Online Appendices

Contents

A Data and context

A.1 Dataset Construction

The data construction process combines several datasets. The contract datasets are:

- IHS contract dataset
- Rigzone contract dataset
- Rigzone order book

The well datasets are all from the regulator (BSEE). These are:

- The borehole dataset
- The permit dataset
- The lease dataset
- The well activity report dataset

I begin by merging the borehole, permit, lease, and well activity report datasets on the unique API well number to obtain a single dataset with all well characteristics at the well level which I call the 'well database'.

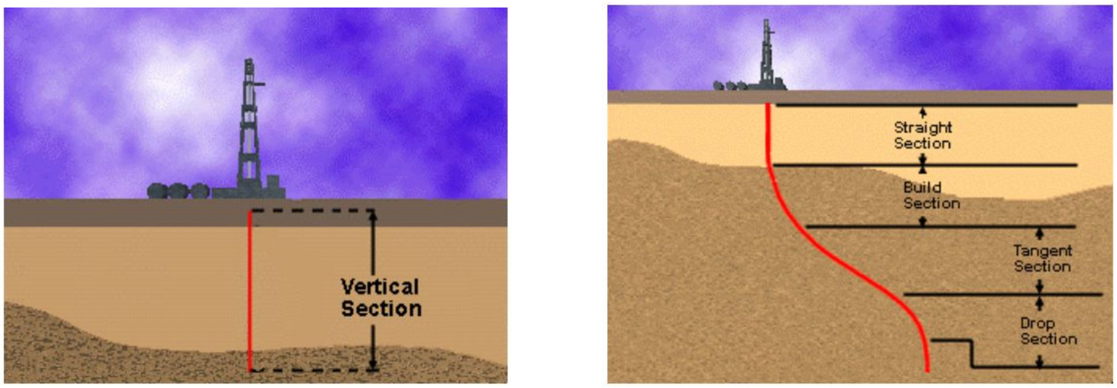
Next I create the 'contract database'. To keep the analysis focused on one market I analyze jackup rigs that drilled wells between 01 January 2000 and 31 December 2009. That is, I remove contracts drilled by deepwater semisubmersibles, drillships, and fixed platforms. There are 3380 contracts for jackup rigs in total. I further remove 17 'workover' rigs, whose main purpose is to reenter wells, typically for maintenance. These rigs rarely drill new wells and so are not in direct competition with drilling rigs. I identify workover rigs as any rig offered by the drilling company Nabors as well as rigs whose status is 'workover' in the Rigzone contract $>$ 80% of the time.

Sometimes the rig name differs between the contract data and the well data due to ownership changes and so I first map rig names between the two databases using the Rigzone order book (which has previous rig names), and the websites *maritime-connector.com*, *marinetraffic.com*, and *vesselfinder.com*. I also use these websites plus the Rigzone order book to find the maximum drilling depth of each rig. I merge all but one of the rig names in the IHS contract dataset to the well dataset in this way. In total, removing the workover rigs and merging with the rig name crosswalk reduces the contract dataset to 3103 contracts.

In total there are 3103 contracts and 5129 wells in the datasets for the years 2000-2009. The procedure successfully results in matching 2355 contracts and 4937 wells. I further impute the characteristics of 300

Figure A-1: Diagram of a simple vs complex well

(a) A simple well (b) A complex well



Note: This figure gives an example of a simple well design and a complex well design. Simple wells will rank low on the Mechanical Risk Index whereas complex wells will rank high. Panel (a) illustrates a simple well - in this case it is just a short vertical hole. Panel (b) illustrates a complex well. In this case there are many connected sections and curves. A more complex well design increases the risk the rig will encounter a problem and high-specification rigs are more suited to drilling these types of wells. Source: https://directionaldrilling.wordpress.com/

contracts if the contract was an extension/renegotiation. In total 2655 contracts (86%) and 4937 wells (96%) are matched.

Sometimes a drilling contract will contain two or more wells. Therefore, I collapse multi-well contracts by taking the mean well complexity, the mean well water depth, and the mean well value. The resulting and final dataset that I use for estimation is at the contract level. Finally, I compute the 'number of rigs' as the average number of rigs of each type per month in the sample.

A.2 Computing the Mechanical Risk Index

This section draws directly from Kaiser (2007). The Mechanical Risk Index was developed by Conoco engineers in the 1980s (Kaiser (2007)). The idea behind the index is to collapse the many dimensions that a well can differ on into a one-dimensional ranking of well complexity. Well complexity is directly related to the cost of drilling a well: these wells run an increased risk of technical issues which may require new materials or result in blowouts. Figure A-1 contains an example of a complex versus a simple well.

The Mechanical Risk Index is computed by first computing 'component factors':

$$\phi_1 = \left(\frac{TD + WD}{1000}\right)^2$$
$$\phi_2 = \left(\frac{VD}{1000}\right)^2 \left(\frac{TD + HD}{VD}\right)$$
$$\phi_3 = (MW)^2 \left(\frac{WD + VD}{VD}\right)$$
$$\phi_4 = \phi_1 \sqrt{NS + \frac{MW}{(NS)^2}}$$

Here TD is total depth in feet, WD is water depth in feet, VD is vertical depth in feet, MW is mud weight in ppg, NS is the number of strings.

Next 'key drilling factors' are computed. These are: $\psi_1 = 3$ if there is a horizontal sections; $\psi_2 = 3$ if there is a J-curve; $\psi_3 = 2$ if there is an S-curve; ψ_4 if there is a subsea well; $\psi_5 = 1$ if there is an H_2S/CO_2 environment; $\psi_6 = 1$ if there is a hydrate environment; $\psi_7 = 1$ if there is a depleted sand section; $\psi_8 = 1$ if there is a salt section; $\psi_9 = 1$ if there is a slimhole, $\psi_{10} = 1$ if there is a mudline suspension system installed; $\psi_{11} = 1$ if there is coring; $\psi_{12} = 1$ if there is shallow water flow potential; $\psi_{13} = 1$ if there is riserless mud to drill shallow water flows; $\psi_{14} = 1$ if there is a loop current.

The Mechanical Risk Index is then computed as:

$$MRI = \left(1 + \frac{\sum_j \psi_j}{10}\right) \sum_i \psi_i$$

In my data I have excellent information for all wells on TD, WD, VD, HD using the BSEE permit data and the BSEE borehole data. I have data for MW, NS for a subset of wells and I impute the remainder based on geological proximity (whether they are in the same 'field') - based on the fact that geological conditions are usually similar for nearby wells. I use a similar imputation procedure for the bid data.

Computing the 'key drilling factors' ψ_j presents a greater challenge because the data are either not recorded (e.g. if there is shallow water flow potential) or would need to be imputed from well velocity surveys (e.g. if there is an S-curve). Rather than guess I set all $\psi_j = 0$. The implication for the index is that there will be a less accurate measure of complexity which will result in measurement error.

A.3 Other factors which might affect rig choice

In this sub-section I provide a longer discussion about two alternative factors which might affect rig choice.

One factor I consider is distance between a rig and a well. I initially perform tests about whether distance is a factor for rig choice in the data. After finding that it is not an important determinant, I then discuss institutional details about why this is the case.

In order to test whether distance is a factor I first need a measure of distance. This is challenging because, although I have the latitude and longitude coordinates of each well, I do not have positional data on rigs. As a result I construct a measure in the following way: I take the position of the final well that a rig is drilling on a contract and take the distance to the first well on the subsequent contract. This will be a good proxy for the distance between a rig and a well particularly when the two contracts immediately follow each other in time. Sometimes this measure for distance is not available (for example, the contract is the first in the data for a particular rig) or corresponds to a contract extension where the rig is mechanically near the next well, and so I also interact the distance measure with an indicator for if (i) the distance measure is observed and (ii) it corresponds to a new contract.

I then include this distance metric in hedonic regressions of price on characteristics in Table A-1. Overall, distance does affect prices (about $1.3 thousand dollars/day for a distance of 100 miles) but this magnitude is extremely small: 100 miles is the average value for 'distance' (when it is observed) but this changes prices by only 2%.

Why is distance not a major factor for within-field rig moves? Transportation costs in the offshore oil and gas industry mainly come in the form of lost drilling time. However, drilling rigs are extremely mobile and take around 1 day to move across the Gulf of Mexico. When compared to the average new contract length (around 65 days), a back-of-the-envelope calculation implies that choosing a far away rig over a nearby rig would increase costs by around 1.5% for the average contract, which is consistent with the above value for the extent to which distance is capitalized in prices. Therefore, I do not include distance to a well as a factor for rig choice in the model.

The second factor - past experience between a rig operator and a well owner - has been shown to be a consideration for rig choice in the *onshore* oil and gas industry (Kellogg (2011)). I capture repeated contracting in my model by allowing for contract extensions. However, for new contracts, I assume that agents' decisions about who to match with are independent of past experience. This modeling assumption seems to be supported by the data: I find 66 percent of new contracts are between a rig-owner pair who have not worked together in the previous 2 years.[1]

B Proofs

B.1 Microfoundation for the targeting weights

In this sub-section I provide more details for the micro-foundation for how projects contact capital.

Denote each unit of available capital by j and the corresponding type as y_j. Similarly, denote each searching project by i and its corresponding type by x_i. Using this notation, for a project of type x_i, the

[1] I use 2 years as my cutoff for a 'relationship' because that is the definition used by Kellogg (2011).

Table A-1: Hedonic regressions of price on characteristics

	(1) Price/day	(2) Price/day	(3) Price/day
1[Low-spec]	45.5***	45.4***	37.8***
	(0.6)	(0.6)	(1.4)
1[Low-spec] × Complexity	-4.6***	-4.4**	-5.2***
	(1.7)	(1.7)	(1.7)
1[Mid-spec]	52.0***	51.8***	44.3***
	(0.6)	(0.6)	(1.4)
1[Mid-spec] × Complexity	2.2*	2.1*	1.8
	(1.1)	(1.1)	(1.1)
1[High-spec]	70.4***	70.3***	62.7***
	(0.8)	(0.8)	(1.6)
1[High-spec] × Complexity	10.7***	10.7***	10.6***
	(1.6)	(1.6)	(1.7)
Distance (100s of miles)	1.3**	1.3**	1.3**
	(0.5)	(0.5)	(0.5)
Value of hydrocarbons		0.7*	0.7*
		(0.4)	(0.4)
State and rig-type interactions	Yes	Yes	Yes
Contract-type controls	No	No	Yes
N	2655	2655	2655

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Prices are in thousands of US dollars (per day). Contract duration is in months. Value is the quantity of hydrocarbons in a well multiplied by the current natural gas price. State and rig type interactions correspond to including controls for four state variables: gas price, and the number of available rigs of each type, as well as the interactions of these variables with an indicator for each of the three rig types. I de-mean the complexity index variable, as well as the state variables, so that the indicators 1[Low-spec], 1[Mid-spec], 1[High-spec], correspond to the average price per day of each rig type. Robust standard errors are in brackets.

(expected) value of targeting capital j is $\pi_{x_i y_j t}$.

