factor would produce a per kilometer estimate of $\$355,555$ for distances less than 1km and $\$55,555$ per km for distances greater than 3km. Predicted construction in box 3 was three times lower than the actual length constructed, implying a per kilometer cost of $\$3.2$ million for distances less than 1km and $\$500,000$ for distances greater than 3km.

1.8.4 Spillover internalization does not appear to be a major driver of flaring

To asses how the importance of the spillover internalization parameter, I run counterfactual simulations in Box 1 where I switch the parameter to 1 – implying perfect internalization. These are plotted in figure 17. This figure plots the number of wells of each type (gathered or flared) drilled in each time period, as well as the percentage of production relative to the baseline simulation. Perfect internalization does seem to induce some wells that had previously been flaring to gather, leading to a small increase in gathered gas volumes and a small decrease in flared gas volumes. I also present total production figures in table 7. Perfect internalization leads to a 5% increase in sold gas volumes and a 6% decrease in flared volumes. Overall, the effects of spillover internalization seem to be relatively small. Nonetheless, given the magnitude of the social costs of flaring, it's plausible that perfect internalization could decrease social costs by tens of millions of dollars per year.

One explanation for the seemingly small impact of the internalization parameter is that the costs of gathering appear to be relatively small compared to the very high fixed costs of gathering. I run a comparison of perfect versus zero internalization where I double the costs of pipelines. These results are also plotted in figure 17. This seems to increase the wedge between the two internalization scenarios ($\alpha_{\tilde{V}} = 0$ and $\alpha_{\tilde{V}} = 1$), but not dramatically.

1.8.5 The effect of a flaring tax, flaring ban, and gas subsidy on drilling and flaring

To assess the performance of potential policies I run simulations in Box 1 with various levels of flaring taxes, a gas subsidy, and a flaring ban. These results are plotted in figure 18, and aggregate production figures can also be found in table 7. A subsidy poorly targets flaring. A $\$1$ per Mcf gas subsidy increases total gas production by 4% but only decreases flaring by 2% at a cost (in just this box) of $\$600$ million. A tax targets flaring quite well. A $\$5$/Mcf tax reduces flared volumes by 39%, increases gathered volumes by 5%, decreases oil produced by 2%, and raises $\$150$ million in revenue.

Flaring volumes seem to respond linearly to taxes at low levels of the tax, then less than linearly as the tax get higher. A $\$1$/Mcf tax reduces flared volumes by 9% while a $\$2$/Mcf tax reduces flared volumes by 20%. More than doubling that tax to $\$5$/Mcf less than doubles the reduction in flared volumes to 39%. Lastly, a ban eliminates all flaring, leads to a small increase in gas production (and hence only a small increase in pipeline investment), and causes a small

	Gas Gathered (10 Bcf)	Oil From Gathered Wells (10 million barrels)	Gas Flared (10 Bcf)	Oil From Flared Wells (10 million barrels)	Total Oil Production (10 million barrels)
Baseline Simulation	57.85	7.58	4.91	0.69	8.27
Subsidy of $\$1$	60.30	7.89	4.82	0.67	8.56
Flaring Ban	59.20	7.78	0.00	0.00	7.78
Tax of $\$0.55$ per Mcf	58.03	7.61	4.68	0.66	8.27
Tax of $\$1$ per Mcf	58.07	7.62	4.45	0.63	8.26
Tax of $\$2$ per Mcf	58.04	7.61	3.90	0.55	8.16
Tax of $\$3$ per Mcf	58.19	7.63	3.48	0.50	8.13
Tax of $\$5$ per Mcf	58.22	7.63	3.00	0.43	8.07
Spillovers Internalized	60.73	7.91	4.62	0.65	8.55
Gathering Costs Doubled	49.53	6.49	5.16	0.72	7.21
Spillovers Internalized & Gathering Costs Doubled	54.73	7.16	4.74	0.67	7.82
Network Fixed at $t = 0$	48.32	6.29	5.11	0.72	7.01

Table 7: Total oil and gas production in Box 1 in various simulations

decrease in oil production. Theoretically, a ban might be attractive since there are multiple market failures (spillover effects and the external costs of flaring). It might be the case that a ban could simply induce flaring wells to wait to drill. However, my empirical results suggest that a ban is a very blunt policy tool in this context, leading to an overall decrease in oil production.

1.8.6 Endogenous network growth has a substantial effect on drilling and flaring

Lastly, I show that allowing the network to grow endogenously is actually quite important for simulation outcomes. I run the simulation but fix lease distances to their values at $t =$2010-01-01. Flared wells increase modestly, but connected wells decrease drastically. Results from this simulation are plotted in figure 19, and aggregate production in this simulation can be found in table 7. Total gas production is 16% lower if the network cannot grow, demonstrating that growth over time is an important determinant of drilling outcomes.

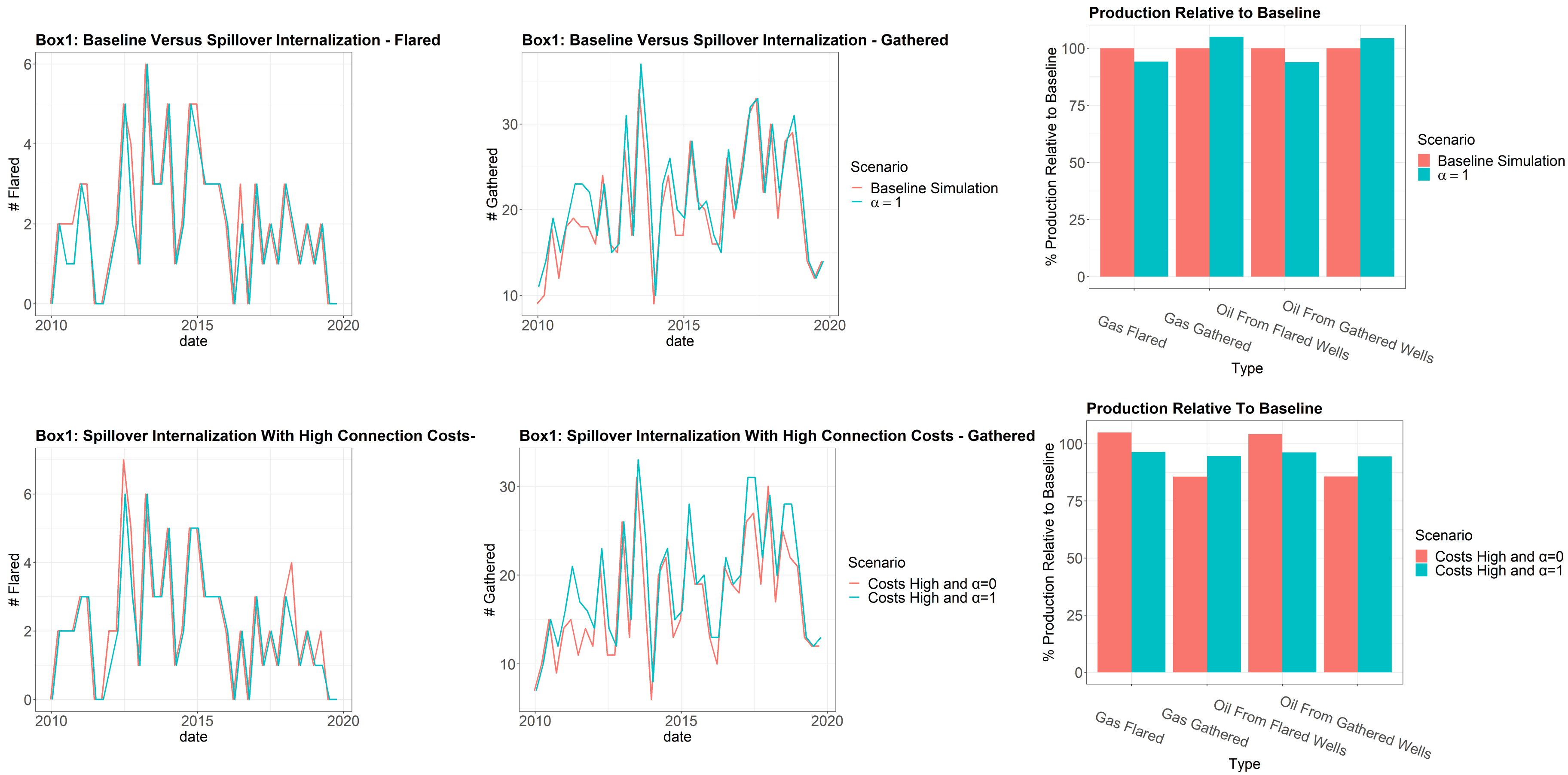


Figure 17: Simulated outcomes with perfect spillover internalization. The line plots present number of wells drilled in each category over time, while the bar charts present total production from each well type relative to the baseline. The graphs in the top panel of this figure only alter the spillover internalization parameter. The graphs in the lower panel multiply the costs of pipelines by 2 to determine whether this would increase the wedge between $\alpha_{\tilde{V}} = 0$ and $\alpha_{\tilde{V}} = 1$.

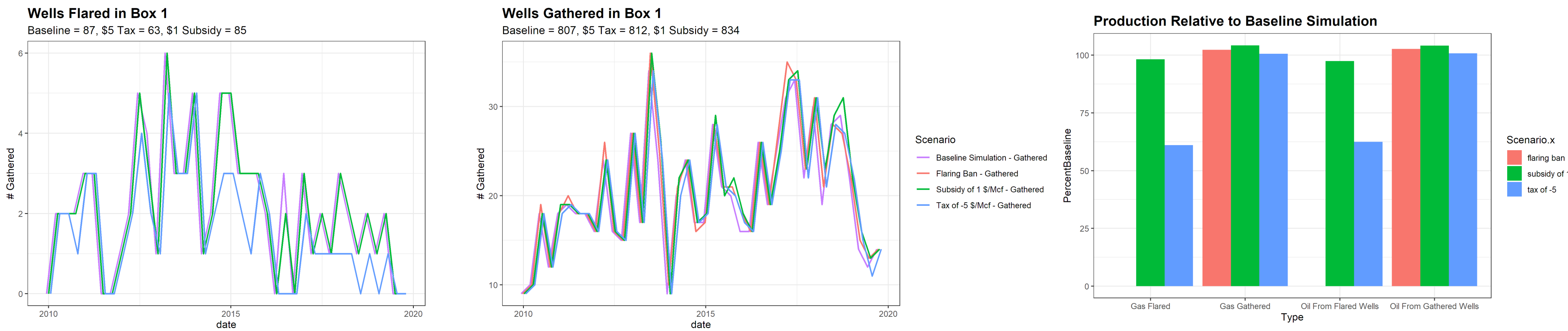


Figure 18: Outcomes of a Flaring Ban, Flaring Tax, Gas Subsidy

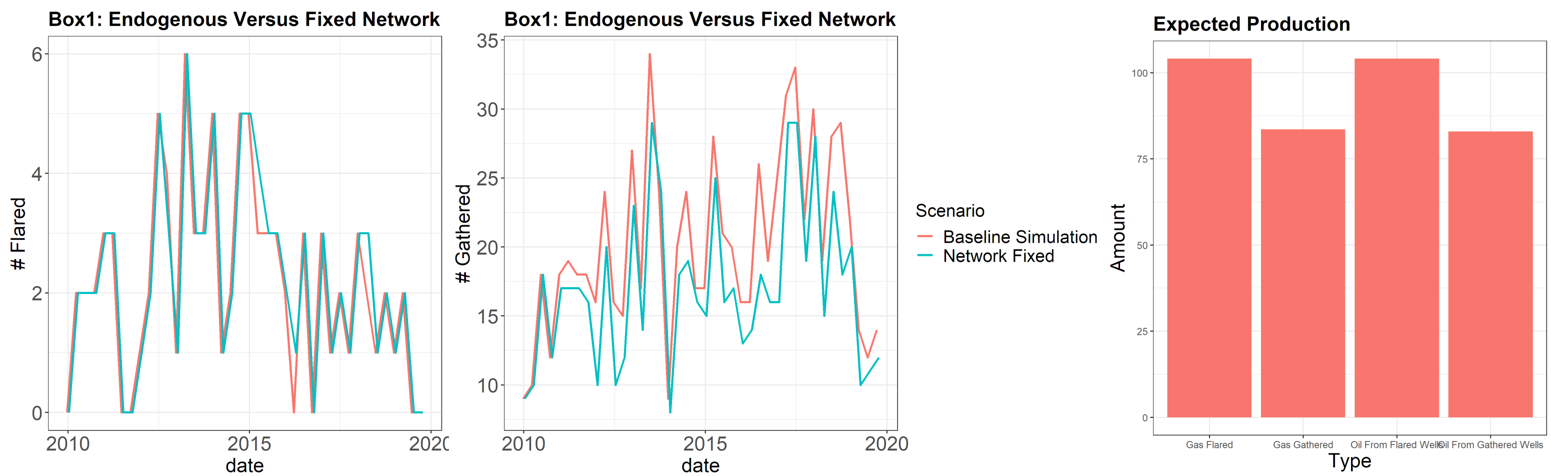


Figure 19: Outcomes With Network Fixed at t =2010-01-01 Versus Endogenous Network

1.9 Conclusion

I show that distance to pipeline infrastructure matters for flaring outcomes and that as the network grows over time, connections become less costly. My model indicates that flaring abatement is relatively costly, but that a tax near the likely social cost could substantially reduce flaring. A $\$5$ per Mcf tax would reduce flared volumes due to wells not connecting by 39%. Moreover, my results find that the majority of the decrease in flaring wells comes from flaring wells switching to not drilling rather than switching to connecting. In addition, I also find that subsidies poorly target flaring, and that bans only lead to a small increase in pipeline investment. My model estimates also suggest that spatial spillovers are not internalized. However, my counterfactual simulations suggest that the lack of spillover internalization does not have a substantial effect on flared volumes and that the lack of internalization might simply be due to the benefits of internalization being relatively small.

This paper makes a few important contributions to the literature. First, this paper is the first to explore the endogenous relationship between drilling and gas pipeline construction. Second, this paper is the first paper which models flaring *and* allows policies to affect the drilling margin. Third, this paper estimates pipeline costs directly instead of relying on industry reported costs. Combined, these contributions make this paper the most accurate model of the drilling and connecting decisions of firms.

These decisions have large consequences for U.S. greenhouse gas emissions, making this paper a timely contribution for policymakers interested in addressing climate change. The U.S. EIA estimates that in 2020 about 420 million Mcf of natural gas was vented or flared in the U.S. (U.S. Energy Information Administration, 2021b). Assuming a social cost of flaring between $\$3$

and $\$6$ implies that flaring accounts for billions of dollars in climate damages annually.