In the special case where search is perfectly directed then each potential project i will choose j to solve $\max_j \pi_{x_i y_j t}$. In my setting, I allow for a more flexible search technology by instead modeling potential projects targeting capital based on a *perceived value* $\hat{\pi}_{x_i y_j t}$ which is defined as:

$$\hat{\pi}_{x_i y_j t} = \pi_{x_i y_j t} - \gamma_1 1\left[x_i \notin A_{y_j t}\right] + \epsilon_{ijt}^{target} \tag{A-1}$$

I assume that ϵ_{ijt}^{target} are drawn from an i.i.d. type-1 extreme value distribution with scale parameter $1/\gamma_0$. The interpretation of γ_0 and γ_1 is that they are 'targeting parameters' that index how precisely a project can target capital. I allow for targeting to be responsive to both whether the match will be rejected (the parameter γ_1) as well as the overall quality of the match: γ_0.[2]

The conditional choice probabilities that result from $\max_j \hat{\pi}_{x_i y_j t}$ are given as follows:

$$P_{ijt} = \frac{\exp\Big(\gamma_0\left[\pi_{x_i y_j t} - \gamma_1 1\left[x_i \notin A_{y_j t}\right]\right]\Big)}{\sum_k \exp\Big(\gamma_0\left[\pi_{x_i y_k t} - \gamma_1 1 1\left[x_i \notin A_{y_k t}\right]\right]\Big)} \tag{A-2}$$

Finally, noting that $P_{ijt} = P_{ij't}$ for all capital of the same type $y_j = y_{j'}$, and that there are n_{yt} rigs of type y, Equation (A-2) can be aggregated to form the probability of a project of type x targeting a rig of type y:

$$\omega_{yt}(x) = \frac{n_{yt}\exp\Big(\gamma_0\Big[\pi_{xyt} - \gamma_1 1\left[x \notin A_{yt}\right]\Big]\Big)}{\sum_{k\in Y} n_{kt}\exp\Big(\gamma_0\Big[\pi_{xkt} - \gamma_1 1\left[x \notin A_{kt}\right]\Big]\Big)} \tag{A-3}$$

B.2 Constructing U_{yt} from the data

The aim is to show that the value of search U_{yt} can be computed from the data.

Writing out the value of searching for rig y at time t:

$$\begin{aligned} U_{yt} &= \int_z \max\Big\{V_{zyt}, \beta\mathbb{E}_t U_{y,t+1}\Big\} h_{zyt}dz + h_{\emptyset yt}\beta\mathbb{E}_t U_{y,t+1} && \text{(A-4)}\\ &= \int_{z\in A_{yt}} h_{zyt}V_{zyt}dz + \left(\int_{z\notin A_{yt}} h_{zyt}dz + h_{\emptyset yt}\right)\cdot\beta\mathbb{E}_t U_{y,t+1} && \text{(A-5)}\\ &= \sum_{n\in\{2,3,4\}} \mathbb{P}_{\tau=n,yt}\cdot\mathbb{E}_t\mathbb{E}_{x_{-\tau}|\tau=n,t}V_{xyt} + \mathbb{P}_{\tau=0,yt}\cdot\beta\mathbb{E}_t U_{y,t+1} && \text{(A-6)}\end{aligned}$$

where, in the third equality, n indexes an element in the set of possible contract lengths and $\mathbb{E}_{x_{-\tau}|\tau=n,t}$ is an expectation taken over all the covariates in a potential project (except τ) conditional on $\tau = n$ at

[2]Whether a match could be rejected might be more salient to capital owners than other features of the match - I allow the data to determine whether this is the case.

time t (i.e. the expectation is over the covariates x_{quantity} and $x_{\text{complexity}}$). Recall that value of a match to the capital owner is:

$$V_{xyt} = \sum_{k=0}^{\tau-1} \beta^k p_{xyt} + \beta^\tau \mathbb{E}_t \Big[\eta_{xy,t+\tau} V_{xy,t+\tau} + (1-\eta_{xy,t+\tau}) U_{y,t+\tau} \Big] \tag{A-7}$$

Taking conditional expectations of both sides of Equation (A-7) with respect to the covariates (except τ):

$$\begin{aligned}
\mathbb{E}_{x_{-\tau}|\tau=n,t} V_{xyt} &= \mathbb{E}_{x_{-\tau}|\tau=n,t} \Big[\sum_{k=0}^{n-1} \beta^k p_{xyt} \Big] \\
&\quad + \beta^n \mathbb{E}_t \mathbb{E}_{x_{-\tau}|\tau=n,t} \Big[\eta_{xy,t+n} V_{xy,t+n} + (1-\eta_{xy,t+n}) U_{y,t+n} \Big] && \text{(A-8)} \\
&= \mathbb{E}_{x_{-\tau}|\tau=n,t} \Big[\sum_{k=0}^{n-1} \beta^k p_{xyt} \Big] \\
&\quad + \beta^n \mathbb{E}_t \mathbb{E}_{x_{-\tau}|\tau=n,t} \Big[\eta_{xy,t+n} \sum_{k=0}^{n-1} \beta^k p_{xy,t+n} + (1-\eta_{xy,t+n}) U_{y,t+n} \Big] \\
&\quad + \beta^{2n} \mathbb{E}_t \mathbb{E}_{x_{-\tau}|\tau=n,t} \Big[\\
&\quad\quad \eta_{xy,t+n} \eta_{xy,t+2n} \sum_{k=0}^{n-1} \beta^k p_{xy,t+2n} + \eta_{xy,t+n} (1-\eta_{xy,t+2n}) U_{y,t+2n} \Big] \\
&\quad + \ldots && \text{(A-9)} \\
&= \sum_{k=0}^{n-1} \beta^k \mathbb{E}_{x_{-\tau}|\tau=n,t} \Big[p_{xyt} \Big] \\
&\quad + \beta^n \mathbb{E}_t \sum_{k=0}^{n-1} \beta^k \mathbb{E}_{x_{-\tau}|\tau=n,t} \Big[\eta_{xy,t+n} p_{xy,t+n} \Big] \\
&\quad + \beta^n \mathbb{E}_t \mathbb{E}_{x_{-\tau}|\tau=n,t} \Big[(1-\eta_{xy,t+n}) U_{y,t+n} \Big] \\
&\quad + \beta^{2n} \mathbb{E}_t \sum_{k=0}^{n-1} \beta^k \mathbb{E}_{x_{-\tau}|\tau=n,t} \Big[\eta_{xy,t+n} \eta_{xy,t+2n} p_{xy,t+2n} \Big] \\
&\quad + \beta^{2n} \mathbb{E}_t \mathbb{E}_{x_{-\tau}|\tau=n,t} \Big[\eta_{xy,t+2n} (1-\eta_{xy,t+n}) U_{y,t+2n} \Big] \\
&\quad + \ldots && \text{(A-10)}
\end{aligned}$$

Note that the per-period payoff $\mathbb{E}_{x_{-\tau}|\tau=n,t}\Big[p_{xyt}\Big] = \bar{p}_{\tau=n,yt}$, which is the average price (conditional on the rig type, state, and contract length), and is directly observed in the data. Furthermore, the values of extending the contract $\mathbb{E}_{x_{-\tau}|\tau=n,t}\Big[\eta_{xy,t+n} p_{xy,t+n}\Big]$, and $\mathbb{E}_{x_{-\tau}|\tau=n,t}\Big[(1-\eta_{xy,t+n}) U_{y,t+n}\Big]$, etc, can be constructed, since they are just a function of the observed probabilities of extension, observed prices of extensions, or future values of U.

Although the probabilities of extension and the prices of extensions are observed directly in the data and so could be computed in principle, I only have limited data on extensions and a large dimensional space over which these components need to be computed. Therefore, I need to make some further assumptions about how to aggregate these components. The main assumption that I make is that these components are well-approximated by combinations of the mean empirical probabilities of extensions and mean empirical probabilities of prices computed only at the state where the extension is occurring. Therefore, for example, I approximate the second component of Equation (A-10) by:

$$\mathbb{E}_{x_{-\tau}|\tau=n,t}\left[\eta_{xy,t+n}p_{xy,t+n}\right] = \bar{\eta}_{\tau=n,y,t+n} \cdot \bar{p}_{\tau=n,y,t+n} \tag{A-11}$$

In the above equation $\bar{\eta}_{\tau=n,y,t+n}$ denotes the probability of extending a contract of length n for a rig of type y at the state at time $t+n$. Similar, $\bar{p}_{\tau=n,y,t+n}$ denotes the average price. The assumption that the average probability of extending a contract, and the average price if an extension occurs, are well-approximated by these components computed only at the state where the extension is occurring, is restrictive. For example, it rules out the dependence of the probability of extending a contract for a second time at $t+2n$ on the state at the first extension at $t+n$. This implicitly ignores that different states at $t+n$ might cause different types of matches to survive at $t+n$, which might lead to different probabilities of extension at $t+2n$. However, allowing the average probabilities of extension (and also the extension prices) to be dependent on the entire sequence of states from when the contract was initially signed is not feasible in this setting due to limited data. I emphasize that in practice this assumption is probably not too strong in this application. For example, contracts typically do not survive after one or two extensions, and so most of the variation in the value functions is driven by the state affecting the probabilities of being matched in a new contract (and the corresponding price).

Overall:

$$\begin{aligned} U_{yt} = \sum_{n\in\{2,3,4\}} \mathbb{P}_{\tau=n,yt} \cdot \Bigg\{ & \\ & \sum_{k=0}^{n-1} \beta^k \bar{p}_{\tau=n,yt} + \beta^n \mathbb{E}_t \left[\bar{\eta}_{\tau=n,y,t+n} \Big(\sum_{k=0}^{n-1} \beta^k \bar{p}_{\tau=n,y,t+n} + ... \Big) + (1-\bar{\eta}_{\tau=n,y,t+n}) U_{y,t+n} \right] \\ \Bigg\} & + \mathbb{P}_{\tau=0,yt} \cdot \beta \mathbb{E}_t U_{y,t+1} \end{aligned} \tag{A-12}$$

Notice that the above expression implies that U_{yt} can be written in terms of the average price of extensions, the average price of new contracts, the probability of matching different length contracts, the extension probability, and future values of U, which proves the result.

B.3 Proof that prices can be written as in Section 4.3

Under Nash bargaining the surplus is split in the following way:

$$V_{xyt} - \beta \mathbb{E}_t U_{y,t+1} = \delta S_{xyt} \tag{A-13}$$

Substituting in Equation (8) into V_{xyt}:

$$\sum_{k=0}^{\tau-1}\beta^k p_{xyt} + \beta^\tau \mathbb{E}_t\Big[\eta_{xy,t+\tau}V_{xy,t+\tau} + (1-\eta_{xy,t+\tau})U_{y,t+\tau}\Big] - \beta\mathbb{E}_t U_{y,t+1}$$
$$= \delta\Big[W_{xyt} + V_{xyt} - \beta\mathbb{E}_t U_{y,t+1}\Big] \tag{A-14}$$

Further substituting in for V_{xyt} and W_{xyt}:

$$\sum_{k=0}^{\tau-1}\beta^k p_{xyt} + \beta^\tau \mathbb{E}_t\Big[\eta_{xy,t+\tau}V_{xy,t+\tau} + (1-\eta_{xy,t+\tau})U_{y,t+\tau}\Big] - \beta\mathbb{E}_t U_{y,t+1}$$
$$= \delta\Big[\sum_{k=0}^{\tau-1}\beta^k v_{xyt,k} + \beta^\tau \mathbb{E}_t\Big[\eta_{xy,t+\tau}V_{xy,t+\tau} + \eta_{xy,t+\tau}W_{xy,t+\tau} + (1-\eta_{xy,t+\tau})U_{y,t+\tau}\Big] - \beta\mathbb{E}_t U_{y,t+1}\Big] \tag{A-15}$$

Substituting in for surplus in the above equation:

$$\sum_{k=0}^{\tau-1}\beta^k p_{xyt} + \beta^\tau \mathbb{E}_t\Big[\eta_{xy,t+\tau}\Big[\delta S_{xy,t+\tau} + \beta U_{y,t+\tau+1}\Big] + (1-\eta_{xy,t+\tau})U_{y,t+\tau}\Big] - \beta\mathbb{E}_t U_{y,t+1}$$
$$= \delta\Big[\sum_{k=0}^{\tau-1}\beta^k v_{xyt,k} + \beta^\tau \mathbb{E}_t\Big[\eta_{xy,t+\tau}\Big[S_{xy,t+\tau} + \beta U_{y,t+\tau+1}\Big] + (1-\eta_{xy,t+\tau})U_{y,t+\tau}\Big] - \beta\mathbb{E}_t U_{y,t+1}\Big] \tag{A-16}$$

Rearranging the equation further:

$$\sum_{k=0}^{\tau-1}\beta^k p_{xyt} = \delta\Big[\sum_{k=0}^{\tau-1}\beta^k v_{xyt,k}\Big] + (1-\delta)\mathbb{E}_t\Big[\beta U_{y,t+1} - \beta^\tau(1-\eta_{xy,t+\tau})U_{y,t+\tau} - \beta^{\tau+1}\eta_{xy,t+\tau}U_{y,t+\tau+1}\Big] \tag{A-17}$$

Finally, rearranging the above equation produces the result:

$$p_{xyt} = (1-\delta)z_{xyt} + \delta\left[\frac{\sum_{k=0}^{\tau-1}\beta^k v_{xyt,k}}{\sum_{k=0}^{\tau-1}\beta^k}\right] \tag{A-18}$$
$$= (1-\delta)z_{xyt} + \delta m_{0,y} + \delta m_{1,y}x_{\text{complexity}} + \delta\left[\frac{\sum_{k=0}^{\tau-1}\beta^k\mathbb{E}_t[g_{t+k}]}{\sum_{k=0}^{\tau-1}\beta^k}\right]x_{\text{quantity}} \tag{A-19}$$

where $z_{xyt} = \frac{1}{\sum_{s=0}^{\tau-1}\beta^s}\mathbb{E}_t\left[\beta U_{y,t+1} - \beta^\tau(1-\eta_{xy,t+\tau})U_{y,t+\tau} - \beta^{\tau+1}\eta_{xy,t+\tau}U_{y,t+\tau+1}\right]$.

B.4 Bargaining parameter equation

I explain the steps to compute the bargaining parameter. I first drop time subscripts t and denote with a 'bar' the steady state average values in the sample. I work with the steady-state equivalent of Equation (A-19). Taking averages on both sides of this equation results in: $\bar{p} = (1-\delta)\bar{z} + \delta\bar{v}$ where $\bar{z} = \frac{1}{\sum_{s=0}^{\tau-1}\beta^s}\mathbb{E}[\beta U - \beta^\tau(1-\eta)U - \beta^{\tau+1}\eta U]$. Finally, substituting in for the average match value using $\bar{v} = \bar{p}/(1-margin)$ and rearranging leads to the result that $\delta = 1 - \frac{margin\cdot\bar{p}}{\bar{p}-(1-margin)\cdot\bar{z}}$.

I calibrate the bargaining parameter from data from the year 2005 which is a period where the industry is approximately at its long-run steady-state. To do this I then obtain margins from the 2005 annual reports of the three largest non-major oil and gas companies operating in the Gulf of Mexico (I do not use the majors since it is difficult to disentangle their deepwater vs shallow water operations). I then compute the above equation for each contract in 2005 and then take the average to recover the bargaining parameter.

B.5 Details about identification in Section 4.3

The aim is to show that the targeting parameters γ_0, γ_1, the matching efficiency parameters a_y, the distribution of potential projects f_x, and the potential project draw parameters k_0, k_1, are identified. The observed distribution of projects that type y capital matches with is:

$$\tilde{f}_{xyt} = \begin{cases} q_y^{capital}(\theta_{yt})\frac{\omega_{xyt}e_{xt}f_x}{\int_z \omega_{zyt}e_{zt}f_z dz} & \text{if } x \in A_{yt} \\ 1 - q_y^{capital}(\theta_{yt})\frac{\int_{z\in A_{yt}} \omega_{zyt}e_{zt}f_z dz}{\int_z \omega_{zyt}e_{zt}f_z dz} & \text{if } x = \emptyset \end{cases} \tag{A-20}$$

where $x = \emptyset$ corresponds to the capital being unmatched. At this stage the probability of projects meeting a match $q_y^{project}(\theta_{yt})$, the probability of capital meeting a match $q_y^{capital}(\theta_{yt})$, and the weights ω_{xyt}, are not known.

Note that I previously showed that the value function for capital can be constructed from the data. Similarly, as discussed in the main text, the match value parameters can be identified from prices. Therefore I assume that the value of a match to a project, W_{xyt}, as well as acceptance sets, are known for this proof. Furthermore, I assume that the entry cost c is calibrated from external data.

I complete the proof in several steps:

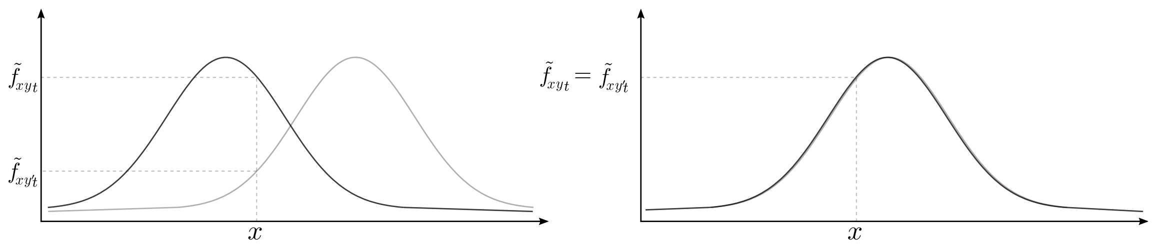
1. Identify the targeting weights ω_{xyt}
2. Identify the targeting parameter γ_0
3. Identify the matching efficiency parameters a_y
4. Identify the distribution of potential projects f_x and the parameters that underlie the potential project draws k_0, k_1 where $K_t = k_0 + k_1 g_t$
5. Identify the targeting parameter γ_1

Part 1 I show how the targeting weights ω_{xyt} can be identified. I begin with some intuition about how the targeting parameter γ_0 can be identified from observed matches. Consider Figure A-2. Both panels show the observed distributions of matches within a period conditional on a capital type y: $\tilde{f}_{xyt}$.

Under random search where $\gamma_0 = 0$ (the right panel), for a given project type x, different types of capital y, y' should have the same probability of matching: $\tilde{f}_{xyt} = \tilde{f}_{xy't}$. Under partially directed search (the left

Figure A-2: Identification of the targeting parameter $\gamma_0 > 0$

(a) Matches, $\gamma_0 > 0$ (b) Matches, $\gamma_0 = 0$



Note: This figure gives intuition about how the targeting parameter γ_0 can be identified. Further explanation is in the text.

panel), the probability that a particular project matches with different types of capital may not be the same: $\tilde{f}_{xyt} \neq \tilde{f}_{xy't}$.

To make this intuition more formal, start by rewriting the equation for the targeting weights:

$$\omega_{xyt} = \frac{n_{yt} \exp\left(\gamma_0 \left[\pi_{xyt} - \gamma_1 1\left[x \notin A_{yt}\right]\right]\right)}{\sum_{k \in Y} n_{kt} \exp\left(\gamma_0 \left[\pi_{xkt} - \gamma_1 1\left[x \notin A_{kt}\right]\right]\right)} \tag{A-21}$$

Recall that the acceptance sets are known at this stage. Consider one project type x that will accept a match with any capital type (i.e. $x \in A_{yt}$ for each $y \in \{\text{low}, \text{mid}, \text{high}\}$ and therefore $1\left[x \notin A_{yt}\right] = 0$). Comparing the probability of the same type of project x matching conditional on different capital y, y' at t:

$$\begin{aligned}
\ln\left(\tilde{f}_{xyt}\right) - \ln\left(\tilde{f}_{xy't}\right) &= \ln\left(\omega_{xyt}/\omega_{xy't}\right) \\
&\quad + \ln\left(q_y^{capital}(\theta_{yt})/q_{y'}^{capital}(\theta_{y't})\right) \\
&\quad + \ln\left(\int_z \omega_{zy't} e_{zt} f_z dz \Big/ \int_z \omega_{zyt} e_{zt} f_z dz\right) \qquad \text{(A-22)} \\
&= \gamma_0\left(\pi_{xyt} - \pi_{xy't}\right) \\
&\quad + \ln\left(n_{yt}/n_{y't}\right) \\
&\quad + \ln\left(q_y^{capital}(\theta_{yt})/q_{y'}^{capital}(\theta_{y't})\right) \\
&\quad + \ln\left(\int_z \omega_{zy't} e_{zt} f_z dz \Big/ \int_z \omega_{zyt} e_{zt} f_z dz\right) \qquad \text{(A-23)}
\end{aligned}$$

Differencing Equation (A-23) over two points x and x':

$$\left(\ln\left(\tilde{f}_{xyt}\right) - \ln\left(\tilde{f}_{xy't}\right)\right) - \left(\ln\left(\tilde{f}_{x'yt}\right) - \ln\left(\tilde{f}_{x'y't}\right)\right)$$

$$= \gamma_0\Big(\pi_{xyt} - \pi_{xy't} - \big(\pi_{x'yt} - \pi_{x'y't}\big)\Big) \tag{A-24}$$

$$= \gamma_0 q_y^{project}(\theta_{yt}) \cdot \Big(W_{xyt} - W_{x'yt}\Big) - \gamma_0 q_{y'}^{project}(\theta_{y't}) \cdot \Big(W_{xy't} - W_{x'y't}\Big) \tag{A-25}$$

Here the second equality follows from substituting the expression:

$$\pi_{xyt} = q_y^{project}(\theta_{yt}) W_{xyt} \tag{A-26}$$

In Equation (A-25) the left-hand-side is data and the W_{xyt} terms are known. Therefore, I can identify $\gamma_0 q_y^{project}(\theta_{yt})$ for each $y \in \{\text{low}, \text{mid}, \text{high}\}$. Hence, $\gamma_0 \pi_{xyt}$ can be constructed for any match (x, y) using Equation (A-26) and $\gamma_0 q_y^{project}(\theta_{yt})$. Finally, I can recover the weights ω_{xyt} because they are just a function of $\gamma_0 \pi_{xyt}$ (for the case here where all rigs accept a match with a well of type x).

Part 2: Next I show how γ_0 can be identified from the data. Denote $\widehat{q_y}^{project}(\theta_{yt}) = \gamma_0 q_y^{project}(\theta_{yt})$. As in Part 1, continue to consider one project type x that will accept a match with any capital type. Denote the expected value from entering as $(1/\gamma_0) g_{xt} = (1/\gamma_0) \sum_{k \in Y} \omega_{xkt} W_{xkt} \widehat{q_k}^{project}(\theta_{kt})$. Rearranging the entry condition:

$$e_{xt} = \frac{\exp\left(\sum_{k\in Y} \omega_{xkt}\pi_{xkt} - c\right)}{1 + \exp\left(\sum_{k\in Y} \omega_{xkt}\pi_{xkt} - c\right)} = \frac{\exp\Big((1/\gamma_0) g_{xt} - c\Big)}{1 + \exp\Big((1/\gamma_0) g_{xt} - c\Big)} \tag{A-27}$$

Note that g_{xt} can be constructed from the data and also note that I calibrate the entry cost c using external data. Therefore, all the components of Equation (A-27) are known with the exception of the parameter γ_0.