Chapter 2: Performance of an Input-Based Fee to Reduce Methane Emissions

2.1 Introduction

While most of the focus on climate change mitigation has been focused on reducing CO_2 output, methane emissions have received increasing scrutiny over the past few years. Methane is an extremely potent, though relatively short-lived greenhouse gas. Over a 20-year period, the global warming potential is a whopping 72-times that of CO_2 (U.S. Environmental Protection Agency, Table A-238). Therefore, methane emissions abatement can be an extremely powerful tool to reduce climate change – especially over shorter time-horizons.

One major source of methane emissions in the U.S. is from oil and gas production. The U.S. Environmental Protection Agency (EPA) estimates that oil and gas production emits over 125 million metric tons of CO_2e per year in methane. However, recent scientific papers have shown that these EPA estimates are likely far too low. For example, Alvarez et al. (2018) estimate that methane emissions from oil and gas production are 60% higher than EPA estimates while Zhang et al. (2020) estimate that emissions are over twice that of EPA estimates in the Permian Basin.

Methane emissions are a non-point source pollution problem where the regulator cannot directly observe site-level emissions. Rather, the EPA estimates emissions by combining site-level equipment inventories and self-report flared and vented gas volumes with estimated emissions factors. This type of estimate is known as a "bottom-up" estimate. In contrast, "top-down" estimates come from measuring methane air concentrations (either via aircraft, towers, or satellites) then inferring emissions. "Top-down" estimates (such as Zhang et al. (2020)), tend to estimate higher emissions than the EPA inventory. This likely happens because EPA's imputation method concentrates on emissions from normal operations and misses emissions from a small number of "super-emitters."

Methane emissions come from both the normal functioning of equipment like pumps and pneumatic devices and routine venting and flaring, but also from faulty equipment leaks. Therefore, emissions can be abated either through the installation of more efficient capital equipment or from leak detection and repair efforts. In this paper, I explore how a tax levied on the EPA's imputation would compare to a true (but technologically infeasible) Pigouvian tax by estimating the relative importance of each source of abatement (capital upgrades versus leak detection and repair). The central insight of the paper is that a tax on imputed methane will primarily cause changes in capital equipment and not affect firm decisions to find and repair leaks.

Natural gas is primarily composed of methane, so methane emissions from oil and gas production *are* leaked and vented natural gas.[1] This means that operators do have price incentives to abate emissions. However, these incentives will be far below the social cost of emissions.[2]

The economics literature on methane emissions from oil and gas production is relatively sparse and hasn't yet jointly investigated the relative importance of both sources of emissions abatement. Marks (2022) uses EPA estimated methane emissions from the Greenhouse Gas Reporting Program (GHGRP) and variation in natural gas prices to estimate the the methane

[1]Typically vented gas is referred to as intentional emissions while leaked gas refers to unintentional emissions.

[2]The social cost of methane in 2020 using a 3% discount rate is $1500 per metric ton (Interagency Working Group on Social Cost of Greenhouse Gases, 2021a). An Mcf of gas weights approximately 53 lbs (at 60°F and 14.7psia), which implies that the social cost of leaking an Mcf of gas is around $35 – far higher than the typical spot price at major trading hubs (usually between $2 and $6).

abatement cost curve from oil and gas production. However, since he's using EPA's imputed emissions, these estimates primarily capture the effect of prices on capital upgrades (e.g. buying more efficient pumps or pneumatic devices) and venting and flaring. Equally important are the unknown emissions from "super emitters" that the EPA GHGRP largely misses. On the other end, Dunkle Werner and Qiu (2020) estimate the methane emissions abatement curve of emissions from leaks by using data collected from plane fly-overs. This paper captures abatement from leaks but doesn't consider how prices affect capital investment decisions.

This paper aims to bridge these two papers by jointly considering both sources of emissions, and how the existence and relative importance of both channels of abatement affects policy efficacy. In particular, the aim of this paper is to understand how the type of fee proposed in the "Inflation Reduction Act of 2022" (henceforth IRA), signed into law by President Joe Biden in August of 2022, would compare to a Pigouvian tax. This bill included a provision to place a fee of $\$900$ per metric ton of methane which will increase to $1,500 after two years (Congressional Research Service, 2022). The fee will be calculated using EPA's GHGRP, so will suffer from the problem discussed earlier – that true emissions are imperfectly correlated imputed emissions.

The estimation strategy in this paper starts with the assumption that capital choices which affect emissions will be made at the time of drilling, and that leak abatement decisions will be made continuously over time. Since methane is a valuable commodity, this means that a well's current period emissions will be affected by the current period price, as well as the price at the time of drilling. I estimate the sensitivity of methane emissions from oil and gas wells to the current price and the price at the time of a well's completion (the date when a well comes into production) using output from an inverse atmospheric modeling tool which uses satellite data on methane concentrations to create a gridded inventory of emissions. I argue that current-period price sensitivity is primarily capturing leak abatement while the price sensitivity at the time of completion is capturing operator capital decisions.

Methane emissions from oil and gas production share many characteristics with more traditional non-point source pollution problems such as water pollution from excess fertilization or manure production. Like agricultural non-point sources, the pollution comes from many, diffuse sources and it is incredibly difficult or impossible for the regulator to directly observe individual-level pollution. The regulator can observe aggregate quantities of pollution, but cannot attribute it to individual sources (Xepapadeas, 2011). Since the regulator cannot impose a Pigouvian tax that directly taxes pollution, the economics literature has focused on other means to internalize the pollution externality. The two main instruments discussed in the literature are input-based schemes, introduced by, Griffin et al. (1982) and ambient schemes, introduced by Segerson (1988). In the former, when pollution is perfectly correlated with an observable production input, the regulator can calibrate a tax to efficiently internalize the pollution externality. In the latter, the regulator sets an ambient standard, then taxes (or subsidizes) all potential polluters if aggregate pollution is above (or below) the threshold.

A tax on methane emissions based on GHGRP estimates, like the one proposed in the IRA, would be considered an input-based scheme. However, since there isn't a perfect correlation between input choices and pollution, the policy will likely generate deadweight loss relative to a first-best policy. First, the tax will only incentivize capital upgrades rather than leak detection effort since the regulator doesn't directly observe detection effort and it doesn't affect the tax. Second, bias in the imputed emissions estimates will translate into biased levels of abatement investment. Thus, this paper also makes a novel contribution to the literature on non-point source pollution by analyzing the cost effectiveness of an input-based scheme relative to a Pigouvian tax in a new context: methane emissions from oil and gas.

I find that both prices at the time of completion and current period prices decrease methane emissions from wells – implying that firms use both capital upgrades and leak detection and repair to decrease emissions when gas prices increase. The effect of prices at the time of completion is larger than the effect of current period prices, however it is fairly imprecisely estimated, so I cannot rule out the effects being equal. In simple counterfactual simulations, my model estimates imply that an input-based scheme would lead to significantly less emissions reductions than an equivalent Pigouvian tax. In the base simulation, a $\$2$ per Mcf tax on imputed emissions would lead to a 4% reduction in emissions while a $\$2$ per Mcf Pigouvian tax would lead to a 12% reduction in emissions.

2.2 Data

2.2.1 Methane Emissions

The most significant barrier to economic research on methane emissions has been the difficulty observing methane emissions. In this paper, I use posterior emissions estimates generated by the Integrated Methane Inversion (IMI) tool. The IMI uses inverse atmospheric analysis to estimate implied emissions by using satellite observations on methane concentrations gathered by the TROPOMI satellite (Varon et al., 2022). This tool allows researchers to create gridded inventories of emissions at a $0.25° \times 0.3125°$ resolution. Satellite methane observations from TROPOMI become available starting in May of 2018.

I run the IMI monthly using the boundaries for the Permian Basin used by the EPA's GHGRP. I plot an example of the inversion output below in figure 2.1. The region of interest

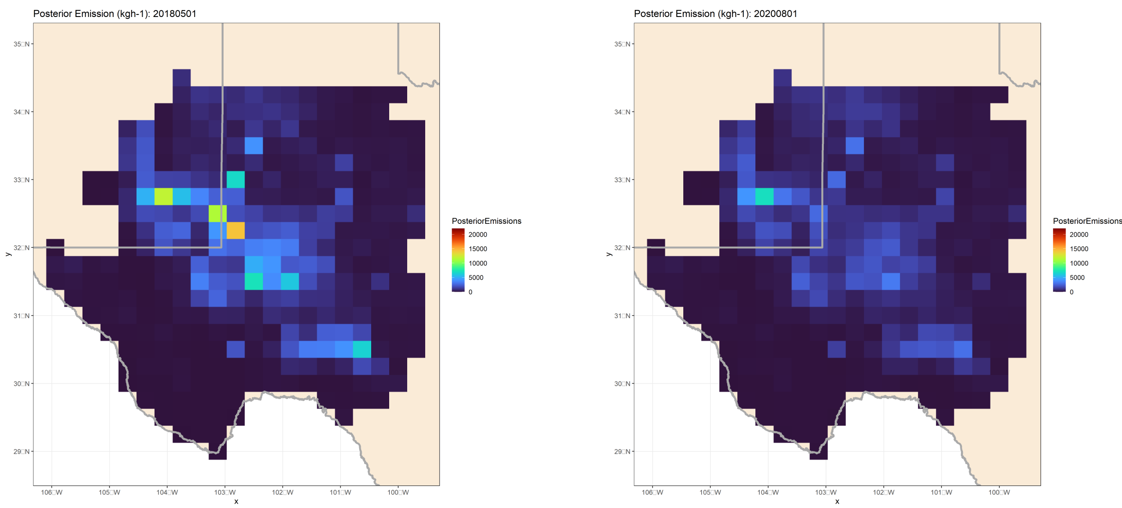


Figure 2.1: Examples of IMI output

contains 334 grid squares. Methane emissions appear to have large seasonal variations. Below in figure 2.2 I plot aggregate methane emissions by month of the year. As you can see, the summer months are associated with lower levels of emissions.

2.2.2 Production Data

I obtain data on well characteristics and well production from Enverus (formerly DrillingInfo). I observe well characteristics such as the completion date (when the well comes into production), the well type, and the monthly production from each well. There are 154,060 individual wells, and 7,642,185 well-month observations.

Summary statistics for the well-month data are found below in table 2.1, and summary statistics for the wells are found in table 2.2. In table 2.1, you can see that the production data has a strong negative skew (lots of very small producers and a few very large producers). In table 2.2, you can see that most wells are primarily oil wells.

I include the number of completions in each month as a control in my empirical estimation.

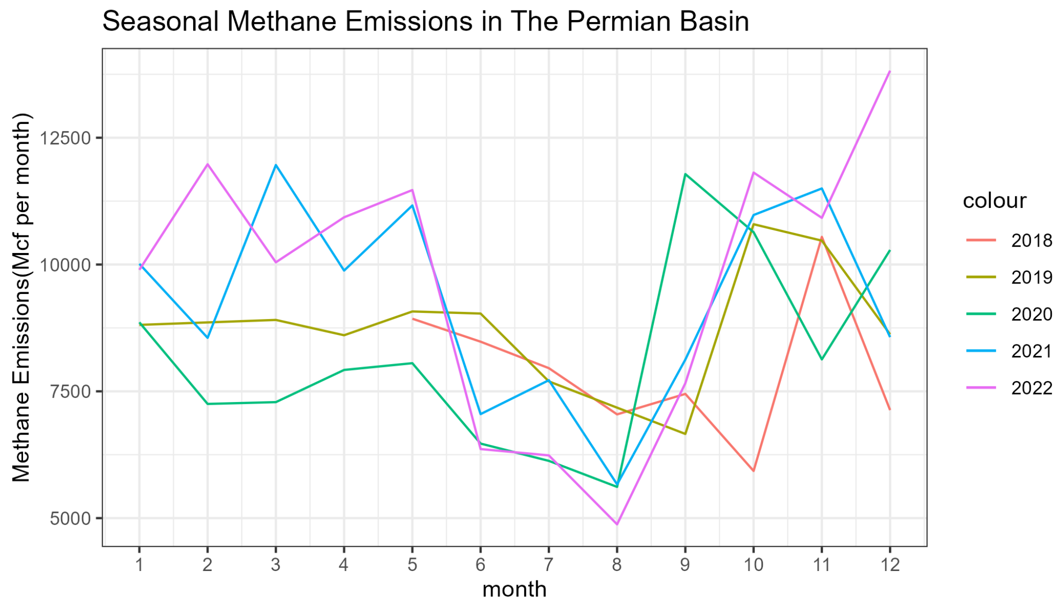


Figure 2.2: Methane Emissions by Month of the Year

Table 2.1: Production Summary Statistics

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
Oil (BBL)	7642185	1013	3589	0	11	330	229317
Gas (Mcf)	7642185	3783	13546	0	4	1142	1457574

Table 2.2: Monthly Summary Statistics

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
OilWell	154060	0.82	0.38	0	1	1	1
PriceatCompletion	97295	4.1	2	1.6	2.8	4.8	13
OilPriceCompletion	113495	61	30	11	32	87	134

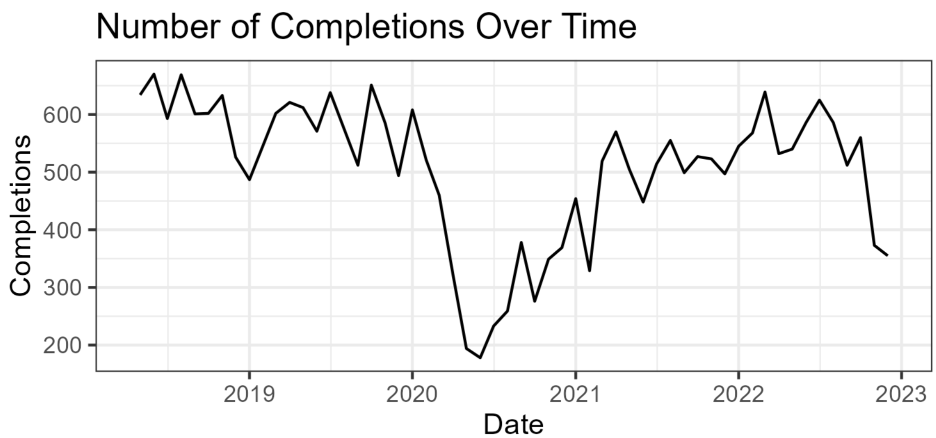


Figure 2.3: Number of Completions Over Time

Completions are major sources of both flaring and venting. In both Texas and New Mexico (the two states in the study region), flaring and venting are allowed for limited amounts of time after a completion. In Texas under Statewide Rule 32, operators are allowed to flare for 10 days after a completion, and vent any controlled release lasting less than 24 hours (U.S. Department of Energy, 2020). In New Mexico, operators are allowed to vent and flare gas for 60 days after a well completion (U.S. Department of Energy, 2022). I've plotted the number of completions in the study region in figure 2.3.

2.2.3 Prices

Finally, I gather natural gas and crude oil spot prices from the EIA. I use natural gas spot prices from trading at Henry Hub, and oil spot prices from the WTI trading hub in Cushing, OK. Unfortunately, these are prices at trading hubs, rather than the prices producers receive for the commodity at the wellhead. Therefore, there will be deviations from the wellhead price due to idiosyncratic price differences across space. In addition wellhead prices are typically lower because of transportation and processing costs.