Next, consider the relative probability of two projects x and x' entering at different time periods t and t'. Differencing Equation (A-20) over these two points x and x' and over the two time periods, and rearranging:

$$\left(\ln\left(\tilde{f}_{xyt}/\omega_{xyt}\right) - \ln\left(\tilde{f}_{x'yt}/\omega_{x'yt}\right)\right) - \left(\ln\left(\tilde{f}_{xyt'}/\omega_{xyt'}\right) - \ln\left(\tilde{f}_{x'yt'}/\omega_{x'yt'}\right)\right)$$

$$= \ln\left(\frac{e_{xt}/e_{x't}}{e_{xt'}/e_{x't'}}\right) \tag{A-28}$$

$$= (1/\gamma_0)\big((g_{xt} - g_{x't}) - (g_{xt'} - g_{x't'})\big) + \ln\left(\frac{1 + \exp((1/\gamma_0)g_{x't} - c)}{1 + \exp((1/\gamma_0)g_{xt} - c)} \cdot \frac{1 + \exp((1/\gamma_0)g_{xt'} - c)}{1 + \exp((1/\gamma_0)g_{x't'} - c)}\right) \tag{A-29}$$

Here, the left-hand-side can be constructed from the data: the $\tilde{f}_{xyt}$ are directly observed and the weight terms ω_{xyt} are identified using Part 1. Furthermore, since the right-hand-side is known up to the parameter γ_0, the parameter γ_0 is identified. Therefore, the $q_y^{project}(\theta_{yt})$ are also identified.

Part 3: Suppose that there is a period in the data where the acceptance sets for capital type y is all wells x (that is, suppose no matches are rejected). Then, $q_y^{capital}(\theta_{yt})$ is directly observed in the data since it is the probability of capital type y matching. Since $\theta_{yt} = q_y^{project}(\theta_{yt})/q_y^{capital}(\theta_{yt})$, and $q_y^{capital}(\theta_{yt})$ is directly observed in the data, the market tightness terms θ_{yt} are also identified. Therefore, the matching efficiency terms a_y are also identified, through the effect of variation in θ_{yt} on the (observed) probability of capital matching $q_y^{capital}(\theta_{yt})$.

Part 4: The distribution of potential projects f_x can be identified using the time periods in Part 3 where no matches are rejected. Concretely, using these time periods, f_x can be identified by inverting the distribution of observed matches $\tilde{f}_{xyt}$ through the targeting weights and the entry condition. Similarly, in these periods the number of potential well draws K_t can be identified because the market tightness θ_{yt} is known and $\theta_{yt} = \frac{n_{yt}}{K_t \cdot \int_z \omega_{zyt} e_{zt} f_z dz}$, where the number of available capital n_{yt} is known from the data and $\displaystyle\int_z \omega_{zyt} e_{zt} f_z dz$ is identified from the previous steps. The potential well draw parameters (k_0 and k_1) can then be recovered by exploiting variation of K_t across time.

Part 5: Finally, there is one parameter remaining to identify: the targeting parameter γ_1. This parameter is identified by ensuring that capital's probability of matching is not too low in periods where the acceptance sets are narrow (that is, periods where many potential matches are rejected).

C Computation

C.1 Algorithm: computing instances of mismatch in the data

I use the following algorithm to compute Table 3:

1. Start at the first period (January 2000). Take the empirical number of available rigs of each type and match them optimally to the set of matched wells using a linear sum assignment algorithm. To compute the optimal matches use the match values that are later computed in the model.
2. Update the number of available rigs in the next period:
   - For each unemployed rig that is allocated a new match, remove one unemployed rig from the empirical number of available rigs of that type, for the duration of the match.

- For each employed rig that is now unemployed, add one unemployed rig to the empirical number of available rigs of that type, for the duration of the contract.

3. Update to the next period and return to 1.

C.2 Computing capital's value of searching U_{yt}

I estimate the empirical objects used to construct the value functions in Section 4.2 as follows.

Prices To estimate $\bar{p}_{\tau=n,yt}$ for each $y \in \{low, mid, high\}$ I regress observed prices on first and second order polynomial combinations of the state vector plus an extra term for the contract length.

Extension probability I estimate $\bar{\eta}_{\tau=n,yt}$ (the average probability of extending a τ length contract given the state at time t for capital type y) as follows. For each capital type $y \in \{low, mid, high\}$ I estimate a logit model for whether a contract is extended on first and second order polynomial combinations of the state vector plus an extra term for the contract length.

The probability of matching To estimate $\mathbb{P}_{\tau=n,yt}$ I estimate a separate multinomial logit equation for each $y \in \{low, mid, high\}$. The dependent variable alternatives are matching with contract lengths $\tau \in \{0, 2, 3, 4\}$ where $\tau = 0$ denotes the probability of not matching. The independent variables are first and second order polynomial combinations of the state vector.

Algorithm The algorithm that I use is based on the results where I show how U_{yt} can be constructed from the data in Appendix B.2.

Overall, I use forward simulation to compute the value function over a set of points (nodes). I linearly interpolate the value of searching over the state space using a grid with 10 nodes for the gas price dimension and 5 nodes for each available capital state dimension (so there are $5^3 \times 10 = 1250$ grid nodes in total). I perform the algorithm separately for each $y \in \{\text{low}, \text{mid}, \text{high}\}$. For each node s in the state space grid, repeat the following forward simulation algorithm:[3]

1. Initialize the state as the state corresponding to the node.
2. Given the state, draw a contract from the estimated multinomial logit for the probability of matching. (Note that this incorporates the possibility that the rig will be unmatched i.e. draw a 0 length contract.)

[3] I end the algorithm after 1200 periods have elapsed. Since a time that is k periods in the future is discounted by β^k, periods in the later values of the forward simulation are given very low weight.

3. If a contract is drawn:
   (a) Get the corresponding price using the empirical object estimated for prices.
   (b) Update the state; count down the number of periods remaining on the contract by 1.
   (c) If the number of periods left on the contract is 0, simulate the probability of extension using the empirical object computed above. If the contract is extended, begin again from Part (a).[4]
   (d) If an extension does not occur: return to 2.
4. If a contract is not drawn: update the state and return to 2.

I repeat the above forward simulation algorithm 200 times. I then take the average value of these 200 simulations to compute U_{yt}.

C.3 Computing the match surplus

When simulating the model, I compute the match surplus S_{xyt} using forward simulation. I explain the details in this section.

Extension probability I estimate η_{xyt} as follows.[5] For each capital type $y \in \{low, mid, high\}$ I estimate a logit model for whether a contract is extended on first and second order polynomial combinations of the state vector, as well as two extra terms: one for the contract length and one for well complexity.

Algorithm: overview Overall, I use forward simulation to compute the surplus over a set of points (nodes). I linearly interpolate the surplus using a grid with 10 nodes for the gas price dimension, 5 nodes for each available capital state dimension, and 5 nodes for the well complexity dimension (so there are $5^3 \times 10 \times 5 = 6250$ grid nodes in total).[6] I perform the algorithm separately for each $y \in \{\text{low}, \text{mid}, \text{high}\}$

[4] Since this step occurs after the state is updated, it is consistent with the timing outlined in the model section that the extensions occur first. Also recall that the extension will be of the same match (including the contract duration) as the original contract.

[5] When computing the rig's value of searching recall that I computed a different object for extensions based on the average probability of extending a contract (where the average is taken over the distribution of potential matches at a future state in time). In computing the match surplus the exact (x,y) match is fixed. Therefore, for computational simplicity, I estimate the probability of extending a contract conditional on a particular match.

[6] By assumption, in the model the quantity of hydrocarbons $x_{quantity}$ is assumed to by a polynomial function of the well complexity variable $x_{complexity}$. Therefore, given this polynomial function, I can compute $x_{quantity}$ from a particular $x_{complexity}$. That is, I do not need to include $x_{quantity}$ as an additional dimension to be interpolated over.

and for each contract length $\tau \in \{2, 3, 4\}$.[7]

When searching over the objective function in the simulated method of moments, I need to quickly compute the match surplus for different candidate parameter values of $m_{0,y}, m_{1,y}, m_2, \rho_0, \rho_1, \rho_2, \rho_3$. The algorithm I set out below exploits that the match surplus is a linear function of these parameters and that the nodes of the surplus interpolation grid are fixed. Specifically, the algorithm computes the coefficients on these parameters in the match surplus for each node. The computational benefit is that at a given node in the interpolation grid, I only need to perform the forward simulation algorithm once.

The coefficients on the parameters that I simulate can be derived in the following way:

$$\begin{aligned} S_{xyt} &= W_{xyt} + V_{xyt} - \beta \mathbb{E}_t U_{y,t+1} \\ &= \sum_{k=0}^{\tau-1} \beta^k v_{xyt,k} + \beta^\tau \mathbb{E}_t \Big[\eta_{xy,t+\tau} \Big(\sum_{k=0}^{\tau-1} \beta^k v_{xy,t+\tau,k} + ... \Big) + (1 - \eta_{xy,t+\tau}) U_{y,t+\tau} \Big] - \beta \mathbb{E}_t U_{y,t+1} \end{aligned}$$

and therefore:

$$S_{xyt} = m_{0,y} b_0 + m_{1,y} b_1 + m_2 \cdot (\rho_0 b_2 + \rho_1 b_3 + \rho_2 b_4 + \rho_3 b_5) + b_6 - \beta \cdot b_7$$

where the coefficients are:

$$\begin{aligned} b_0 &= \mathbb{E}_t \left[\sum_{k=0}^{\tau-1} \beta^k + \eta_{xy,t+\tau} \sum_{k=0}^{\tau-1} \beta^{\tau+k} + \eta_{xy,t+\tau} \eta_{xy,t+2\tau} \sum_{k=0}^{\tau-1} \beta^{2\tau+k} + ... \right] \\ b_1 &= \mathbb{E}_t \left[\sum_{k=0}^{\tau-1} \beta^k + \eta_{xy,t+\tau} \sum_{k=0}^{\tau-1} \beta^{\tau+k} + \eta_{xy,t+\tau} \eta_{xy,t+2\tau} \sum_{k=0}^{\tau-1} \beta^{2\tau+k} + ... \right] \cdot x_{complexity} \\ b_2 &= \mathbb{E}_t \left[\sum_{k=0}^{\tau-1} \beta^k g_{t+k} + \eta_{xy,t+\tau} \sum_{k=0}^{\tau-1} \beta^{\tau+k} g_{t+\tau+k} + \eta_{xy,t+\tau} \eta_{xy,t+2\tau} \sum_{k=0}^{\tau-1} \beta^{2\tau+k} g_{t+2\tau+k} + ... \right] \\ b_3 &= \mathbb{E}_t \left[\sum_{k=0}^{\tau-1} \beta^k g_{t+k} + \eta_{xy,t+\tau} \sum_{k=0}^{\tau-1} \beta^{\tau+k} g_{t+\tau+k} + \eta_{xy,t+\tau} \eta_{xy,t+2\tau} \sum_{k=0}^{\tau-1} \beta^{2\tau+k} g_{t+2\tau+k} + ... \right] \cdot x_{complexity} \\ b_4 &= \mathbb{E}_t \left[\sum_{k=0}^{\tau-1} \beta^k g_{t+k} + \eta_{xy,t+\tau} \sum_{k=0}^{\tau-1} \beta^{\tau+k} g_{t+\tau+k} + \eta_{xy,t+\tau} \eta_{xy,t+2\tau} \sum_{k=0}^{\tau-1} \beta^{2\tau+k} g_{t+2\tau+k} + ... \right] \cdot x^2_{complexity} \\ b_5 &= \mathbb{E}_t \left[\sum_{k=0}^{\tau-1} \beta^k g_{t+k} + \eta_{xy,t+\tau} \sum_{k=0}^{\tau-1} \beta^{\tau+k} g_{t+\tau+k} + \eta_{xy,t+\tau} \eta_{xy,t+2\tau} \sum_{k=0}^{\tau-1} \beta^{2\tau+k} g_{t+2\tau+k} + ... \right] \cdot x^3_{complexity} \\ b_6 &= \mathbb{E}_t \Big[(1 - \eta_{xy,t+\tau}) U_{y,t+\tau} + \eta_{xy,t+\tau} (1 - \eta_{xy,t+2\tau}) U_{y,t+2\tau} + ... \Big] \\ b_7 &= \mathbb{E}_t U_{y,t+1} \end{aligned}$$

[7] Therefore, there are 9 different components that I compute for surplus. Later, when I need to compute the surplus at a particular match, the code first chooses the component relating to the capital type and contract length of the match, and then linear interpolates surplus over the continuous variables i.e. the state and well complexity.

Algorithm: computing the surplus at each node For each node s in the state space grid, I compute b_j for $j \in \{0, 1, ..., 7\}$ by forward simulation. I take the average value of each b_j over 200 simulations. For each simulation, I stop the algorithm after 120 periods have elapsed. Since the values of later periods in the surplus equation are weighted by the probability that the contract has been extended until that period, the simulation error from ending the algorithm after 120 periods is extremely low.

C.4 Computing prices

Algorithm: overview Similar to the surplus computation above, when estimating the model I need to quickly compute prices for different values of the parameters. In order to do this, I use Equation (A-19) which shows how prices can be written as a function of the bargaining parameter, capital's value of searching, and the value of a match. For every contract in the data, I simulate the following components:

$$
\begin{aligned}
z_{xyt} &= \frac{1}{\sum_{s=0}^{\tau-1} \beta^s} \mathbb{E}_t \left[\beta U_{y,t+1} - \beta^{\tau}(1-\eta_{xy,t+\tau}) U_{y,t+\tau} - \beta^{\tau+1} \eta_{xy,t+\tau} U_{y,t+\tau+1} \right] \\
b_0 &= \frac{1}{\sum_{s=0}^{\tau-1} \beta^s} \cdot \sum_{k=0}^{\tau-1} \beta^k \\
b_1 &= \frac{1}{\sum_{s=0}^{\tau-1} \beta^s} \cdot \sum_{k=0}^{\tau-1} \beta^k \cdot x_{complexity} \\
b_2 &= \frac{1}{\sum_{s=0}^{\tau-1} \beta^s} \cdot \sum_{k=0}^{\tau-1} \beta^k \cdot \mathbb{E}_t[g_{t+k}] x_{quantity}
\end{aligned}
$$

and, therefore, prices can be constructed in the following way (recalling that the bargaining parameter δ is calibrated):

$$
p_{xyt} = (1-\delta) z_{xyt} + \delta(m_{0,y} b_0 + m_{1,y} b_1 + m_2 b_2) \tag{A-30}
$$

Therefore, when estimating the model, the algorithm requires simulating the components z_{xyt}, b_0, b_1, b_2 only once for a given match. I use 200 simulations to compute these components. Prices can then be quickly computed using Equation (A-30) for different candidate parameters $m_{0,y}, m_{1,y}, m_2$.

C.5 Algorithm: simulated method of moments

In this section I describe the algorithm used to simulate the market which I use to estimate the parameters in the simulated method of moments. I first describe how to compute the equilibrium within each period (a month) for a given set of the parameters. I then describe how to nest the computation of the per-period equilibrium to simulate the market over the period 2000-2009.

C.5.1 Computing the per-period equilibrium

The state that agents take into account when computing their value functions is $s_t = (g_t, n_{low,t}, n_{mid,t}, n_{high,t})$ where n_{yt} denotes the number of rigs of type y that are available to match the following period. Given s_t the per-period equilibrium is computed using the following algorithm:

1. Guess the share of potential well draws $K_t = k_0 + k_1 g_t$ that choose to enter and target a rig of type y. Denote this share as $share^i_{yt}$ where the share of wells that do not enter is $1 - \sum_{k\in\{low,mid,high\}} share^i_{kt}$. The variable i denotes the iteration (so the guess initializes at $i = 0$).
2. Get the submarket tightness θ^i_{yt} using:

$$\theta^i_{yt} = \frac{n_{yt}}{share^i_{yt} \cdot K_t} \tag{A-31}$$

   and note that this pins down the expected surplus of project type x to searching in the type y capital submarket:

$$\pi^i_{xyt} = q^{project}_y(\theta^i_{yt})W_{xyt} \tag{A-32}$$

3. Update the shares using:

$$share^{i+1}_y = \int_z \omega^i_{zyt} e^i_{zt} f_z dz \tag{A-33}$$

   where the weights are defined using Equation (5):

$$\omega^i_{xyt} = \frac{n_{yt} \exp\left(\gamma_0 \left[\pi^i_{xyt} - \gamma_1 1\left[x \notin A_{yt}\right]\right]\right)}{\sum\limits_{k\in Y} n_{kt} \exp\left(\gamma_0 \left[\pi^i_{xkt} - \gamma_1 1\left[x \notin A_{kt}\right]\right]\right)} \tag{A-34}$$

   and the entry condition is defined using Equation (6):

$$e^i_{xt} = \frac{\exp\left(\sum_{k\in Y} \omega^i_{xkt}\pi^i_{xkt} - c\right)}{1 + \exp\left(\sum_{k\in Y} \omega^i_{xkt}\pi^i_{xkt} - c\right)} \tag{A-35}$$

4. Repeat steps 2-3 until the targeting shares converge. Denote the equilibrium shares as $share^*_{yt}$, the equilibrium weights as ω^*_{xyt}, and the equilibrium probability of entry as e^*_{xt}.
5. The distribution and total number of matches (amongst other things) can now be computed from this targeting equilibrium and the acceptance sets. Recall that the acceptance set is the set of projects x where the match surplus with capital type y is positive at the time t state. To simplify the computation I use the empirical state at time t (rather than the model predicted state) to compute the surplus and the acceptance sets. In addition, since some components of the empirical state can be noisy (e.g. the number of available rigs of each type), I smooth the empirical state using a local polynomial regression.

C.5.2 Computing the market over the 2000-2009 period

I nest the preceding algorithm to compute the market over the 2000-2009 period. I use the empirical evolution of the gas price g_t (since this is assumed to be exogenous) but I update the number of available rigs in accordance with the model equilibrium. To do this I introduce a 'detailed state' which is the current natural gas price g_t and, for each capital type y, a distribution of current matches.

For the distribution of current matches, I discretize $x_{complexity}$ into 30 bins. In addition, each match also has a corresponding contract length τ, and a number of remaining periods on the contract. So, for example, the number of possible detailed states is divided into $30 \times 2 = 60$ bins for $\tau = 2$ contracts, 90 bins for $\tau = 3$ contracts, and 120 bins for $\tau = 4$ contracts.

Note that agents' expectations about future states (and their corresponding value functions) are computed over the state $s_t = (g_t, n_{low,t}, n_{mid,t}, n_{high,t})$ rather than the detailed state. Therefore, the value functions are not subject to a curse of dimensionality.

I compute the market as follows (starting from a guess of the detailed state for January 2000 - which includes the empirical natural gas price - where I first burn-in the simulation).

1. Compute the probability that each match is extended and update the 'detailed state' with these contract extensions.
2. Compute the equilibrium at the current state using the per-period equilibrium algorithm described in the preceding section.
3. For this equilibrium, compute the equilibrium probabilities of each type of rig matching with each duration contract. Use these equilibrium values to update the 'detailed state'.
4. Repeat Steps 2-4 over the monthly natural gas price evolution from January 2000 - December 2009.

C.5.3 Other implementation details for the simulated method of moments

I stack the simulated moments: $m_s(\boldsymbol{\lambda})$, where I denote $\boldsymbol{\lambda}$ as the vector of parameters to be estimated. I fit the simulated moments m_s to the empirical moments m_d by minimizing the following objective function (denoting Ω as the weighting matrix): $(m_d - m_s(\boldsymbol{\lambda}))'\Omega(m_d - m_s(\boldsymbol{\lambda}))$.

I set the weighting matrix Ω to be a diagonal matrix. I set the weights for the medium-specification rig matches, as well as the low- and mid-specification mean utilization moments, to 0.1. I set the 2006 high-specification utilization moment weight to 0.01. I set the utilization variance for the high-specification rig to 10, and the moment on the coefficient β_3 to 100. I set all the remaining weights to 1.[8]

[8] These weights are chosen to ensure that the model closely replicates moments such as the sorting patterns, which are of primary interest, as well as to ensure the moments have a similar scale.

C.6 Algorithm: no sorting counterfactual

Note that no extra computation needs to be performed for the no sorting counterfactual. This is because the three components that depend on the value function in the model are constrained to not depend on the value function in the counterfactual. These three components are as follows. First, the entry decision is assumed to be the same in the no sorting equilibrium as in the market baseline (this is to avoid compositional effects). Second, the targeting parameters are set so that $\gamma_0 = \gamma_1 = 0$ which implies that projects do not take into account value functions when making their targeting decisions. Finally, acceptance sets are computed as if the future value of searching $E_t U_{y,t+k} = 0$ for $k \in \{1, 2, ...\}$ (that is, agents do not reject matches due to high future outside options).

C.7 Algorithm: intermediary

In this section I describe the algorithm I use to compute the equilibrium with an intermediary. Given that the match value is supermodular, positive assortative matching is optimal in a static model. Motivated by this idea, I look for cutoff solutions in the following form: one cutoff $\bar{x}_{complexity}$ where all wells where $x_{complexity} > \bar{x}_{complexity}$ are assigned to target high-specification rigs, and another cutoff $\underline{\mathrm{x}}_{complexity}$ where if $\bar{x}_{complexity} \geq x_{complexity} > \underline{\mathrm{x}}_{complexity}$ then the well is assigned to target a mid-specification rig, and if $x_{complexity} \leq \underline{\mathrm{x}}_{complexity}$ then the well is assigned to target a low-specification rig. Finally, I assume that if a rig and well meet under the intermediary's protocol then the match will be accepted, and also that matches will be extended randomly (i.e. with probability η).[9] Note that since I consider a myopic algorithm, as well as also keeping the composition of entered wells the same as in the baseline model and assuming that matches will be accepted, I do not need to re-solve for the value functions or agents' beliefs over the state evolution in this counterfactual.

An alternative algorithm would be an intermediary which also internalizes dynamic considerations. I choose to not implement this alternative for several reasons. First, in the myopic algorithm, rigs are already strongly assortatively matched. This blunts the incentives behind the sorting effect as there is less of a benefit to waiting for a better match. Second, the greedy matching algorithm does not require re-solving for value functions and beliefs that are changing across time and so is not computationally intensive. Third the matching protocol is relatively simple and so could be feasibly implemented by a real-world intermediary. Finally, to the extent that there are additional benefits from incorporating dynamic considerations, this simple greedy algorithm can be interpreted as a lower bound on the potential benefits of an intermediary.