2.3 Descriptive Regressions

I begin my empirical analysis by presenting some descriptive regressions to confirm expected correlations between methane emissions and production statistics and to directly compare some basic results to that of Marks (2022). I estimate:

$$M_{it} = \beta_1 C_{it} + \beta_2 p_t^g + \beta_3 p_t^o + \beta_4 O_{it} + \beta_5 G_{it} + \gamma_i + \eta_{m(t)} + \omega_{y(t)} + \epsilon_{it} \tag{2.1}$$

where M_{it} is methane emissions in grid i during month t (measured in Mcf per month), C_{it} are the number of completions, p_t^g is the natural gas price, p_t^o is the price of oil, G_{it} is aggregate gas production (measured in millions of mcfs), and O_{it} is aggregate oil production (measured in millions of barrels). I include grid-level fixed effects (γ_i) to control for time-invariant idiosyncratic differences across space as well as month-of-year and year fixed-effects ($\eta_{m(t)}$ and $\omega_{y(t)}$, respectively) to control for seasonality.

Results are presented in table 2.3. Column one presents the results from equation 2.1 and coefficients can be interpreted as the marginal effect on emissions in grid i in Mcf per month. Column two presents results when equation 2.1 is estimated using first-differences. Column three presents results when the dependent variable is logged, so coefficients can be interpreted as the percentage change in emissions from a marginal change in the independent variables. Column four presents results from a log-log specification, so results can be interpreted as elasticities. Finally, column 5 present results when the dependent variable is emissions rate, calculated as methane emissions divided by total gas production. The results from this specification can be directly compared to the results of Marks (2022).

Results are broadly in agreement across specifications. Completions seem to have a sizable effect on emissions. Both column (1) and (2) suggest that a completion releases approximately 200 mcfs of gas. Gas prices have a significant, negative effect on emissions. However, columns (3) and (4) suggest that emissions are relatively inelastic to changes in natural gas prices. Column (3) suggests that a $\$1$ increase in the price of natural gas is associated with a 4% decrease in emissions while column (4) suggests that the elasticity of emissions with respect to natural gas prices is -.23.

It is theoretically ambiguous whether using satellite data would lead to more or less elastic price responses than those estimated by Marks (2022). Since I am estimating shorter-run price responses, you might expect that these estimates would be less elastic. However, since the GHGRP only captures some of the mechanisms producers use to abate emissions, you also might expect that using satellite data would lead to more elastic estimates. Marks (2022) estimates that a $\$1$ increase in the price of natural gas leads to a 0.23 percentage point decrease in emissions. I estimate that price responses are more elastic and that a $\$1$ increase leads to a 0.8 percentage point decrease in emissions rates. This suggests that producers are actually more responsive to prices than GHGRP imputed emissions would imply. Alone, this is evidence that producers respond to prices in ways that do not show up in the GHGRP, and hence, that a tax on imputed emissions won't incentivize important sources of abatement.

Table 2.3: Descriptive Regressions of Grid-Level Methane Emissions

	Methane Emissions (1)	First-Differences (2)	log(Methane Emissions) (3)	log-log (4)	Emissions Rate (5)
Completed	194.8*	233.9**	0.001*	0.007	0.0005**
	(102.4)	(94.0)	(0.0008)	(0.006)	(0.0002)
Gas Price p^g	-2,738.4***	-1,725.8***	-0.040***	-0.235***	-0.008**
	(387.8)	(494.9)	(0.003)	(0.017)	(0.003)
Oil Price p^o	24.2	-167.2***	0.0006***	0.166***	-0.0007***
	(17.9)	(45.5)	(0.0002)	(0.019)	(0.0002)
Oil Production (mmBBL)	-2,198.4	5,471.3**	-0.011	-0.042**	-0.0008
	(2,266.2)	(2,263.4)	(0.018)	(0.021)	(0.006)
Gas Production (mmMcf)	810.1	-461.5	0.006	0.020*	-0.004**
	(562.8)	(493.3)	(0.004)	(0.011)	(0.002)
Standard-Errors	x-y	x & y		x-y	
Observations	12,222	11,996	12,208	4,471	9,112
R^2	0.860	0.047	0.979	0.975	0.811
Within R^2	0.010	0.006	0.021	0.040	0.004
x-y fixed effects	✓		✓	✓	✓
month(Date) fixed effects	✓	✓	✓	✓	✓
year(Date) fixed effects	✓	✓	✓	✓	✓

Standard-errors are clustered at the grid-level. Methane Emissions are measured in Mcf per month.

2.4 Distortions from Inefficient Policy

A social planner knows that oil and gas production is creating costly of methane emissions which have a marginal external cost of c^e. A firm holds a drilling prospect which will produce a fixed amount of oil o_i, and gas g_i. Emissions, e_i, are a decreasing function, $f(k_i, l_i)$, of capital (k_i) and leak detection effort (l_i). Moreover, emissions decrease the firm's volume of marketable gas, for which it receives market price p^g. The firm's problem is to maximize is profits:

$$\max_{k,l} p^o o_i + p^g(g_i - f(k_i, l_i)) - p^k k - p^l l \tag{2.2}$$

Clearly, if the social planner was able to observe emissions directly the optimal policy would be to place a tax of c^e on emissions, changing the firm's maximization problem to:

$$\max_{k,l} p^o o_i + p^g(g_i - f(k_i, l_i)) - p^k k - p^l l - c^e f(k_i, l_i) \tag{2.3}$$

with the following first-order conditions that equate the marginal cost of abatement with the marginal benefit (market value of increased gas plus the reduction in the tax):

$$\begin{aligned} k_i :& - p^g \frac{\partial f(k^*, l^*)}{\partial k} - p^k - c^e \frac{\partial f(k^*, l^*)}{\partial k} = 0 \\ l_i :& - p^g \frac{\partial f(k^*, l^*)}{\partial l} - p^l - c^e \frac{\partial f(k^*, l^*)}{\partial l} = 0 \end{aligned}$$

However, the planner cannot observe e, directly. Instead, the planner observes k, but cannot observe l and imputes $\tilde{e}_i = g(k_i)$, and places a tax of c^e on $\tilde{e}_i$. Then the firm's first-order conditions become:

$$
\begin{aligned}
k_i :& - p^g \frac{\partial f(k^*, l^*)}{\partial k} - p^k - c^e \frac{\partial g(k^*)}{\partial k} = 0 \\
l_i :& - p^g \frac{\partial f(k^*, l^*)}{\partial l} - p^l = 0
\end{aligned}
$$

Without putting structure on the shape of $f(\cdot)$ and $g(\cdot)$ it is impossible to sign whether the imputation leads to over or under-investment in k and l. However, economic intuition would suggest that the tax on imputed emissions would likely lead to under-investment in leak-detection. For example $\frac{\partial f}{\partial l \partial k} = 0$ would be enough to imply that leak detection would be lower in the policy with imputed emissions than in the optimal policy. This is a somewhat reasonable assumption that would simply imply that a firm's capital choices have no effect on the marginal benefit of leak detection effort. This would be true if, for example, two differently efficient pneumatic devices had the same potential leak sizes and probability of leaking.

2.5 Model and Estimation

In the main model to be estimated, a firm has made the decision to drill a well that will generate a know amount of oil and gas from period $t = 1$ to $t = T$. Gas production in each period is given by g_{it}. In period $t = 0$, the firm chooses a base level of emissions e_i^k (emissions from capital) that will persist over the entire life of the well. In all subsequent periods, the firm will observe prices and choose a secondary emissions rate (emissions from leakage), e^l.

First the firm chooses the base level of emissions to maximize the sum of expected profits over the life of the well:

$$\max_{e^k} E\Bigg[\sum_{t=1}^{T} \beta^t (p_t^g(g_t - e^k - e^l(p_t^g)) - c^l(e^l(p_t^g)))\Bigg] - c^k(e^k) \tag{2.4}$$

where $e^l(p_t^g)$ is the optimal amount of emissions from leakage given price p_t^g, $c^l(\cdot)$ and $c^k(\cdot)$ are the costs of achieving levels of leakage emissions and capital emissions, respectively, and β is the discount rate. The first-order condition is:

$$E\Bigg[\sum_{t=1}^{T} -\beta^t p_t^g\Bigg] = \frac{\partial c^k(e^{k*})}{\partial e^k}$$

This first-order condition simply states that the firm will choose a level of emissions from capital which equates their marginal cost of equipment with the expected discounted benefit over the life of the well of installing that equipment. I'm going to assume that firms form their expectation of future prices using only today's price (which would be consistent with the idea that gas price movements are determined by a first-order Markov processes). This will imply that emissions rates vary with the price of gas at the time of a well's completion.

Once the firm has chosen e^{k*}, next, the firm chooses in each period an additional leak size, e_t^l.[3] This leak size only affects one period. Again, they will select a leak size which equates their marginal benefit of abatement with the marginal cost by maximizing:

$$\max_{e_t^l} p_t^g(g_t - e^{k*} - e^l) - c^l(e^l) \tag{2.5}$$

[3]We can also think of the firm as choosing a leak probability. The empirical strategy will be summing over leaks from all wells, so in expectation, having fixed leak sizes, and choosing leak probabilities will work the same as directly choosing deterministic leak sizes.

with the associated first-order condition:

$$e^l : -p_t^g = \frac{\partial c^l(e^{l*})}{\partial e^l} \tag{2.6}$$

The very strong assumption in this model is that emissions from leak detection and repair, e^l, and emissions from capital choices, e^k are independent and separable.

2.5.1 Estimation

Recall that I'm using gridded data on emissions rather than observing well-level emissions directly. The empirical challenge is that since I'm observing aggregate emissions, I cannot simply regress well-level emissions on well-level characteristics. Instead, I parametrically specify well-level emissions and select parameters which match observed aggregate emissions with predicted aggregate emissions using maximum likelihood. I specify emissions from each well as:

$$e_{it} = \alpha_0 + \alpha_1 p_t^g + \alpha_2 p_{\tau(i)}^g + \alpha_3 p_t^o + \alpha_4 p_{\tau(i)}^o + \alpha_5 G_{it} + \alpha_6 O_{it} + \alpha_7 WellType_i \tag{2.7}$$

In this specification, per-well emissions are a function on the current price of oil and natural gas, the prices at the time of the well's completion, and well characteristics. Namely, whether the well is classified as an oil or gas well, and how much oil and gas the well produces at t. I include oil prices since it's plausible that changes in oil prices have an effect on the willingness of operators to find and repair leaks and install more efficient capital. For example, if an operator is capital constrained and oil prices are very high, then they might focus less effort on methane abatement.

Given emissions at each well, observed emissions in each grid will be given by the sum of e_{it} over all wells i in grid j, the number of completions multiplied by emissions per completion, time-invariant grid-level emissions from other sources, β_j, seasonal methane emissions from other sources, $\gamma_{m(t)}$, and a random error term, ϵ_{jt}:

$$E_{tj} = \beta_j + \beta_1 NComp_{tj} + \sum_{i \in J_t} e_{it} + \gamma_{m(t)} + \epsilon_{tj} \tag{2.8}$$

I assume the error term, ϵ_{jt} is i.i.d normal with mean 0 and variance $\sigma^2/2$.

To eliminate the need to estimate all β_j, I begin by taking differences of equation 2.8:

$$\Delta E_{tj} = \beta_1 \Delta NComp_{tj} + \sum_{i \in J_t} e_{it} - \sum_{i \in J_{t-1}} e_{it-1} + \Delta\gamma_{m(t)} + \Delta\epsilon_{tj} \tag{2.9}$$

Using this equation, I form a log-likelihood function to estimate the parameters of the model. Since ϵ_{jt} are assumed to be i.i.d. draws from $\mathcal{N}(0, \sigma^2/2)$, then $\Delta\epsilon_{jt} \sim \mathcal{N}(0, \sigma^2)$. Denote the entire set of parameters (excluding σ) $\boldsymbol{\beta}$. Then the likelihood function can be formulated as:

$$LL(\boldsymbol{\beta}, \sigma) = \frac{-1}{2\sigma^2} \sum_j \sum_t \sigma^2 \ln(2\pi) + \sigma^2 \ln(\sigma^2) + (\Delta E_{tj} - \Delta\hat{E}_{tj}(\boldsymbol{\beta}))^2 \tag{2.10}$$

The parameters to be estimated are $\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, \beta_1, \sigma$, and the set of monthly differences, which I estimate via maximum likelihood. For each interation of the likelihood function, I calculate expected emissions given the set of parameters for each well, then sum over wells for each grid.

2.6 Results

The two main parameters of interest are the parameter on the current price of natural gas, and the parameter on the price of natural gas at the time of completion. If emissions are more sensitive to prices at the time of completion than to current period prices, then it would suggest that firms primarily abate methane through capital upgrades. That would imply that a tax on imputed emissions would be fairly cost-effective relative to a first-best, but technologically infeasible, policy.

Results are presented in table 2.4. As expected both the coefficient on gas price, and the coefficient on gas price at time of completion are negative. The coefficient on gas price at completion is about six times as large as the coefficient on the current gas price. However the coefficient on gas price at time of completion is not significant at a 5% level and is far less precisely estimated than the current period gas price. As a result, the two coefficients have overlapping 95% confidence intervals.

Since some wells have missing values in the price at completion (typically means that the well was completed before the EIA spot price series started), I throw in dummies for these missing values since I cannot simply throw them out from estimation.

I also estimate a version of the model where instead of estimating single slopes on prices, I divvy up oil and gas prices into bins based on observed price quartiles, then estimate coefficients on the price bins. This allows me to examine the shape of the abatement curve over different prices. Results are found in table 3.3. The coefficients on the gas price and oil price bins are also plotted in figures 2.4 and 2.5, respectively.