The algorithm proceeds as follows *within* each period:

[9] The rationale for this assumption - as well as computational tractability - is that since the intermediary's allocation results in high quality matches, there is little incentive to reject a match (or to not extend a match) in order to wait for a better match.

1. Given a set of entered wells and available rigs, and potential cutoffs ($\underline{x}_{complexity}, \bar{x}_{complexity}$), allocate wells to rig submarkets.

2. Given the above allocations, compute the total expected value of the (static) matches.

3. Repeat Steps 1.-2. and recompute over all potential cutoffs.

4. Given the above steps, choose the solution which maximizes the total expected value of the (static) matches.

Then, using the above solution, I update the state of the market and move to the next period where I run the above algorithm again.

C.8 Algorithm: demand smoothing

In this section I detail the algorithm I use to compute the demand smoothing counterfactual. I need to recompute value functions. In addition, I also need to recompute agents' beliefs about the evolution of the state space (since these beliefs were computed when estimating the model from the empirical state evolution which will change in the counterfactual). Since by assumption in this counterfactual the state evolution is in a steady state at the mean natural gas price over 2000-2009 ($\bar{g}$), the transitions are in the following form:

$$R_0 = \begin{bmatrix} \bar{g} \\ \bar{n}_{low} \\ \bar{n}_{mid} \\ \bar{n}_{high} \end{bmatrix}, \quad R_1 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \sigma_\epsilon = 0 \tag{A-36}$$

In the above equations the mean number of available rigs of each type $\bar{n}_{low}, \bar{n}_{mid}, \bar{n}_{high}$ will change. Therefore, these components need to be recomputed in the algorithm, amongst other things.

Overall, the algorithm can be viewed as featuring an inner loop and outer loop. In the inner loop, I recompute value functions and re-simulate the model using the same algorithm as the original model. In the outer loop I treat the results of the re-simulated model as 'data'. From these 'data' I recompute the state evolution beliefs and other empirical objects used to construct value functions, and continue to iterate until convergence.

1. Denote the state evolution beliefs in the j-th outer loop iteration as R_0^j, R_1^j. Initialize R_0^0, R_1^0.

2. Initialize a guess of the objects used to compute the value of searching: the average prices for each capital type, the average probability of matching for each capital type, and the average probability of extending a contract for each capital type.[10]

[10]When implementing this procedure for computational simplicity I fix the probability of extending the contract as η. Note that this restriction is made without loss of generality. Since the market is in steady state the surplus

3. Denote the value of searching in the j-th iteration as U_y^j. Construct this value of searching for each capital type y using the current guess of the objects used to compute the value of searching using the procedure in Section C.2.

4. Using U_y^j, simulate the model using the algorithm in Section C.5.

5. From the simulated model, recompute the objects used to compute the value of searching (the average prices for each capital type, the average probability of matching for each capital type, and the average probability of extending a contract for each capital type) and update the state evolution beliefs R_0^j, R_1^j.

6. Check for convergence.[11] If the algorithm has not converged, repeat from Step 3.

D Additional Graphs, Tables, and Results

D.1 Why do I assume the extension is the same length as the initial contract?

I assume that an extended contract is the same length as the original contract because this is typically the case in the data. Concretely, using the aggregated measures of duration that I use in the model ($\tau \in \{2, 3, 4\}$), I test how often the duration changes between subsequent extended contracts. I find that the probability that the duration remains the same across extended contracts is 64.8% . Furthermore, since the majority of contracts are new contracts, extended contracts with a different duration from the original contract represent only 12.2% of all contracts.

D.2 Do rigs change ranking between contracts?

In order to answer this question I split the data up into two five-year periods: January 2000 to December 2004 and January 2005 to December 2009. I begin by de-meaning the price by the average in each month (to remove price cyclicality). Using the de-meaned average prices I then rank rigs into three categories (low, medium, and high) in each of the five-year periods for the rigs that are present in both periods. I then investigate the extent to which rig rankings stay constant between the two periods.

I find that rig ranking remains remarkably constant throughout the two time periods, with 84.6% of

of a match is fixed over time. Therefore, if a match is initially accepted, the match surplus is positive and so will also be positive at the time when the contract is extended. Hence, each initial match is extended with probability $\eta_{xyt} = \eta 1[x \in A_{yt}] = \eta$.

[11] My metric for convergence is the maximum absolute difference in the average prices. I define convergence as whether this metric is < 0.001.

rigs maintaining their ranking across periods.[12] In terms of the small fraction of rigs that change their category under this metric, note that average prices are a noisy measure of rig type since they are also dependent on the well that the rig is matched to. In addition, the average price for each rig in each period might only be computed from a few contracts.

D.3 Synergies between high efficiency rigs and high complexity wells during boom periods

I show that there are larger synergies between high efficiency rigs and high complexity wells during boom periods, as reflected in prices. To do so I run the following regression at the contract level (for a match between well i and rig j at time t):

$$\begin{aligned} price_{ijt} = \beta_{0,y_j} + \beta_{1,y_j} 1[\text{High-complexity}_i] + \beta_{2,y_j} g_t + \beta_{3,low} 1[\text{High-complexity}_i] g_t \\ + \beta_{3,y_j} 1[y_j \in \{\text{mid}, \text{high}\}] 1[\text{High-complexity}_i] g_t + \epsilon_{ijt} \end{aligned} \quad \text{(A-37)}$$

In Equation (A-37), $1[\text{High-complexity}_i]$ is an indicator for whether the well complexity for well i is in the top 10% of complexity. The component g_t is the natural gas price. The coefficients are dependent on the type of rig $y_j \in \{low, mid, high\}$. The component $\beta_{3,low}$ is included in all regressions, while if $y_j \in \{mid, high\}$ then the components $\beta_{3,mid}$ or $\beta_{3,high}$ are additionally included.

Equation (A-37) captures how high-specification rig and complex well synergies change in busts vs booms through the coefficent $\beta_{3,high}$. I report regressions based on Equation (A-37) in Table A-2. In column (1) I report the value of $\beta_{3,high}$ for an initial specification without controls. In column (2) I include project controls (a control for the value of hydrocarbons) and in column (3) I additionally include contract-type controls (contract duration and whether the contract is an extension). Overall, the results show that the price premium for a 'well-matched' high-specification rig is higher in a boom than a bust. Concretely, moving from a bust (e.g. a natural gas price of \$3) to a boom (e.g. a natural gas price of \$9) results in a price premium of 'well-matched' high-specification rigs of around \$30 thousand dollars per day. For comparison, the average price per day for a high-specification rig over the cycle is around \$70 thousand dollars per day, so the \$30 thousand dollars per day number is substantial.

[12] I also experimented with computing the rankings by splitting the data up into the boom versus the bust. This results in a similarly high number of rigs that maintain their price ranking, with 85.5% of rigs keeping the same rank.

Table A-2: Regressions of synergies

	(1) Price/day	(2) Price/day	(3) Price/day
Δ synergies in boom: complex wells/high-spec rigs	5.42***	5.44***	4.98***
	(1.22)	(1.22)	(1.26)
Project controls	No	Yes	Yes
Contract-type controls	No	No	Yes
N	2655	2655	2655

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. This regression reports the estimated $\beta_{3,high}$ from the equation: $price_{ijt} = \beta_{0,y_j} + \beta_{1,y_j} 1[\text{High-complexity}_i] + \beta_{2,y_j} g_t + \beta_{3,low} 1[\text{High-complexity}_i] g_t + \beta_{3,y_j} 1[y_j \in \{\text{mid}, \text{high}\}] 1[\text{High-complexity}_i] g_t + \epsilon_{ijt}$. (The component $\beta_{3,low}$ is included in all regressions, while if $y_j \in \{mid, high\}$ then the components $\beta_{3,mid}$ and $\beta_{3,high}$ are additionally included.) Prices are in thousands of USD per day. The Δ synergies in boom term corresponds to $\beta_{3,high}$ i.e. the change in a price per day for a 'well-matched' high-specification rig for a \$1 increase in the natural gas price. Project controls indicates a control for the value of hydrocarbons (i.e. the quantity of hydrocarbons multiplied by the current natural gas price). Contract-type controls corresponds to controls for contract duration and also whether the contract is an extension or not. Robust standard errors are in brackets.

D.4 Additional results

Table A-3: Regressions of well drilling duration on whether the market is in a boom or bust

	(1)	(2)	(3)
Dependent Variable	Duration	Duration	Duration
1[Boom]	0.14	0.08	-0.03
	(0.79)	(0.8)	(0.8)
Third order polynomial of complexity	Yes	Yes	Yes
Rig type FEs	No	Yes	Yes
Rig type FEs interacted with complexity polynomial	No	No	Yes
N	5652	5652	5652

Note: * p < 0.1, ** p < 0.05, *** p < 0.01. Drilling duration is measured in days. The variable 'boom' is an indicator for whether the market is in a boom or not (defined as the gas price being above or below average). The results illustrates that drilling speed (after controlling for the complexity of a well) does not change in booms vs busts. Robust standard errors are in brackets.

Figure A-3: Composition of matches in booms and busts

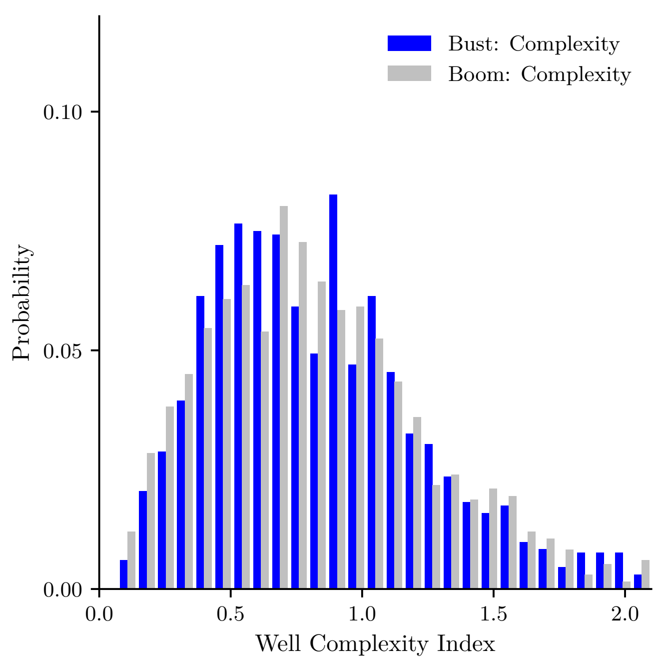


Table A-4: Regressions of sorting patterns

	(1) Complexity	(2) Complexity	(3) Complexity
1[Low-spec]	0.787***	0.79***	0.688***
	(0.02)	(0.02)	(0.038)
1[Low-spec] × 1[Boom]	-0.063**	-0.062**	-0.06**
	(0.027)	(0.027)	(0.027)
1[Mid-spec]	0.844***	0.848***	0.749***
	(0.02)	(0.02)	(0.037)
1[Mid-spec] × 1[Boom]	0.016	0.016	0.013
	(0.028)	(0.028)	(0.028)
1[High-spec]	0.9***	0.905***	0.801***
	(0.023)	(0.023)	(0.042)
1[High-spec] × 1[Boom]	0.083**	0.081**	0.086**
	(0.038)	(0.038)	(0.038)
Project controls	No	Yes	Yes
Contract-type controls	No	No	Yes
N	2655	2655	2655

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The variable 'boom' is an indicator for whether the market is in a boom or not (defined as the gas price being above or below average). Project controls indicates a control for the quantity of hydrocarbons. Contract-type controls corresponds to controls for contract duration. Robust standard errors are in brackets.

Table A-5: Utilization by rig specification and whether the market is in a boom or bust

	Bust	Boom
Utilization: low-specification rigs	0.549	0.657
Utilization: mid-specification rigs	0.661	0.804
Utilization: high-specification rigs	0.894	0.976

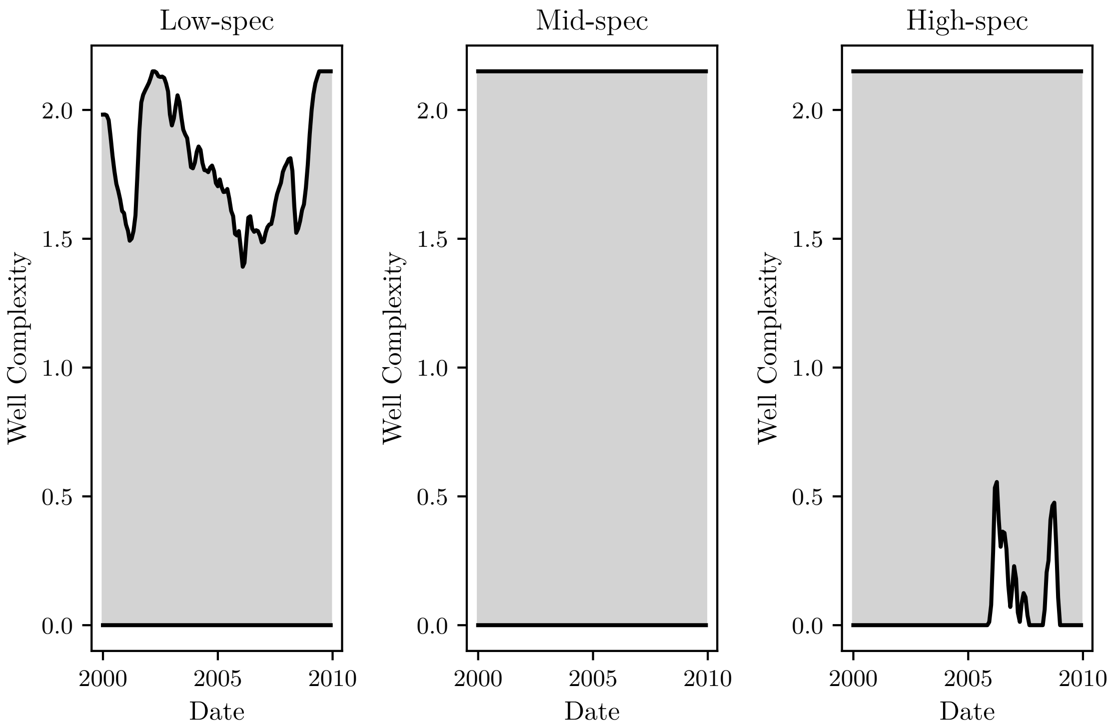
Note: The variable 'boom' is an indicator for whether the market is in a boom or not (defined as the gas price being above or below average).

Table A-6: Fit of the simulation to the moments used in the estimation

		Data	**Simulated**		**Data**	**Simulated**
Matches: Complexity				**Matches: Duration**		
Mean, Bust:	Low	0.79	0.79	Prob. τ_2	0.7	0.7
	Mid	0.84	0.87	Prob. τ_3	0.19	0.19
	High	0.9	0.9	**Price**		
Mean, Boom:	Low	0.72	0.72	$\hat{\beta}_0$	0.4	0.3
	Mid	0.86	0.87	$\hat{\beta}_{1,low}$	-0.09	-0.1
	High	0.98	0.98	$\hat{\beta}_{0,mid}$	-0.06	-0.06
Variance		0.26	0.26	$\hat{\beta}_{1,mid}$	0.02	0.003
Utilization				$\hat{\beta}_{0,high}$	-0.1	-0.1
Mean:	Low	0.6	0.61	$\hat{\beta}_{1,high}$	0.1	0.1
	Mid	0.73	0.71	$\hat{\beta}_2$	0.01	0.01
	High	0.94	0.94	Difference: High-Mid	0.1	0.2
Variance:	Low	0.032	0.0074	Difference: Mid-Low	0.07	0.09
	Mid	0.024	0.0094	**Extensions**		
	High	0.0086	0.005	Mean	0.35	0.35
Covariance:	Low	0.18	0.19			
	Mid	0.22	0.21			
	High	0.12	0.12			
Level in 2006:	High	0.97	0.94			

Note: This table contains the moments used in the simulated method of moments step. In this table I present the 'price' moments in hundreds of thousands of dollars for readability (in the model and throughout the paper I measure prices in millions of dollars). I report both the value observed in the data and the simulated moments at the optimal parameters.

Figure A-4: Acceptance sets for 2 month contracts over time



Note: This table contains the acceptance sets computed for a contract of $\tau = 2$ months. The vertical distance between the two black lines (shaded gray) is the acceptance set at a given date. I plot the acceptance sets for the three rig types computed at the empirical state value and with the minimum bound 0 and the maximum bound 2.15. Judging only by the vertical distance, the low-spec rig acceptance set appears to shrink more than for high-spec rigs in booms. However, for the high-spec rig, this rejection occurs in a region where the density of searching wells is higher. Therefore, the high-spec rig may still reject more matches in total.

D.5 Simple version of the model

I now set out a model with two types of rigs and wells, to show the intuition for the result. The main takeaway is that whether stronger sorting in a boom is optimal - and whether the model will predict stronger sorting in a boom - is ultimately an empirical question and could go either way in the theoretical model. This is because drilling a well is more valuable during boom periods because the value of the oil and gas that is extracted from the well is higher. Therefore, in theory, agents might be *less* selective in booms if the value of the project increases more than the dynamic benefits of waiting for a better match.

Simple model setup Suppose that capital can take two types $y \in \{low, high\}$ and wells can take two types $x \in \{simple, complex\}$. Assume that capital takes on each type with probability 0.5, and similarly wells take on each type with probability 0.5. Assume that contract length is $\tau = 2$, the market is in a steady state (that is, if the market is in a boom then agents expect it to remain in a boom in the following period), and that agents meet under random search.

Denote the value of a match as $\bar{\mathrm{m}}$ if it is between a simple well and a low-type capital, or between a complex well and high-type capital. Alternatively, denote the value of a match as $\underline{\mathrm{m}} < \bar{\mathrm{m}}$ if it is between a complex well and a low-type rig, or a simple well and a high-type rig. Therefore, given one rig of each type and one well of each type, without any search frictions it is optimal to allocate the complex well to the high-efficiency capital and the simple well to the low-efficiency capital. In addition to this match value agents also receive the value of hydrocarbons in the well which in this simple model I set to the gas price g (that is, $x_{quantity} = 1$).

Solution The aim is to determine under what primitives rejections of bad matches ($\underline{\mathrm{m}}$), and hence stronger sorting patterns, will occur. I begin by assuming that bad matches are rejected and find conditions on the parameters that support this equilibrium. The rig's value of searching is:[13]

$$U = 0.5qV_{\bar{\mathrm{m}}} + \Big(1 - 0.5q\Big)\beta U \tag{A-38}$$

In Equation A-38, a rig is matched with probability q and if matched then encounters a match it will accept with probability 0.5. Since only matches with value $\bar{\mathrm{m}}$ are accepted and prices are determined by Nash bargaining the payoff to matching is: $V_{\bar{\mathrm{m}}} = \delta S_{\bar{\mathrm{m}}} + \beta U$. Match surplus is $S_{\bar{\mathrm{m}}} = (1+\beta)(\bar{\mathrm{m}}+g) + \beta^2 U - \beta U$. Substituting these objects into Equation A-38 and rearranging results in:

$$U = \left(\frac{1+\beta}{1-\beta}\right)\left(\frac{0.5q\delta}{1+0.5q\delta}\right)(\bar{m}+g) \tag{A-39}$$

It is optimal to reject bad matches if $S_{\underline{\mathrm{m}}} < 0$. Substituting Equation A-39 into this surplus condition and rearranging:

$$q > \frac{\underline{\mathrm{m}} + g}{0.5\delta\beta(\bar{\mathrm{m}} - \underline{\mathrm{m}})} \tag{A-40}$$

Similarly, it can be show that if the inequality in Equation A-40 does not hold then it is optimal to accept the match $\underline{\mathrm{m}}$.

[13] Note that I avoid notation about rig type.

Comparative statics Equation A-40 allows me to perform some simple comparative statics that illustrate the main forces at work in the model. First, as the bargaining parameter δ increases or the difference between a good vs bad match $\bar{\mathrm{m}} - \underline{\mathrm{m}}$ increases, it becomes more beneficial to reject bad matches (for fixed values of q and g).

Second, to illustrate the sorting effect, consider how changing the probability of matching q and the gas price g affects the decision to reject bad matches in Equation A-40. As the market moves from a bust to a boom both q and g increase. Since both sides of Equation A-40 will increase in a boom it is an empirical question as to whether rejections of bad matches (and therefore the sorting effect) will be pro-cyclical or counter-cyclical.

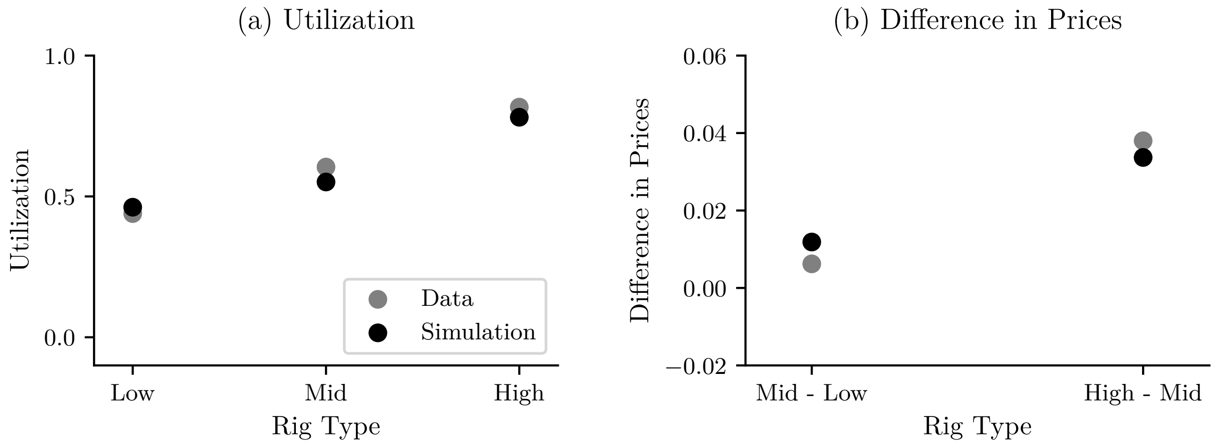
D.6 Out-of-sample fit

I evaluate out-of-sample fit in Figure A-5. To do so I run the model for the period 2010-2013, removing the dates from April to December 2010 when the shallow-water market was either in an official moratorium after the 2010 Deepwater Horizon oil spill, or a 'defacto moratorium' where no new permits were awarded.[14] This period is out-of-sample in terms of the date range. Also, the average natural gas price in this period is low (\$3.65) when compared to the sample data used for estimation (where the average natural gas price is around \$6.66). Therefore, it presents an arguably onerous out-of-sample test.