There is broad agreement between the two version of the model. Like the simpler model

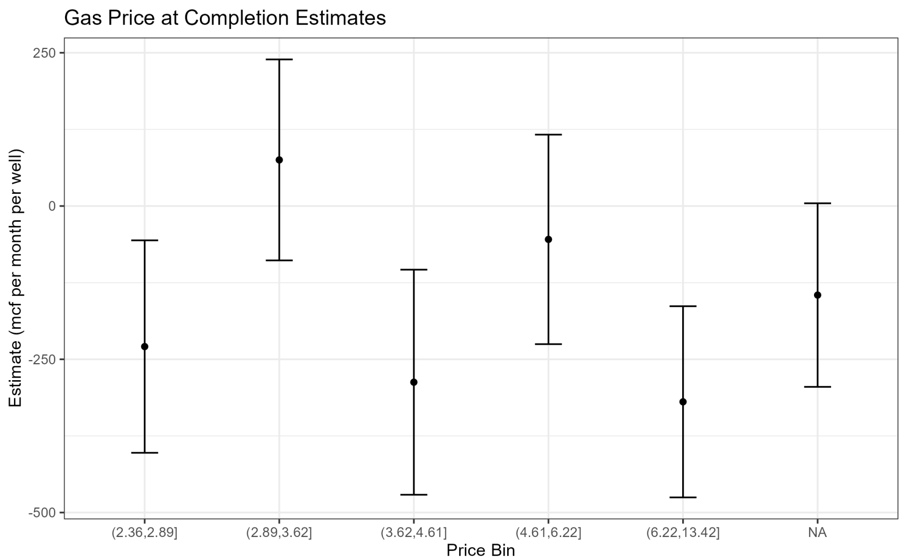
	name	Estimate	Std. Err	t-val	p-val
1	σ	26030.37	166.63	156.22	0.00
2	α_0	-410.27	116.04	-3.54	0.00
3	Completions	147.85	61.64	2.40	0.02
4	Gas	-8.12	6.91	-1.18	0.24
5	Oil	0.00	0.00	0.20	0.84
6	OilWell	-84.79	41.68	-2.03	0.04
7	GasPrice	-3.91	0.31	-12.54	0.00
8	OilPrice	-0.25	0.03	-7.29	0.00
9	GasPriceCompletion	-24.26	21.09	-1.15	0.25
10	GasCompletionNA	195.88	122.79	1.60	0.11
11	OilPriceCompletion	10.91	1.76	6.19	0.00
12	OilCompletionNA	368.44	107.06	3.44	0.00
13	month2	1055.85	1019.77	1.04	0.30
14	month3	106.80	873.37	0.12	0.90
15	month4	-769.89	871.95	-0.88	0.38
16	month5	4415.40	863.62	5.11	0.00
17	month6	-10069.90	770.47	-13.07	0.00
18	month7	-1723.52	778.75	-2.21	0.03
19	month8	-4414.98	864.26	-5.11	0.00
20	month9	9005.98	770.12	11.69	0.00
21	month10	6930.27	781.64	8.87	0.00
22	month11	1154.36	770.34	1.50	0.13
23	month12	-4063.05	780.82	-5.20	0.00

Table 2.4: Model Parameter Estimates (Slope)

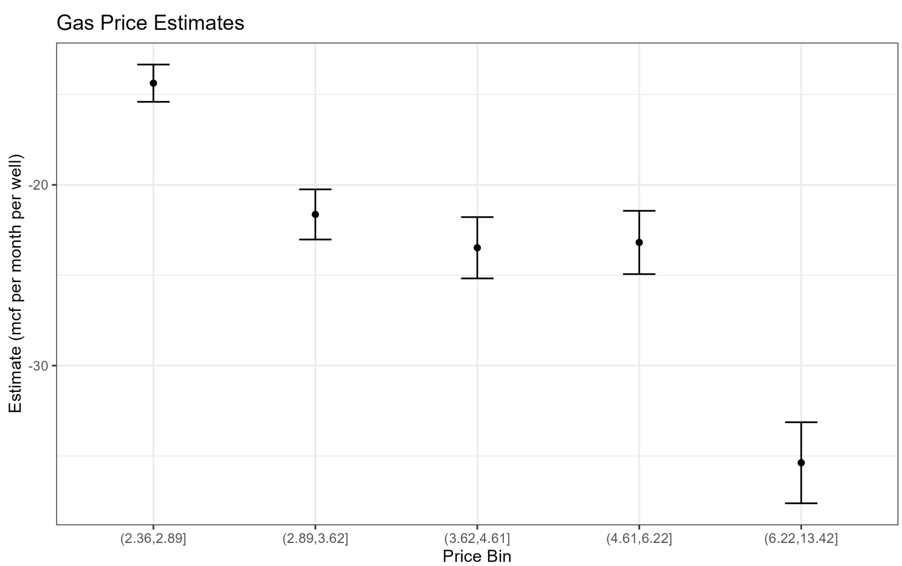
with just a slope on prices, the coefficients on price at time of completion are larger in magnitude but far less precise.

2.6.1 Policy Implications

The focus of this paper was to understand the efficiency of a hypothetical tax on imputed emissions. This exercise hinges on the assumptions that (1) capital choices are made and fixed at the time of drilling and (2) that the only way producers abate emissions of already-existing wells is by finding and fixing leaks. If, for example, I had data on capital choices at wells, I could test

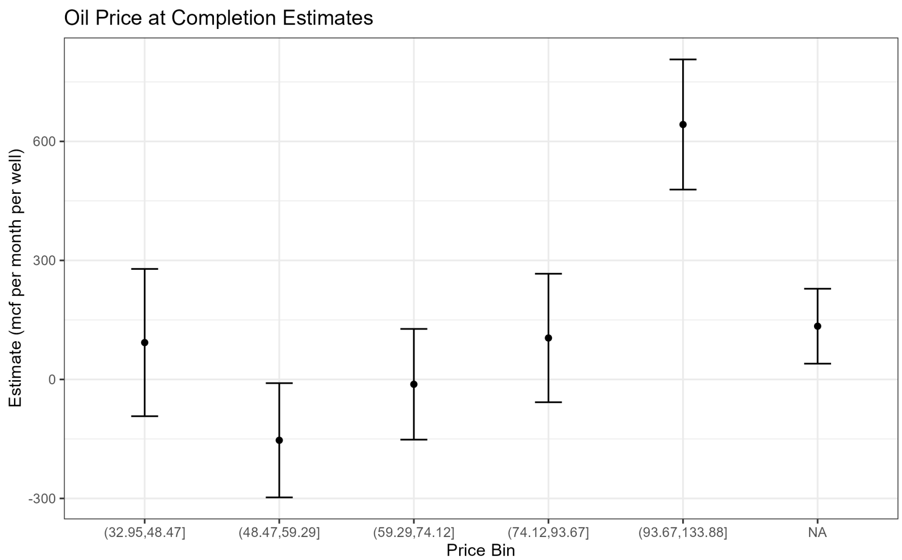


(a) Price at Completion

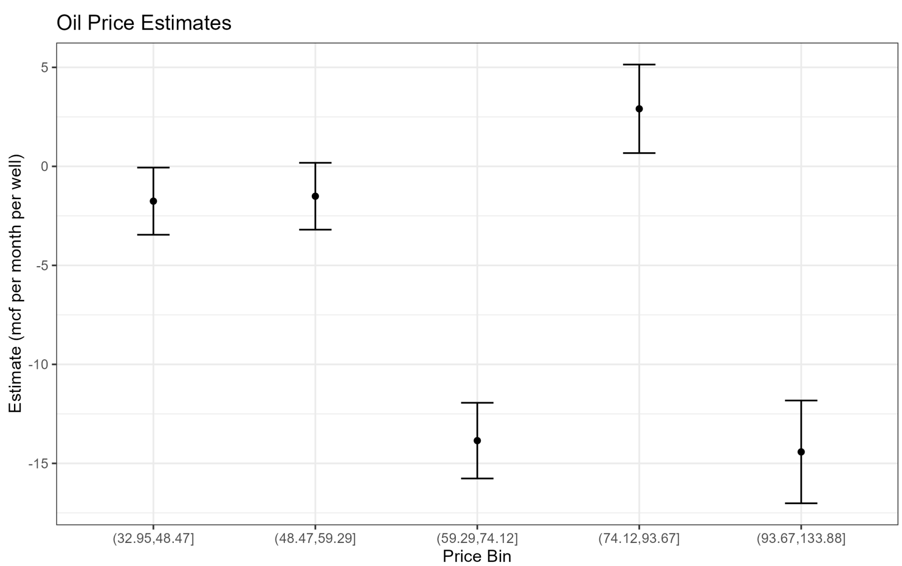


(b) Current Period Price

Figure 2.4: Gas Price Bin Coefficients



(a) Price at Completion



(b) Current Period Price

Figure 2.5: Oil Price Bin Coefficients

these assumptions and more directly decompose the strategies firms use for emissions abatement. Data on capital choices is available from the EPA's GHGRP, however it is provided at the year-firm-basin level, which is too coarse to be used in conjunction with data from TROPOMI (which didn't launch until mid-2018).

The point estimates on the current-period price of natural gas and the price at time of completion were -3.9 and -24.3, respectively. Suppose a fee of $\$1$ per Mcf was placed on imputed emissions. These point estimates would imply that that an imputed would reduce emissions for all new wells by 24.3 mcfs per month. A Pigouvian tax on the other hand would reduce emissions at new wells by 28.2 mcfs per month, and at all existing wells by 3.9 mcfs per month. This implies that, while an imputed tax would lead to substantial reductions in emissions, an infeasible first-best policy would lead to significantly more reductions.

To get an estimate of how these per-well changes would translate into differences in aggregate emissions I run a simple counterfactual simulation. The exercise is to look at predicted methane emissions under two different scenarios. In the first scenario, emissions at wells drilled on or after May 2018 face varying prices at the "time of completion" between $\$2$ and $\$6$ while the current period price is set to a constant $\$2$. This is meant to mimic an input-based scheme where producers are incentivized to make capital upgrades at the time of well completion, but not afterwards for finding leaks. In the second scenario I vary both the price at time of completion (for all wells drilled after May 2018) as well as the current price for all wells. I predict methane emissions using point estimates from table 2.4. I've plotted the outcomes of these counterfactuals. in figure 2.6.

If the market price of gas is set at $\$2$ and we think about comparing an input-based fee of $\$2$ to a Pigouvian tax of $\$2$, this chart suggests that the input-based tax will lead to a 4% drop in

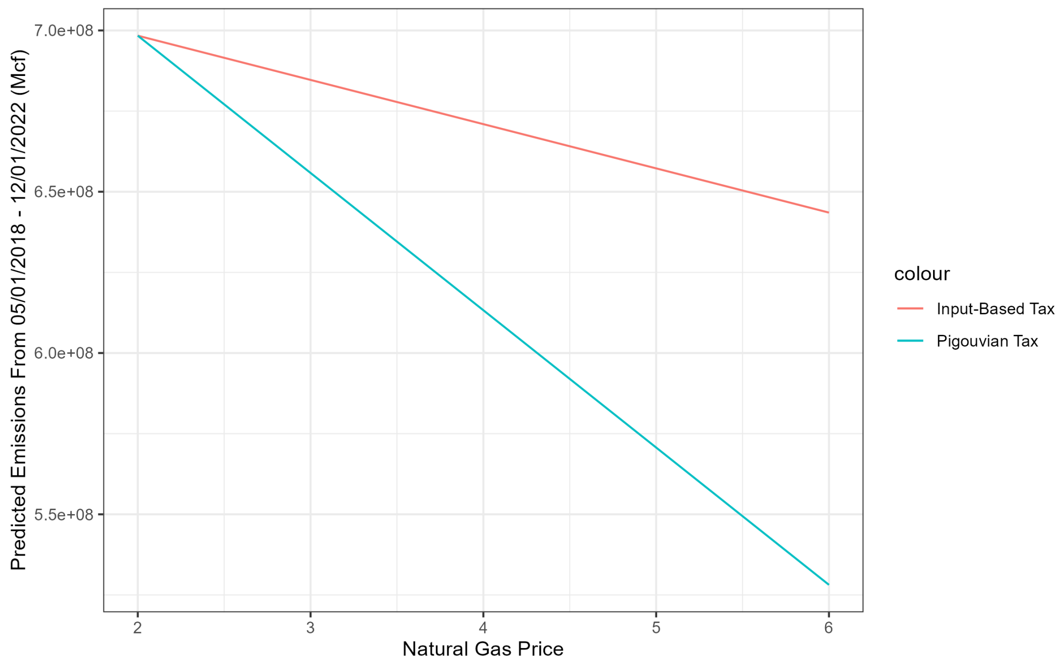


Figure 2.6: Counterfactual Emissions Under Different Scenarios

overall emissions ($7 \times 10^8 \rightarrow 6.7 \times 10^8$ Mcf) while a Pigouvian tax would lead to a 12% drop ($7 \times 10^8 \rightarrow 6.1 \times 10^8$ Mcf) in emissions.

Of course, the assumptions needed to come up with these numbers are quite strong, so further research should be conducted on this topic. Once a longer panel of satellite data becomes available, it may be possible to combine insights from the GHGRP and TROPOMI, and directly test how firms use a mixture of both leak detection and capital upgrades to abate emissions.

2.7 Conclusion

This paper estimates the sensitivity of methane emissions from oil and gas production to natural gas prices at the time of a well's completion and the current period price. I argue that at the time of a well's completion, operators respond to higher natural gas prices by installing more efficient capital, and that during the current period, operators respond to higher prices by searching for leaks. The relative magnitude of these responses informs whether an input-based scheme which levies taxes based on imputed emissions is cost effective relative to an infeasible first-best Pigouvian tax. My point estimates suggest that methane emissions are more sensitive to prices at the time of completion than current prices. However, the slope estimate was imprecisely estimated so I cannot reject the null that the slopes are equal, nor can I reject the null that that emissions are unsensitive to the prices at the time of completion. The point estimates imply that an input-based scheme to reduce emissions could achieve major reductions, but might perform substantially worse than an infeasible first-best policy.

This paper makes an important contribution to the economics literature on methane emissions from oil and gas by considering how responsive firms are to price changes on numerous margins by using satellite data. Previous papers have focused either on the leak detection and repair, or capital investment margins. This paper also makes a novel contribution to the literature on non-point source pollution problems by estimating the performance of an input-based fee relative to an infeasible first-best Pigouvian tax.

Name	Estimate	Std. Err	t-val	p-val
σ	26098.00	171.76	151.95	0.00
α_0	116.85	130.21	0.90	0.37
Completions	175.01	62.13	2.82	0.00
Gas	0.43	7.24	0.06	0.95
Oil	0.01	0.00	1.98	0.05
OilWell	-37.29	42.44	-0.88	0.38
CompletionPriceBin (2.36,2.89]	-229.27	173.29	-1.32	0.19
CompletionPriceBin (2.89,3.62]	75.22	163.91	0.46	0.65
CompletionPriceBin (3.62,4.61]	-287.39	183.57	-1.57	0.12
CompletionPriceBin (4.61,6.22]	-54.42	170.87	-0.32	0.75
CompletionPriceBin (6.22,13.42]	-319.37	155.94	-2.05	0.04
CompletionPriceBin NA	-145.27	149.72	-0.97	0.33
Gas Price Bin (2.36,2.89]	-14.37	1.03	-13.94	0.00
Gas Price Bin (2.89,3.62]	-21.63	1.39	-15.59	0.00
Gas Price Bin (3.62,4.61]	-23.48	1.70	-13.84	0.00
Gas Price Bin (4.61,6.22]	-23.18	1.75	-13.25	0.00
Gas Price Bin (6.22,13.42]	-35.37	2.24	-15.78	0.00
Oil Price Completion (32.95,48.47]	92.91	185.62	0.50	0.62
Oil Price Completion (48.47,59.29]	-153.39	144.01	-1.07	0.29
Oil Price Completion (59.29,74.12]	-12.23	139.53	-0.09	0.93
Oil Price Completion (74.12,93.67]	104.47	161.96	0.65	0.52
Oil Price Completion (93.67,133.88]	642.61	164.04	3.92	0.00
Oil Price Completion NA	134.19	94.41	1.42	0.16
Oil Price (32.95,48.47]	-1.76	1.70	-1.04	0.30
Oil Price (48.47,59.29]	-1.51	1.69	-0.89	0.37
Oil Price (59.29,74.12]	-13.85	1.91	-7.25	0.00
Oil Price (74.12,93.67]	2.91	2.24	1.30	0.19
Oil Price (93.67,133.88]	-14.42	2.60	-5.56	0.00
month2	874.49	1079.26	0.81	0.42
month3	4017.17	954.73	4.21	0.00
month4	1679.26	913.81	1.84	0.07
month5	3898.29	915.81	4.26	0.00
month6	-10973.77	836.02	-13.13	0.00
month7	-2167.23	812.26	-2.67	0.01
month8	-4497.73	899.03	-5.00	0.00
month9	8354.80	824.13	10.14	0.00
month10	4696.05	811.15	5.79	0.00
month11	2331.93	801.06	2.91	0.00
month12	-1207.66	832.20	-1.45	0.15

Table 2.5: Model Parameter Estimates (Bins)

Chapter 3: Policy Options to Reduce State Liabilities from Orphaned Gas Wells

3.1 Introduction

Oil and gas wells pose serious risks to groundwater, air, and safety. These risks are present even after the productive life of the well is over, until the well-site is remediated. This usually involves plugging the well with mud or cement to ensure that associated hydrocarbons or waste cannot seep into groundwater or leak into the air. Sometimes, wells remain inactive for years before the operator chooses to plug the well. If the operator goes bankrupt before the well is plugged, then the well becomes known as an orphan well. These wells continue to risk serious environmental damage even though they are not producing. Moreover, the state or federal government often foots the bill to plug these wells.

The U.S. has a long history of oil and gas production, and orphaned wells have accumulated over time to become a staggering problem. The U.S. EPA estimates that in 2021, there were 2.2 million abandoned oil and gas wells in the country (U.S. Environmental Protection Agency, 2022). Previous work in both the scientific and economics literature has demonstrated that inactive and abandoned wells carry the potential for serious environmental damage. For example, Fischer, Lebel, and Jackson (2020) use ground-based instruments to estimate methane emissions from wells in California and find that 11 of the 17 idle wells surveyed were leaking methane. Kang et al. (2016) measures methane emissions from 88 abandoned wells in Pennsylvania and estimates that leaks from abandoned wells account for a staggering $5-8\%$ of the states anthropogenic greenhouse gas emissions. In addition, these wells also can pose serious safety hazards – for example, in 2008 gas from an abandoned well leaked into a septic system and exploded, killing one person (Kusnetz, 2011).

In this paper, I estimate a simple model of well status, then use the estimated parameters of my models to construct counterfactual simulations of policies aimed at reducing well orphanage. There are a few previous papers in the economics literature focusing on well orphanage. Boomhower (2019) explores the judgement-proof aspect of environmental remediation and exploits a change in bonding requirements in Texas in 2003. He shows that the introduction of bonding requirements led small, environmentally risky firms to exit the market, leading to a decrease in orphan well rates and water protection rule violations. Muehlenbachs (2015) uses a dynamic discrete choice model to show that firms avoid the high fixed-costs of well plugging by leaving wells inactive, even if they have no intention of bringing them back into production in the future. Agerton et al. (2023) estimate that the cost of plugging all non-producing, unplugged wells in the U.S. Gulf of Mexico amounts to \$30 billion. Ho et al. (2018) collect data on state plugging expenditures in 13 states and find that state expenditures far exceed bond amounts.

This paper combines some of the various strengths of previous papers to explore the problem of orphan gas wells in Texas. For example, Muehlenbachs (2015) has data on well plugging but not data on firm survival. Therefore, her paper focuses on showing that producers delay plugging, rather than focusing on whether wells become orphaned. Boomhower (2019) uses data on firm survival but not plugging. Therefore, his analysis shows that the bonding rule reallocates production to larger firms and decreases well orphanage, but doesn't show whether bonding increases plugging. In contrast, since I have data on both firm survival and plugging, I estimate a well-level structural model of well status while focusing on the true outcome of interest, well-orphanage.

I focus my policy counterfactual simulations on a simple policy requiring firms to either bring wells back into production or plug them after they have been inactive for at least two years. My simulations suggest that this policy would decrease the number of orphaned wells by 39% and increase the number of plugged wells by 126%. While these results are likely over-estimates of the potential policy effectiveness due to strong assumptions about policy compliance, they demonstrate that even simple policies may have a substantial effect on orphanage rates without causing a large decrease in production.

3.2 Background

In the state of Texas, there are a couple of rules aimed at ensuring producers remediate their well-sites. First, under the Texas Natural Resource Code Chapter 91, firms are required to post bonds which can be seized in the event of environmental damage or if the firm leaves a well orphaned (91 Tex. Natural Resources Code § 104).[1] Firms have the option to either post a bond equalling $\$2$ per foot of total combined well depth, or they have the option of posting blanket bonds. If the firm has 10 or less wells, they may post $\$25,000$. If they have between 11 and 99 they may post $\$50,000$, and if they have 100 ore more they may post $\$250,000$.

In addition, House Bill 2259 which adopted Statewide Rule §3.15 was passed in 2009. This rule requires operators to address their inactive wells to renew their operational status with the state (McFarland, 2016). Operators are required to plug wells that have been inactive for more than 10 years. For wells that have been inactive more than one year, they have to file a plugging

[1]This is the regulation introduced by Senate Bill 310, which is studied by Boomhower (2019).

extension application. However, the bill had allowed very low producing wells to be considered "active". When the bill was adopted, the threshold for gas wells was 100 Mcf per month for three months. In 2017, this threshold was lowered to 50 Mcf for three months, or 1 Mcf per month for a year. To put this in context, the April 2023 Henry Hub spot price of natural gas was $2.16 per Mcf, so this meant that given those prices a well could produce as little as $25 worth of gas per year to be considered "active" by the state. McFarland (2016) notes that "it is much more practical and financially feasible for an operator to falsify its reporting of inactive wells" since RRC enforcement and accountability of this rule are low.

There is reason to worry that the current pool of shut-in wells poses major risks for the state. In figure 3.1 I plot the number of wells of different statuses over time. The proportion of shut-in wells has been rising substantially for the past decade. Figure 3.2 shows that this increase in well shut-in rates has been driven both by decreased shut-in to plugged transitions (meaning that wells are kept shut-in for longer), decreasing active to plugged transitions, and increasing active to shut-in transitions (meaning active wells are more likely to be shut-in rather than plugged).

While Boomhower (2019) finds evidence that the introduction of well bonding requirements led to a significant decrease in orphaned wells, orphan wells have continued to be a pernicious problem in the state and the increasing inventory of shut-in wells should worry policymakers. In figure3.3 I plot when orphan wells were shut-in. This figure illustrates one of the problems with having a large inventory of shut-in wells – shocks that affect the industry can translate into large jumps in state liabilities. In figure 3.3 there is a large jump in late 2019. You'll notice in figure 3.1, that this date is not associated with a large jump in shut-ins. Instead, a perfectly normal number of wells were shut-in in late 2019, but a wave a bankruptcies from crashing oil and gas prices exactly one year later during the COVID-19 pandemic caused many of these shut-in wells

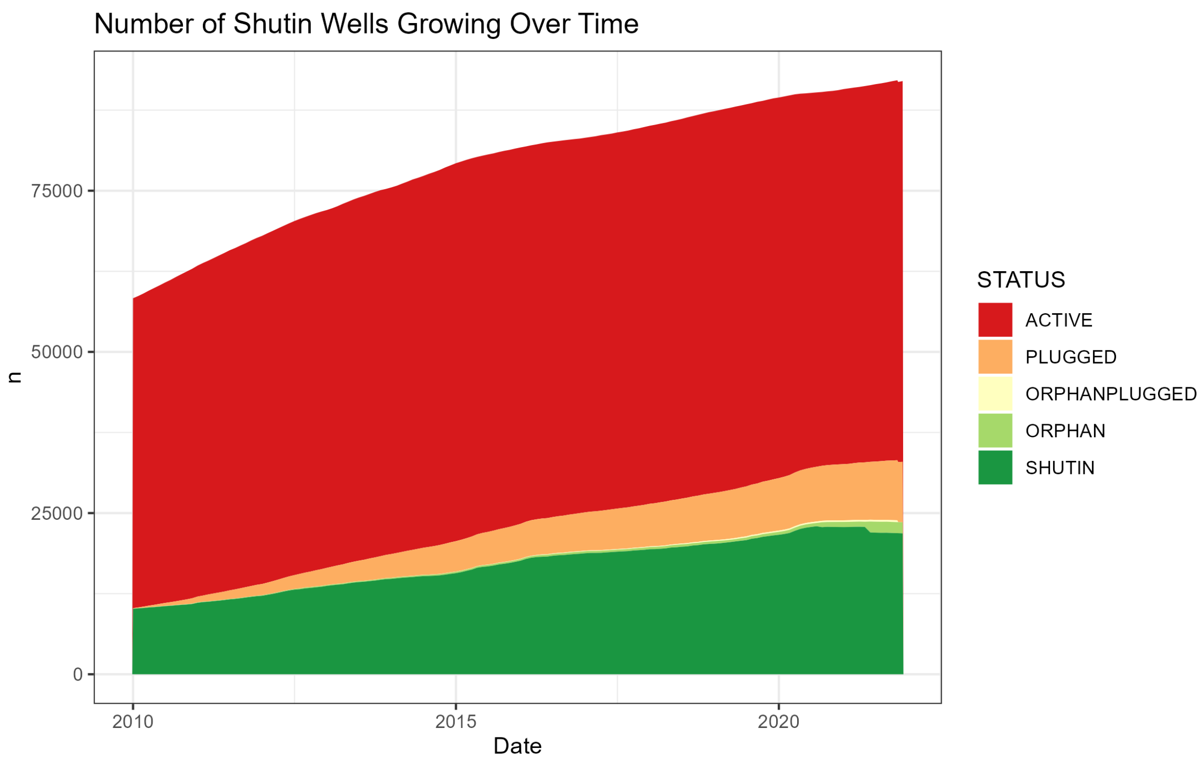
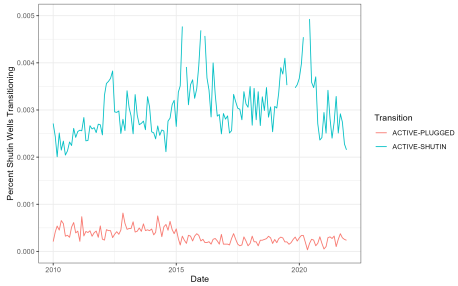
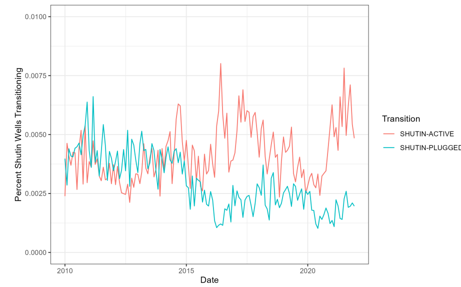


Figure 3.1: Area plot of wells by status. The number of shut-in wells has been steadily increasing since 2010.



(a) Percentage of wells transitioning from active. Active to plugged transitions have gone down while active to shut-in transitions have gone up.



(b) The percentage of shut-in wells becoming plugged has also gone down.

Figure 3.2: Status transition over time. The rise in the number of shut-in wells has been driven by decreasing plugging rates, as well as increasing shut-in rates.

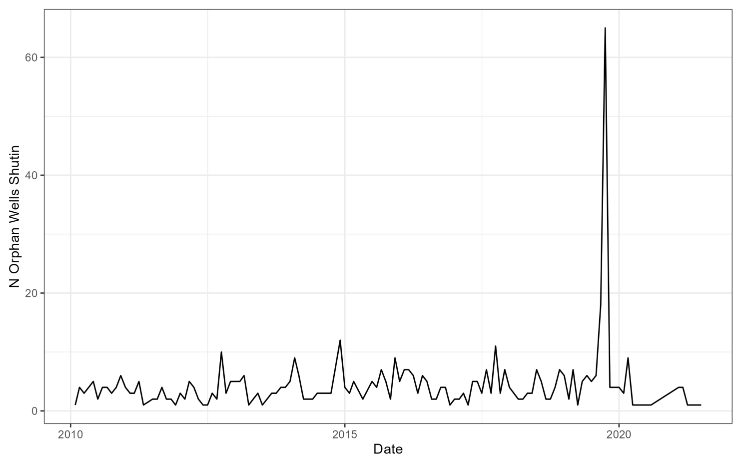
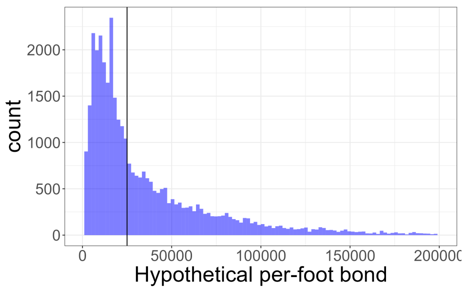


Figure 3.3: When Orphan Wells Shut-in

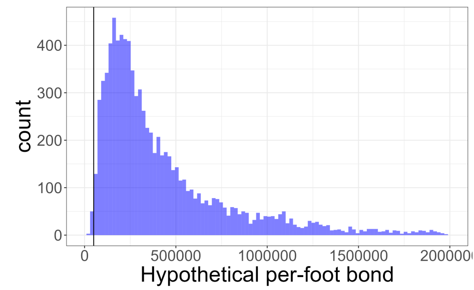
to become orphaned.

Moreover, while bonds are a very effective policy they are currently incredibly low because of producer's ability to cover their entire set of wells with a blanket bond. The insufficiency of blanket bonds is not unique to Texas and has been noted by previous papers such as Raimi et al. (2021). In Texas, it's cheaper for the vast majority of firms to bond their wells with a blanket bond. In figure 3.4 I plot the hypothetical per-foot bonds firms would need to post if blanket bonds were not an option. The blanket bonds are lower than these hypothetical per-foot bonds for the vast majority of firms. In turn, the per-foot bond is far lower than the estimated costs of plugging these wells. This means that the state cannot cover the costs of plugging orphan wells with the posted bonds. The situation looks even more bleak given that the potential environmental damages from orphan wells could far exceed their plugging costs. Boomhower (2019) makes a back-of-the-envelope calculation suggesting that these expected costs are $\$119,000$ per orphan well.

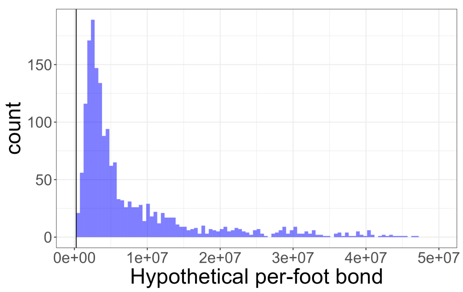
In general, economic theory suggests that optimal bonds will be lower than maximum potential damages since bonding can impose significant costs on producers (Davis (2015), Gerard and Wilson (2009)). However, the bond amounts required by Texas are on average many orders of magnitude lower than the expected damages conditional on needing to seize the bond. Moreover, in Texas bonds are intended to be used either in the event of a well becoming orphaned *or* in the event of some environmental damage (like an oil spill or groundwater contamination event). Well orphanage and potential environmental damage from wells are two separate problems with very different characteristics. For example, all wells need to eventually be plugged, but only a very small proportion are involved in a spill or groundwater contamination event. Moreover, the costs involved with an environmental damage event are much higher than the costs of plugging a single well. It follows that the optimal set of policies addressing these issues would likely target them separately. In this paper, I focus on the issue of well plugging and well orphanage while recognizing that any policy affecting plugging rates will likely have knock-off effects on other environmental damage incidents.



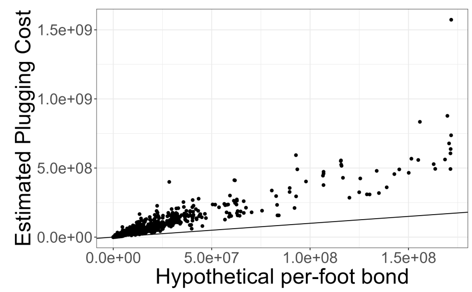
(a) Firms Covered by $25,000 Blanket Bond



(b) Firms Covered by $50,000 Blanket Bond



(c) Firms Covered by $250,000 Blanket Bond



(d) Plot of Hypothetical Per-Foot Bond Versus Estimated Plugging Cost

Figure 3.4: Plot of hypothetical per-foot bonds and blanket bonds. Panels (a), (b), and (c) plot the hypothetical bond amounts firms would be required to post if they used the per-foot calculation rather than a blanket bond. In these panels, a black line is drawn at the blanket bond amount. Blanket bond amounts tend to be far less than the hypothetical per-foot bond. Panel (d) plots the hypothetical per-foot bond against the estimate cost of well plugging by firm. It also plots a line along $y = x$. The hypothetical per-foot bond is far less than the estimated plugging costs.

3.3 Data

I use data obtained from the Texas Railroad Commission (RRC). To construct the main panel, I combine well-level monthly production with well-level characteristics such as the history of well operators, the well depth, the well completion date, and the well's plugging date. I also obtain information on the first and last date at which each producer filed its annual licence renewal with the RRC. The combination of well-level production data, firm status data, and plugging dates allows me to classify each well as either Active, Shut-in, Plugged, or Orphaned. I classify a well as shut-in if it doesn't produce gas for 12 months. The RRC considers wells orphaned once a

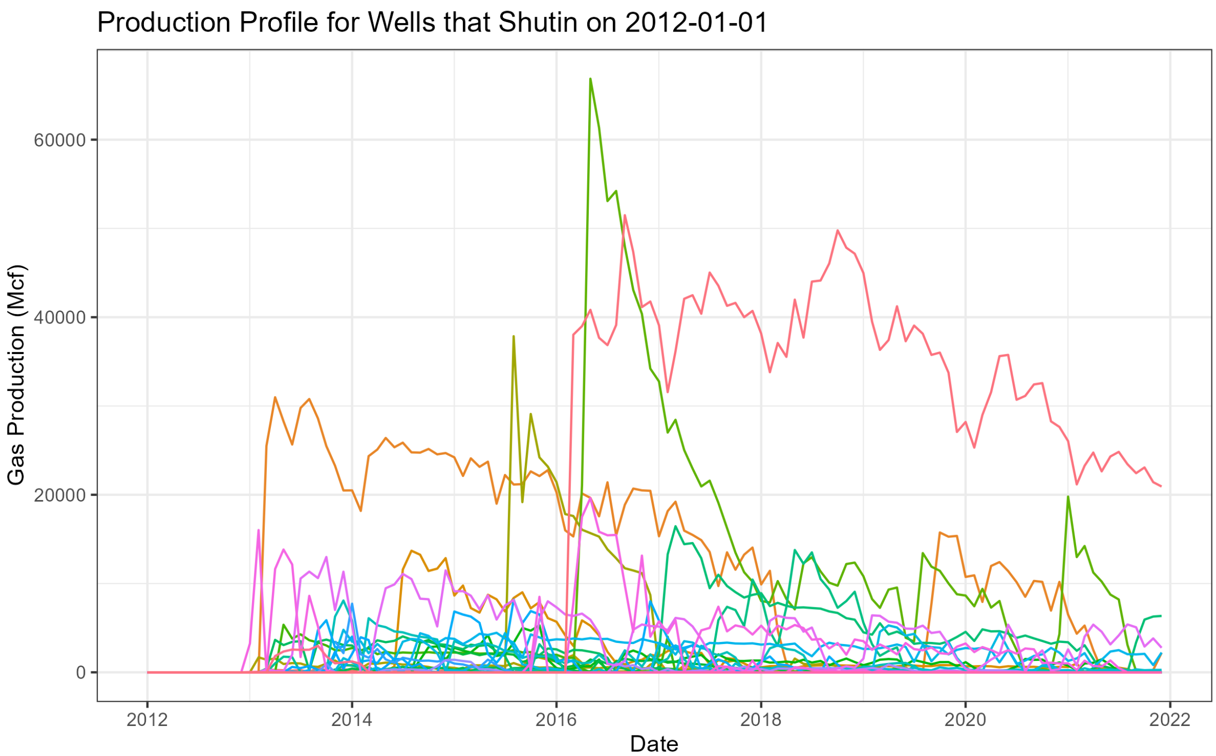


Figure 3.5: Production Profile for Wells that shut-in on 2012-01-01, and came back into production later. Some wells stay shut-in for years then go on to produce substantial amounts of gas, while others never produce substantial amounts again.

well has been inactive for at least twelve months and once the operator of that well has been in delinquent status for twelve months (Biven and Palacios, 2022), so I use the same definition they do.

Importantly, a well can be shut-in for substantial periods of time, then go on to produce large amounts of gas later. This means that shutting-in a well can give the operator option value that it would lose from plugging a well. In my sample, I find that of wells that become shut-in at some point during the period of study, 83% remain shut-in for the entire study period, 10% are brought back into production, 6% get plugged, and 1% become orphaned.[2] Alone, these numbers imply that shutting-in wells *does* have option value, but also that producers may be using shut-ins to avoid plugging costs. Moreover, shut-in wells that come back into production tend to do so relatively quickly. In figure 3.6 I plot the length of time that a well remains shut-in for before a transition. You can see that most shut-in wells that come back into production do so within two years. This is why I use a hypothetical two-year limit in my counterfactual simulations.

The producer data I use only gives me information on the first and last date at which a producer filed their annual license renewal. Importantly, a producer "exiting" does not necessarily mean that they left orphan wells. The operator might have sold all of its assets or been acquired by another firm. To infer whether a producer went bankrupt and could not cover the costs of site remediation, I observe whether the firm left orphan wells. I use a combination of the "February 2023 Orphan Well List" downloaded from the RRC, and data on which wells were plugged with state funds after 2010 obtained via a freedom of information act request.[3]

[2]These number match up quite closely with those found by Muehlenbachs (2015), despite the vastly different study settings. Muehlenbachs (2015) studies a panel of conventional wells in Alberta drilled throughout the 20th and into the 21st century. These wells are much older and typically have much longer lifespans than the wells I study.

[3]The states orphan well list only contains wells currently on the orphan well register. So if, for example, the state pays to plug the well, it is removed from the list. This is why I need both current and historically plugged wells to infer whether a shut-in well became orphaned.

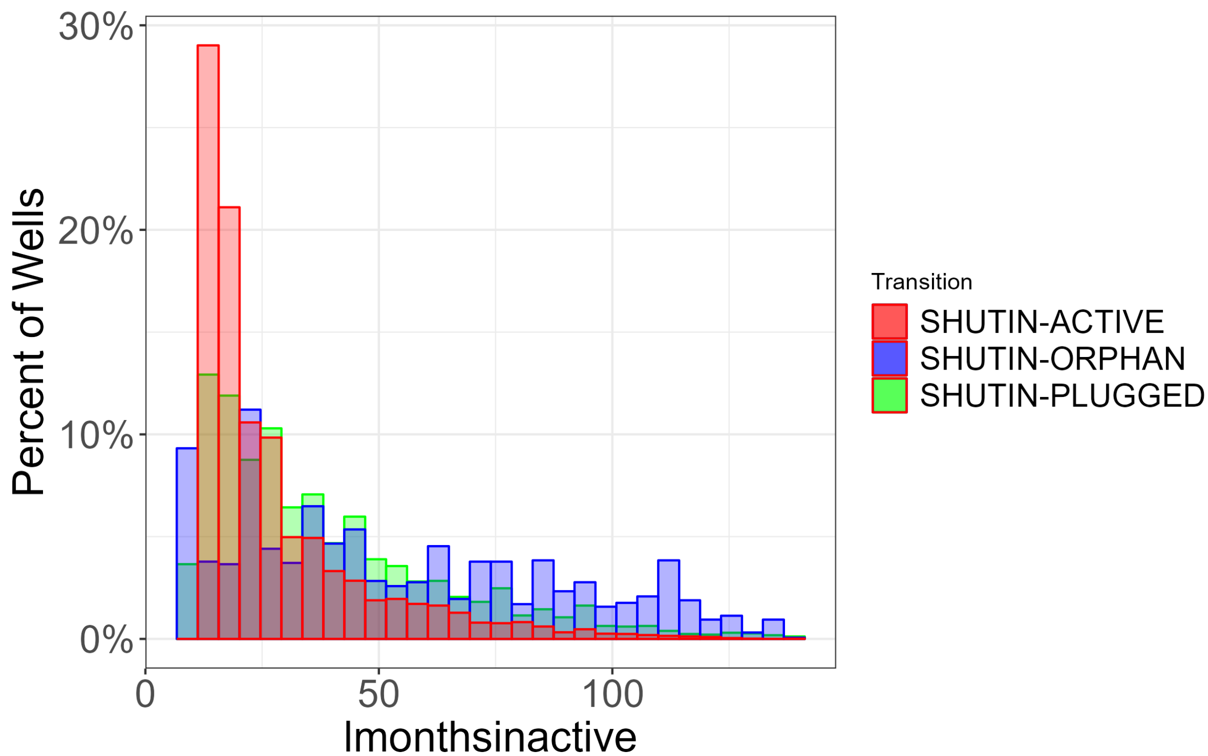


Figure 3.6: Histogram of transitions from shut-in by length of time inactive. Shut-in wells that transition back to active seem to do so much more quickly than wells that go on to be plugged or orphaned.

To calculate plugging costs I multiply well depth by a per-foot costs reported by the RRC [4] Finally, I gather data on monthly spot prices of natural gas from the U.S. Energy Information Administration.

I restrict my sample to gas wells. The Texas RRC reports monthly production at the lease-level. For oil wells, leases can be comprised of multiple wells while gas leases are restricted to one well. Therefore, well-level production is directly observable for gas wells but not for oil wells. I also restrict my sample to wells drilled on or after 2000 since the data on newer wells appears more reliable. Lastly, I restrict the data used in my analysis from January 2010 (the first year plugging costs are available) through December 2021.

I present summary statistics for the panel of wells below in figure 3.1. I also present a table summarising the number of status transitions observed below in figure 3.2. As you can see, conditional on being any given state, a well is most likely to simply remain in the same state in the next period.[5]

	N	mean	median	min	max
Prices (monthly)	144.00	3.26	2.99	1.63	6.00
Plug Costs (District-year)	156.00	8.18	7.50	2.60	31.73
Operator Expiration Year	9408.00	2013.09	2016.00	1983.00	2024.00
Max Well Prod (Mcf)	91923.00	50150.01	23486.00	0.00	1959586.00
Monthly Production (Mcf)	18300285.00	4157.01	562.00	0.00	1561502.00
Completion Year	92298.00	2008.48	2008.00	2000.00	2021.00

Table 3.1: Summary Statistics

[4] The RRC reports district-year level average costs of plugging for wells plugged with state funds (Texas Railroad Commission, a).

[5] The status "ORPHANPLUGGED" is a flag for if the well was plugged by a state-funded plugging program. As you can see, there is some implied mismatch between the statuses I impute and RRC records. For example, some wells make the transition from Orphan to Plugged, meaning that these wells were plugged by an operator that failed to renew its operating licence but went on to plug their well anyways. Simillarly, there are some wells that make the transition from orphan to shut-in which means that an unlicensed operator sold the well after its license expired.

	Transition	n
1	ACTIVE-ACTIVE	12731641
2	ACTIVE-PLUGGED	4619
3	ACTIVE-SHUTIN	43984
4	ORPHAN-ORPHAN	249829
5	ORPHAN-ORPHANPLUGGED	1044
6	ORPHAN-PLUGGED	334
7	ORPHAN-SHUTIN	508
8	SHUTIN-ACTIVE	17845
9	SHUTIN-ORPHAN	4418
10	SHUTIN-ORPHANPLUGGED	413
11	SHUTIN-PLUGGED	16949
12	SHUTIN-SHUTIN	5253580

Table 3.2: Observed Transitions

3.4 Model and Estimation

For each month, t, firm j is responsible for choosing the operating status of well i. The firm chooses a status $k \in \{\text{Active (A), shut-in (S), Plugged (P)}\}$. Once a the firm chooses $k = P$, the well exits the firm's set of wells for $t + 1$ onwards. The operator cannot adjust the production of the well and takes it as a given.

Rather than set up a full dynamic model to estimate, I simply estimate a multinomial logit of well status on well characteristics. The goal of this exercise is to be able to make predictions about well states, rather than understand the underlying structural parameters governing firm choices. I assume that the firm chooses k which maximizes its profit. I normalize the profit of being Active to zero and express the profits of shutting-in or plugging as:

$$
\begin{aligned}
\pi_{itS} =& \beta_{S0} + \beta_{S1} p_{gt} + \beta_{S2} C^p_{it} + \sum_h \beta_{Sh} \mathbb{1}(A_{ith}) + \sum_g \beta_{Sg} \mathbb{1}(Prod_{itg}) + \beta_{S3} \mathbb{1}(k_{it-1} = S) + \\
& \beta_{S4} \mathbb{1}(k_{it-1} = A) + \epsilon_{itS} \\
\pi_{itP} =& \beta_{P0} + \beta_{P1} p_{gt} + \beta_{P2} C^p_{it} + \sum_h \beta_{Sh} \mathbb{1}(A_{ith}) + \sum_g \beta_{Sg} \mathbb{1}(Prod_{itg}) + \beta_{P3} \mathbb{1}(k_{it-1} = S) + \\
& \beta_{P4} \mathbb{1}(k_{it-1} = A) + \epsilon_{itP} \\
\pi_{itA} =& \epsilon_{itA}
\end{aligned}
$$

where p_{gt} is the price of gas, C^p_{it} is the well plugging cost, $\mathbb{1}(A_{ith})$ is an indicator variable for five-year bins of a wells age, $\mathbb{1}(Prod_{itg})$ is an indicator variable for bins of a well's maximum observed production, and $\mathbb{1}(k_{it-1} = S)$ and $\mathbb{1}(k_{it-1} = A)$ are indicators for the well's status in $t-1$. I define the bins for maximum observed production by making five equally sized bins based on the observed quantiles in the data. I use maximum observed production because it's a fairly good proxy for the lifetime production of a well. The production from a well tends to decay exponentially over time as pressure in the well decays (Lade and Rudik, 2020). This means that the well age and maximum observed well production can serve as a decent proxy for potential production.

I assume that the ϵ's are drawn i.i.d. from an extreme value I distribution. This will imply that the choice probabilities are given by logit probabilities:

$$
Pr(i \text{ chooses } k \text{ at } t) = \frac{\exp(\pi_{itk})}{\sum_l \exp(\pi_{itl})} \tag{3.1}
$$

The parameters to be estimated are the set of β's, which can be interpreted as marginal effects on the relative log odds. I estimate the parameters by constructing a likelihood equation from equation 3.1, then using maximum likelihood:

$$LL(\beta) = \sum_i \sum_t \sum_k \log(Pr(i \text{ chooses } k \text{ at } t)\mathbb{1}(i \text{ chooses } k \text{ at } t)) \tag{3.2}$$

I choose to use a multinomial logit rather than a full structural specification because of the difficulty of quantifying the option value for shut-in wells. To estimate a full structural model such as a dynamic discrete choice like that of Muehlenbachs (2015), I would need to construct or estimate how much each well would produce if it switched from Shut-in to Active, even if I don't observe a well make that transition. This is a difficult modelling problem, and one of the real strengths of the context studied by Muehlenbachs (2015).[6]

3.5 Results

3.5.1 Model Estimates

I present results of the multinomial logit below in table 3.3. The coefficients can be interpreted as the marginal effect of a change in the independent variable on the log odds ratio of being either shut-in or plugged versus active. For example, the coefficients on plugging cost have the expected sign and imply that a $\$100,000$ increase in plugging costs is associated with a 0.06 increase in the log odds ratio of being shut-in versus active and a 0.04 decrease in the relative odds ratio of plugging versus active. Higher natural gas prices lead to large decreases in shut-in rates and small increases in plugging rates. This is likely an artifact of the fact that producers,

[6]Muehlenbachs (2015) has estimates of recoverable reserves for each well that I cannot construct in this context.

for the most part, cannot adjust a wells production but must take it as a given. In addition, the coefficients imply that as wells age, producers become less likely to plug them. Lastly, the lagged status coefficients imply that there are likely high switching costs between statuses since the best predictor of a well's status in t is it's status in $t-1$.

coefnames	Estimate	Std. error	t value	Pr(> t)	Estimate	Std. error	t value	Pr(> t)
	SHUTIN				PLUG			
intercept	1.755	0.074	23.749	0	-1.668	0.046	-35.985	0
ngprice	-0.158	0.006	-25.615	0	0.047	0.014	3.378	0.001
plug_cost	0.059	0.01	5.664	0	-0.044	0.026	-1.684	0.092
maxprod_quantbin_2	-0.199	0.014	-13.935	0	-0.299	0.03	-9.874	0
maxprod_quantbin_3	-0.444	0.015	-30.014	0	-0.866	0.034	-25.204	0
maxprod_quantbin_4	-0.721	0.016	-45.526	0	-1.202	0.041	-29.566	0
Wellage_years_five_1	0.096	0.014	6.64	0	-0.127	0.033	-3.868	0
Wellage_years_five_2	0.159	0.015	10.373	0	-0.277	0.037	-7.577	0
Wellage_years_five_3	0.057	0.02	2.833	0.005	-0.745	0.051	-14.547	0
Wellage_years_five_4	-0.107	0.053	-2.036	0.042	-1.037	0.146	-7.084	0
lSTATUSACTIVE	-5.801	0.073	-79.022	0	-4.767	0.047	-101.013	0
lSTATUSSHUTIN	5.019	0.075	66.739	0	2.699	0.052	51.855	0

Table 3.3: Multinomial Logistic Results

3.5.2 Counterfactual Simulations

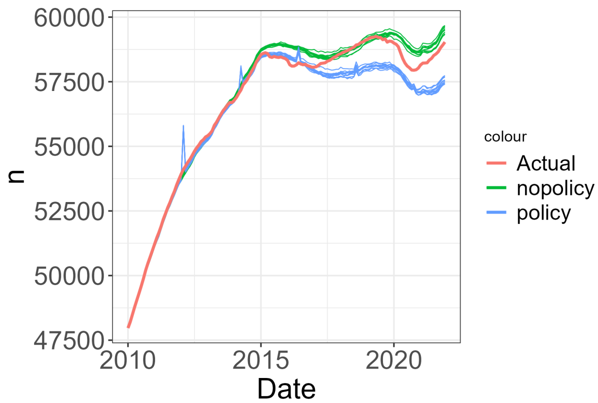
I evaluate a potential policy that requires wells that have been shut-in for two years to either be plugged or come back into production. To evaluate this policy, I take the estimated parameters of the multinomial logit, take draws of the error terms, then forward simulate choices. Once a well is simulated to be "plugged" it is removed from the remaining well set for the remaining periods. Similarly, if a well is "shut-in" while the producer is "delinquent," the well is labeled as "orphaned" and removed from the set for the remaining periods.

In the baseline counterfactual firms can choose between active, shut-in, or plugged. In the policy counterfactual, once firms are shut-in for 24 months, I remove the option to remain shut-in. Therefore, the counterfactual calculation strongly hinges on the IIA assumption implicitly made by the model estimation. Moreover, it also relies strongly on the assumption that if an operator is in business, it is financially able to plug the well.

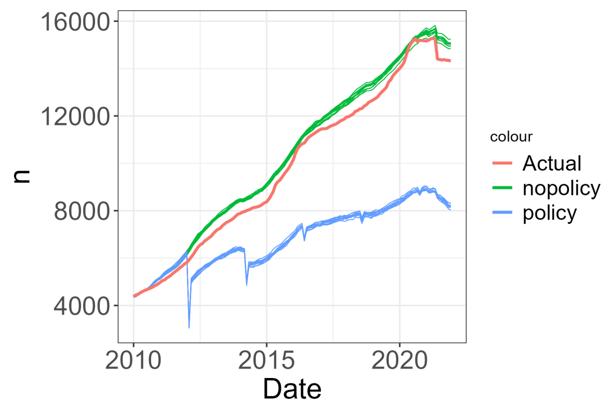
I plot the results of the baseline and counterfactual simulations in figure 3.7 along with the realized outcomes. The realized outcomes is quite close the the simulated outcomes, suggesting that the multinomial logit does a good job of predicting changes over time despite the fact that it's an extremely parsimonious model that includes no time or geographical fixed-effects.

My simulations suggest that this policy would lead to a 39$\%$ drop in the number of orphaned wells by the end of 2021 (averaging across all simulations), a 3% decrease in active wells, a 123% increase in plugged wells, and a 45% decrease in shut-in wells. In the baseline simulations there are on average, 1256.3 orphaned wells by the end of 2021 while in the policy counterfactual there are 761.6. The mean estimated plugging cost in the sample is about $\$71,000$ implying that the policy could have reduced state plugging liabilities by about $\$35$ million over the eleven year sample.

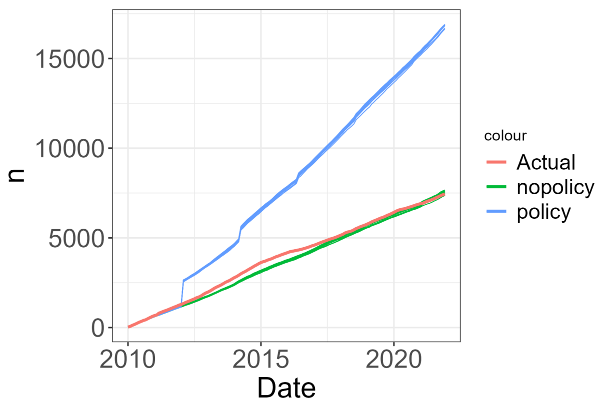
These large estimated policy effects are driven by the simple fact that large numbers of wells remain idle for many years without the policy. Of course, these simulations make two important assumptions about compliance which bias the effect of the policy upwards. First, the simulations assume there will be perfect compliance. Second, the simulations assume that many operators are solvent to pay their plugging costs. Since there are many operators whose plugging costs far exceed their expected discounted stream of revenue from production, these operators may not be solvent in the policy counterfactual.



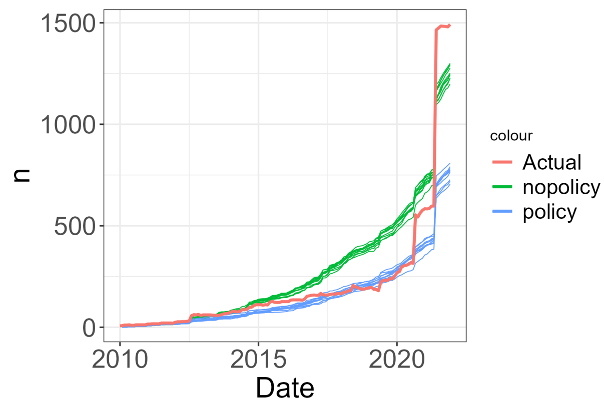
(a) Number of Active Wells



(b) Number of Shut-in Wells



(c) Number of Plugged Wells



(d) Number of Orphaned Wells

Figure 3.7: Simulation Results: These figures plot the actual versus simulated numbers of wells in each state. Observed numbers are plotted in red. Baseline simulations are plotted in green and policy simulation are plotted in blue. Overall, the simulation does a fairly good job of matching well numbers. The policy leads to a substantial drop in well orphanage and a significant increase in well plugging.

3.6 Conclusion

In this paper, I estimate a simple model of transitions between gas well states to inform how a policy requiring firms to either bring inactive wells back into production or plug them after two years of inactivity would affect the number of orphan wells in the state of Texas. I find that this simple policy could lead to substantial decreases in well orphanage and increases in well plugging. I estimate that this policy would have reduced state plugging liabilities by $35 million over my eleven year sample period.

While existing regulations and bonding requirements have helped mitigate the problem of orphan wells to some extent, the current policies fall short in adequately addressing the costs and risks associated with orphan wells. The prevalence of blanket bond usage allow most firms to cover their wells with insufficient financial assurance. Moreover, previous research has shown the environmental and safety hazards associated with these orphaned wells, including increased downstream pollution, methane emissions, and even fatal explosions. This setting is also relevant to other issues of environmental remediation such as in mining or chemicals manufacturing where environmental risks remain even after the end of the useful life of the mine or plant.

Appendix A: Appendix to Chapter 1

A.0.1 Processing the Leases

I run the raw lease data through two rounds of clustering. The first round is intended to identify and remove duplicates. The second round is intended to identify lease amendments and changes to lease ownership.

For the first round of clustering I follow many of the lease processing choices of Herrnstadt, Kellogg, and Lewis (2020). In the Enverus data on leases, there are many likely duplicates that need to be dealt with. There are many cases where the same plot is leased to the same grantee by multiple grantors. These observations are likely a result of undivided mineral interests. For example, multiple members of a family might each show up as a distinct grantor with her own row in the data. I keep all observations labeled as “Leases,” “Lease amendments,” “memos of leases,” and “lease extensions.” I drop observations on mineral rights assignments, lease ratification, royalty deeds, and mineral deeds. Next, I construct a dissimilarity matrix which I pass along to an agglomerative hierarchical clustering method implemented by the agnes function in the `cluster` package, version 2.2.1, in R.

The entries of the dissimilarity matrix are calculated as which I calculate as:

$$d_{ij} = \sum_k m_k(x_i^k, x_j^k)$$

where $m_k(\cdot)$ is equal to 0 if the kth attribute of lease i and j are identical, and positive but less than or equal to 1 otherwise. The attributes and calculation of m_k are as follows:

- $m_k(x_i^k, x_j^k) = |x_i^k - x_j^k|$ for various continuous variables. I use this on both the start and end date of the leases, and the Demerau-Levenshtein string distance between the grantors and grantees. Once I calculate all the pairwise combinations, I then rescale all the observations by dividing through by the max (for each characteristic), so all the date differences lie between 0 and 1.

- $m_A(x_i^A, x_j^A) = \frac{|x_i^A \cap x_j^A|}{2(|x_i^A| + |x_j^A|)}$ - where x_i^A is the spatial polygon of i. This measures the average percentage overlap in area between i and j, and will be bound between 0 (if the polygons do not intersect) and 1 (if the polygons are exactly the same).

Before clustering, there are over 500,000 observations in the data. Constructing a 500,000$\times$500,000 matrix would be infeasible. Even when accounting for the fact that the matrix would be symmetric it would have over 100 billion elements. To deal with this, I run the clustering algorithm at the year of expiration date level.

After this initial round of clustering, primarily intended to identify rows of duplicates, I run this data through one more layer of clustering. This time, I just consider the average percentage overlap between two leases and group them together if the percentage overlap is over 70%. Once two leases are clustered together, I consider the lease with the greatest Instrument Date active at time t to be the active lease and throw out the older ones. Consider the hypothetical scenario below: In this example, for all dates before 12-01-2009, the expiration date would be treated as

Effective Date	Expiration Date	Instrument Date	Polygon
12-01-2007	12-01-2010	12-01-2007	A
12-01-2009	12-01-2012	12-01-2009	A

12-01-2010. So for example at 12-01-2008, there would be 24 months until the lease expiration. Once the new lease becomes active after 12-01-2009, the time until expiration would resent to 36 months.

Bibliography

91 Tex. Natural Resources Code § 104. ???? "NATURAL RESOURCES CODE CHAPTER 91." URL https://statutes.capitol.texas.gov/Docs/NR/htm/NR.91.htm.

Agerton, Mark, Siddhartha Narra, Brian Snyder, and Gregory B. Upton. 2023. "Financial liabilities and environmental implications of unplugged wells for the Gulf of Mexico and coastal waters." *Nature Energy 2023* :1–12URL https://www.nature.com/articles/s41560-023-01248-1.

Agerton, Mark and Gregory B Upton. 2020. "The Economics of Natural Gas Flaring in U.S. Shale: An Agenda for Research and Policy." URL https://www.energy.gov/sites/prod/files/2019/08/f65/.

Alvarez, Ramón A., Daniel Zavala-Araiza, David R. Lyon, David T. Allen, Zachary R. Barkley, Adam R. Brandt, Kenneth J. Davis, Scott C. Herndon, Daniel J. Jacob, Anna Karion, Eric A. Kort, Brian K. Lamb, Thomas Lauvaux, Joannes D. Maasakkers, Anthony J. Marchese, Mark Omara, Stephen W. Pacala, Jeff Peischl, Allen L. Robinson, Paul B. Shepson, Colm Sweeney, Amy Townsend-Small, Steven C. Wofsy, and Steven P. Hamburg. 2018. "Assessment of methane emissions from the U.S. oil and gas supply chain." *Science* 361 (6398):186–188. URL http://science.sciencemag.org/.

Arcidiacono, Peter and Robert A Miller. 2011. "Conditional Choice Probability Estimation of Dynamic Discete Choice Models with Unobserved Heterogeneity." *Econometrica* 79 (6):1823–1867. URL https://www.jstor.org/stable/41336537.

Biven, Megan Milliken and Virginia Palacios. 2022. "Eliminating Orphan Wells and Sites in Texas." Tech. rep., Commission Shift. URL https://commissionshift.org/wp-content/uploads/2022/01/Eliminating-Orphan-Wells-and-Sites-in-Texas_CommissionShift.pdf.

Boomhower, Judson. 2019. "Drilling Like There's No Tomorrow: Bankruptcy, Insurance, and Environmental Risk †." *American Economic Review* 109 (2):391–426. URL https://doi.org/10.1257/aer.20160346.

Chapa, Sergio. 2021. "Texas Oil Body Makes It a Little Harder for Wells to Flare Gas - Bloomberg." URL https://www.bloomberg.com/news/articles/2021-09-14/texas-oil-body-makes-it-a-little-harder-for-wells-to-flare-gas.

Congressional Research Service. 2022. "Inflation Reduction Act Methane Emissions Charge: In Brief." Tech. rep., Congressional Research Service. URL https://crsreports.congress.gov/product/pdf/R/R47206.

Covert, Thomas and Ryan Kellogg. 2017. "Crude by Rail, Option Value, and Pipeline Investment." *Becker Friedman Institute Working Paper* URL https://bfi.uchicago.edu/working-paper/crude-by-rail-option-value-and-pipeline-investment/.

Covert, Thomas R. 2015. "Experiential and Social Learning in Firms: The Case of Hydraulic Fracturing in the Bakken Shale." *SSRN Electronic Journal* URL https://papers.ssrn.com/abstract=2481321.

Covert, Thomas R and Richard L Sweeney. 2019. "NBER WORKING PAPER SERIES RELINQUISHING RICHES: AUCTIONS VS INFORMAL NEGOTIATIONS IN TEXAS OIL AND GAS LEASING." URL https://github.com/rlsweeney/public_cs_texas.

———. 2022. "Secrecy Rules and Exploratory Investment: Theory and Evidence from the Shale Boom *." *Working Paper* URL http://www.richard-sweeney.com/pdfs/cs_disclosure.pdf.

Davis, Lucas W. 2015. "Policy Monitor—Bonding Requirements for U.S. Natural Gas Producers." *https://doi.org/10.1093/reep/reu015* 9 (1):128–144. URL https://www.journals.uchicago.edu/doi/10.1093/reep/reu015.

Décaire, Paul H, Erik Gilje, and Jérôme Taillard. 2019. "Real Option Exercise: Empirical Evidence." *SSRN Electronic Journal* URL https://papers.ssrn.com/abstract=3342565.

Department of Energy. 2022. "Queued Up… But in Need of Transmission Unleashing the Benefits of Clean Power with Grid Infrastructure." Tech. rep. URL https://www.energy.gov/sites/default/files/2022-04/Queued%20Up%E2%80%A6But%20in%20Need%20of%20Transmission.pdf.

Dunkle Werner, Karl and Wenfeng Qiu. 2020. "Hard to Measure Well: Can Feasible Policies Reduce Methane Emissions? Hard to Measure Well: Can Feasible Policies Reduce Methane Emissions?" *Energy Institute as Haas Working Paper* URL https://haas.berkeley.edu/energy-institute/about/funders/.

EDF. 2022. "Methodology for EDF's Permian Methane Analysis Project (PermianMAP) Data Collection and Analysis." Tech. rep. URL https://www.edf.org/sites/default/files/documents/PermianMapMethodology_1.pdf.

Fischer, Marc L., Eric D. Lebel, and Robert B. Jackson. 2020. "Quantifying Methane from California's Plugged and Abandoned Oil and Gas Wells." Tech. rep., California Energy Commission. URL https://www.energy.ca.gov/sites/default/files/2021-05/CEC-500-2020-052.pdf.

Gerard, David and Elizabeth J. Wilson. 2009. "Environmental bonds and the challenge of long-term carbon sequestration." *Journal of environmental management* 90 (2):1097–1105. URL https://pubmed.ncbi.nlm.nih.gov/18619728/.

Gowrisankaran, Gautam, Marc Rysman, Dan Ackerberg, Victor Aguirregabiria, Ana Aizcorbe, Rabah Amir, Lanier Benkard, Steve Berry, Sofronis Clerides, Tim Erickson, Simon Gilchrist, and Avi Goldfarb. 2012. "Dynamics of Consumer Demand for New Durable Goods." *Source: Journal of Political Economy* 120 (6):1173–1219.

Griffin, Ronald C., Daniel Bromley, Ronald C. Griffin, and Daniel Bromley. 1982. "Agricultural Runoff as a Nonpoint Externality: A Theoretical Development." *American Journal of Agricultural Economics* 64 (3):547–552. URL https://EconPapers.repec.org/RePEc:oup:ajagec:v:64:y:1982:i:3:p:547-552.

Henningsen, Arne and Ott Toomet. 2011. "maxLik." URL https://cran.r-project.org/web/packages/maxLik/index.html.

Herrnstadt, Evan, Ryan Kellogg, and Eric Lewis. 2020. "The Economics of Time-Limited Development Options: The Case of Oil and Gas Leases." *SSRN Electronic Journal* URL https://papers.ssrn.com/abstract=3604162.

Hlavac, Marek. 2018. "stargazer: Well-Formatted Regression and Summary Statistics Tables." URL https://CRAN.R-project.org/package=stargazer.

Ho, Jacqueline S., Jhih Shyang Shih, Lucija A. Muehlenbachs, Clayton Munnings, and Alan J. Krupnick. 2018. "Managing Environmental Liability: An Evaluation of Bonding Requirements for Oil and Gas Wells in the United States." *Environmental Science and Technology* 52 (7):3908–3916. URL https://pubs.acs.org/doi/abs/10.1021/acs.est.7b06609.

Hotz, Joseph A. and Robert A. Miller. 1993. "Conditional choice probabilities and the estimation of dynamic models." *Review of Economic Studies* 60 (3):497–529.

ICF. 2018. "North America Midstream Infrastructure through 2035: Significant Development Continues." Tech. rep. URL https://www.ingaa.org/File.aspx?id=34658.

Interagency Working Group on Social Cost of Greenhouse Gases. 2021a. "Technical Support Document: Social Cost of Carbon, Methane, and Nitrous Oxide Interim Estimates under Executive Order 13990." Tech. rep. URL https://www.whitehouse.gov/wp-content/uploads/2021/02/TechnicalSupportDocument_SocialCostofCarbonMethaneNitrousOxide.pdf.

———. 2021b. "Technical Support Document: Social Cost of Carbon, Methane, and Nitrous Oxide Interim Estimates under Executive Order 13990 Interagency Working Group on Social Cost of Greenhouse Gases, United States Government With participation by Council of Economic Advisers Council on Environmental Quality." .

Kang, Mary, Shanna Christian, Michael A. Celia, Denise L. Mauzerall, Markus Bill, Alana R. Miller, Yuheng Chen, Mark E. Conrad, Thomas H. Darrah, and Robert B. Jackson. 2016. "Identification and characterization of high methane-emitting abandoned oil and gas wells." *Proceedings of the National Academy of Sciences of the United States of America* 113 (48):13636–13641. URL `https://www.pnas.org/doi/abs/10.1073/pnas.1605913113`.

Kellogg, Ryan. 2014. "The effect of uncertainty on investment: Evidence from texas oil drilling." *American Economic Review* 104 (6):1698–1734.

Kusnetz, Nicholas. 2011. "Deteriorating Oil and Gas Wells Threaten Drinking Water Across the Country - Scientific American." *Scientific American* URL `https://www.scientificamerican.com/article/deteriorating-oil-gas-wells-threatening-americas-drinking-water/`.

Lade, Gabriel E. and Ivan Rudik. 2020. "Costs of inefficient regulation: Evidence from the Bakken." *Journal of Environmental Economics and Management* 102:102336.

Maechler, Martin, Peter Rousseeuw, Anja Struyf, Mia Hubert, and Kurt Hornik. 2021. "cluster: Cluster Analysis Basics and Extensions." URL `https://CRAN.R-project.org/package=cluster`.

Marks, Levi. 2022. "The Abatement Cost of Methane Emissions from Natural Gas Production." *https://doi.org/10.1086/716700* 9 (2):165–198. URL `https://www.journals.uchicago.edu/doi/10.1086/716700`.

McFarland, John. 2016. "The Problem of Abandoned Wells." URL `https://www.oilandgaslawyerblog.com/problem-abandoned-wells/`.

Méndez-Ruiz, Andrés. 2019. "Coase on Fire: Natural Gas Flaring Abatement Technology Adoption and its Regulation in North Dakota." URL `https://www.perc.org/wp-content/uploads/2019/08/Andres_Workshop_2_Paper.pdf`.

Muehlenbachs, Lucija. 2015. "A DYNAMIC MODEL OF CLEANUP: ESTIMATING SUNK COSTS IN OIL AND GAS PRODUCTION." *International Economic Review* 56 (1):155–185. URL `https://onlinelibrary.wiley.com/doi/full/10.1111/iere.12098https://onlinelibrary.wiley.com/doi/abs/10.1111/iere.12098https://onlinelibrary.wiley.com/doi/10.1111/iere.12098`.

Murphy, Alvin. 2018. "A Dynamic Model of Housing Supply." *American Economic Journal: Economic Policy* 10 (4):243–67.

Paddock, James L., Daniel R. Siegel, and James L. Smith. 1988. “Option valuation of claims on real assets: The case of offshore petroleum leases.” *Quarterly Journal of Economics* 103 (3):479–508.

Plant, Genevieve, Eric A. Kort, Adam R. Brandt, Yuanlei Chen, Graham Fordice, Alan M. Gorchov Negron, Stefan Schwietzke, Mackenzie Smith, and Daniel Zavala-Araiza. 2022. “Inefficient and unlit natural gas flares both emit large quantities of methane.” *Science* 377 (6614):1566–1571. URL https://www.science.org/doi/10.1126/science.abq0385.

Raimi, Daniel, Alan J. Krupnick, Jhih Shyang Shah, and Alexandra Thompson. 2021. “Decommissioning Orphaned and Abandoned Oil and Gas Wells: New Estimates and Cost Drivers.” *Environmental Science and Technology* 55 (15):10224–10230. URL https://pubs.acs.org/doi/full/10.1021/acs.est.1c02234.

Rochet, Jean Charles and Jean Tirole. 2003. “Platform Competition in Two-Sided Markets.” *Journal of the European Economic Association* 1 (4):990–1029. URL https://academic.oup.com/jeea/article/1/4/990/2280902.

Segerson, Kathleen. 1988. “Uncertainty and Incentives for Nonpoint Pollution Control.” Tech. rep.

Texas Railroad Commission. ????a. “Cost Calculation: Plugging Cost Estimates.” URL https://www.rrc.texas.gov/oil-and-gas/compliance-enforcement/hb-2259-hb-3134-inactive-well-requirements/cost-calculation/.

———. ????b. “Flaring Regulation FAQs.” URL https://www.rrc.texas.gov/about-us/faqs/oil-gas-faqs/flaring-regulation/.

U.S. Department of Energy. 2020. “Texas Natural Gas Flaring and Venting Regulations.” Tech. rep. URL https://www.energy.gov/fecm/articles/texas-natural-gas-flaring-and-venting-regulations-fact-sheet.

———. 2022. “New Mexico Natural Gas Flaring and Venting Regulations.” Tech. rep. URL https://www.energy.gov/sites/default/files/2022-06/NewMexico-state-profile.pdf.

U.S. Energy Information Administration. 2021a. “Maps: Oil and Gas Exploration, Resources, and Production - Energy Information Administration.” URL https://www.eia.gov/maps/maps.htm.

———. 2021b. “Natural Gas Vented and Flared.” URL https://www.eia.gov/dnav/ng/ng_prod_sum_a_EPG0_VGV_mmcf_a.htm.

U.S. Environmental Protection Agency. ???? “Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990-2019 — US EPA.” URL https://www.epa.gov/ghgemissions/inventory-us-greenhouse-gas-emissions-and-sinks-1990-2019.

———. 2022. "Inventory of U.S. Greenhouse Gas Emissions and Sinks 1990-2020: Updates for Abandoned Oil and Gas Wells." URL https://www.epa.gov/system/files/documents/2021-09/2022-ghgi-update-abandoned-wells_sept-2021.pdf.

Varon, Daniel J., Daniel J. Jacob, Melissa Sulprizio, Lucas A. Estrada, William B. Downs, Lu Shen, Sarah E. Hancock, Hannah Nesser, Zhen Qu, Elise Penn, Zichong Chen, Xiao Lu, Alba Lorente, Ashutosh Tewari, and Cynthia A. Randles. 2022. "Integrated Methane Inversion (IMI 1.0): a user-friendly, cloud-based facility for inferring high-resolution methane emissions from TROPOMI satellite observations." *Geoscientific Model Development* 15 (14):5787–5805.

Xepapadeas, Anastasios. 2011. "The Economics of Non-Point-Source Pollution." *https://doi.org/10.1146/annurev-resource-083110-115945* 3:355–373. URL https://www.annualreviews.org/doi/abs/10.1146/annurev-resource-083110-115945.

Zhang, Yuzhong, Ritesh Gautam, Sudhanshu Pandey, Mark Omara, Joannes D. Maasakkers, Pankaj Sadavarte, David Lyon, Hannah Nesser, Melissa P. Sulprizio, Daniel J. Varon, Ruixiong Zhang, Sander Houweling, Daniel Zavala-Araiza, Ramon A. Alvarez, Alba Lorente, Steven P. Hamburg, Ilse Aben, and Daniel J. Jacob. 2020. "Quantifying methane emissions from the largest oil-producing basin in the United States from space." *Science Advances* 6 (17):eaaz5120. URL http://advances.sciencemag.org/.