In Figure A-5 I compare two sets of simulated moments to their empirical counterparts. The first set of moments are related to mean rig utilization. These are most sensitive to parameters such as the potential project draws, and the meeting technology/targeting parameters. Even though the market is in a bust and the mean utilization is relatively low for all rig types, the model appears to fit rig utilization well in Panel (a). The second set of moments relate to the differences/orderings in prices between high and mid-specification rigs and mid and low-specification rigs. These moments are quite sensitive to the estimated match values, as well as the rig's value of searching (which enters as an outside option in the Nash bargaining equation). Again, the fit (documented in Panel (b)) appears to be quite good comparing the model to the data.

[14] For a longer discussion of the effects of the moratorium see Vreugdenhil (2021).

Figure A-5: Out-of-sample fit



Note: The out-of-sample fit is performed for the period 2010-2013, removing the dates from April to December 2010 when the shallow-water market was either in an official moratorium after the 2010 Deepwater Horizon oil spill, or a ‘defacto moratorium’ where no new permits were awarded.

E Robustness Exercises

E.1 Robustness to instances of well deepening

I investigate how widespread instances of well deepening are, and also whether it is affecting the sorting patterns in Figure 4(a). Overall, I find that instances of well deepening are at most infrequent in the data and including them or removing them does not make a substantive difference to the results. I now explain in more detail about how I identify instances of well deepening, and illustrate the effects of deleting versus keeping these wells in the data.

In order to identify instances of well deepening I first investigate the ‘proposal to drill’ option in the ‘Applications for Permit to Drill’ database. Amongst the options here is to apply to the regulator to ‘deepen’ a well (as opposed to e.g. drill a ‘new well’). In the data the ‘proposal to drill’ code is not recorded for every permit. However, for the permits in which I do observe this code I do not find any permits that select the ‘deepen’ option, which suggests that this is not a widespread behavior. Since deepening may be still be occurring for the subset of permits which do not directly contain this information, I alternatively try to conservatively identify any instances of deepening by looking at if any additional wellbores were drilled more than 365 days after the original hole.[15] Removing these wells from the dataset results in

[15]To implement this, I look at the ‘wellbore code’ which is contained in the last two digits of the well API number. A value of 00 for the wellbore code indicates the original wellbore. I look at if any additional wellbores were spudded more than 365 days after the initial wellbore depth date.

around 318 (12.0%) fewer contracts, but the sorting patterns in Figure 4(a) do not substantially change. Since this procedure is conservative and results in deleting many wells that are not being deepened, I retain these wells in the dataset.

E.2 Robustness to only using the natural gas price as a state variable

I now explore the implications of lumping both oil and gas together in Figure A-6 and Table A-7. I provide a more detailed discussion below, but overall I argue that because wells in the shallow water of the Gulf of Mexico are mainly producing natural gas it does not seem an unreasonable assumption to abstract away from differences in oil and gas prices in this context.[16]

In Figure A-6(a) I plot the proportion of hydrocarbons produced by the wells drilled under each contract that are natural gas. To compute this measure I use the realized production of oil and natural gas from each well for 5 years after the date of first production (where these data are available). I then aggregate total production of oil and natural gas up to the contract level (since some contracts are for multiple wells). Finally, since oil and natural gas are measured in different units, I convert natural gas production into a 'barrel of oil equivalent' by dividing each cubic foot of natural gas production by 1/6000.[17]

Figure A-6(a) documents that the wells in the shallow water of the US Gulf of Mexico are primarily producing natural gas. This is the motivation of simply using the natural gas price as the state variable rather than both oil and natural gas prices.

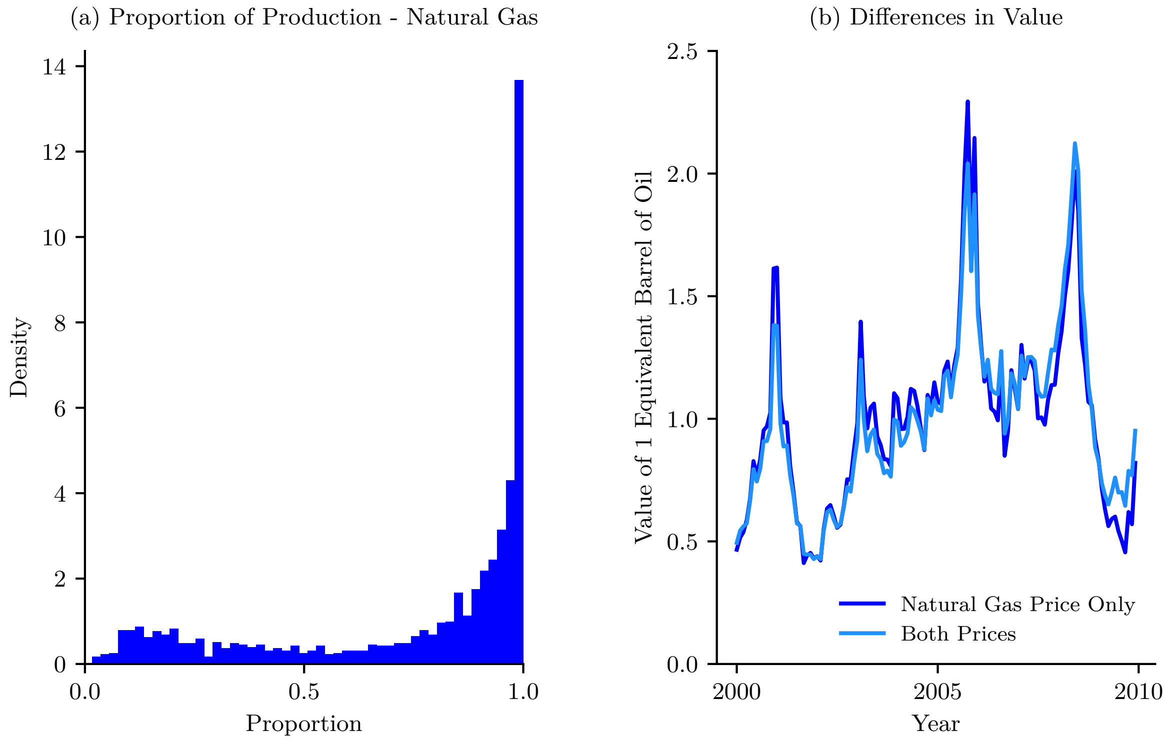
Figure A-6(b) elaborates on the implications of Figure A-6(a) through a thought experiment of how the value of the 'average well' (a well where natural gas is 75% of the total production of hydrocarbons) changes if I use only the natural gas price versus using a weighted average of both the oil and natural gas prices. Concretely, I consider a hypothetical well where the total value of hydrocarbons in the well = 1.0 at the average natural gas price. I then consider how the value of hydrocarbons changes across the cycle in 2000-2009 if I just use the natural gas price versus if I use a weighted average of both prices. The results Figure A-6(b) show that the series are relatively similar, suggesting that using only the natural gas price is not too strong an assumption.

Finally, in Table A-7 I present two regressions that test whether oil and gas heterogeneity makes a

[16]Note that the way that oil and gas enter into the model is through lease bids which proxy for ex-ante expectations over the 'total quantity of hydrocarbons' in the well, rather than the amount of oil and gas produced by a well after it has been drilled. Therefore the primary way that not accounting for differences in oil and gas prices would affect the results is through how this value of hydrocarbons (computed as the quantity of hydrocarbons multiplied by the current natural gas price) differs over the cycle.

[17]This is the value used by the US Geological Survey to compute total hydrocarbon reserves (https://certmapper.cr.usgs.gov/data/PubArchives/WEcont/world/woutsum.pdf). Oil and gas companies also use similar 'barrel of oil equivalent' measures in financial statements when disclosing their reserves to investors.

Figure A-6: Heterogeneity of oil and gas output



Note: This figure documents the degree of heterogeneity in oil/gas output. Panel (a) shows the distribution of the proportion of total production that is natural gas for each contract. Panel (b) show how the value of a hypothetical 'average well' with 75% natural gas would change across the sample period depending on whether the natural gas price is used versus both the oil price and the natural gas price. The hypothetical 'average well' has a value that is normalized to = 1.0 at the average gas price.

difference for the way wells are drilled. These regressions test whether the complexity of the design of the well is a function of whether the hydrocarbons that the well produces are predominately oil or natural gas. Overall, I find that the relationship between the proportion of production that is natural gas and the complexity of the well is not statistically significant. Therefore, I do not find any evidence that heterogeneity of oil/gas output affects the complexity of a well.

Table A-7: Regressions of well complexity on the type of hydrocarbons the well produces

	(1) Complexity	(2) Complexity
Proportion of production - natural gas	0.028	0.025
	(0.031)	(0.031)
Time FEs	No	Yes
N	1796	1796

Note: * p < 0.1, ** p < 0.05, *** p < 0.01. Robust standard errors are in brackets. Time FEs incorporates fixed effects for the year and the month that the contract is signed. These regressions are performed at the contract level for contracts where production data are available.

E.3 Robustness to non-myopic wells

Robustness exercise: overview Overall, the main finding is that allowing projects to have an outside option of negotiation failure - together with re-estimating the parameters of the model to ensure that the model makes sensible in-sample predictions - does not substantially change the results or exacerbate the implied welfare effects from entry. I now provide more details about these robustness exercises, discuss the benefits and limitations, and the results.

Implementation I include an outside option for projects in the following form: $\beta(1-\mathbb{P}_{exit})q_y^{project}(\theta_{yt})W_{xyt}$. Here, if there is a negotiation failure, projects wait for one period. They are then hit by an exogenous exit shock (so that $(1-\mathbb{P}_{exit})$ is the probability that the project survives and is still available to match). Finally, if the project survives, it can rematch with probability $q_y^{project}(\theta_{yt})$ resulting in the value W_{xyt}. For simplicity, I assume that if the well rematches then it will match with the same capital type as the original match.[18]

The above specification allows for the project's outside option to be dependent on the state of the market at time t through the probability of (re-)matching $q_y^{project}(\theta_{yt})$ as well as the value W_{xyt}. Concretely, in a boom the project's probability of re-matching $q_y^{project}(\theta_{yt})$ will decrease but the value of the project W_{xyt} will increase, which lead to overall theoretically ambiguous effects on the well's outside option.

There are at least two implicit assumptions in the above specification of the project's outside option, which I now detail. First, I assume that the well rematches with the same capital type; I make this assumption to keep the robustness check as parsimonious as possible. Second, since I do not observe searching projects (recall I only observe matches and searching capital), I need to make an assumption on the probability of project exit $\mathbb{P}_{exit}$ and the probability of project re-matching $q_y^{project}(\theta_{yt})$. I perform the

[18]A related setup is used for exporters in Brancaccio et al. (2020).

Table A-8: Robustness exercises with non-myopic projects

$\mathbb{P}_{exit}$	**1.0** *(Baseline)*	**0.95**	**0.75**
Value of sorting effect	12.0%	16.9%	14.9%

Note: This table provides an overview of how the results change when projects have an outside option of $\beta(1-\mathbb{P}_{exit})q_y^{project}(\theta_{yt})W_{xyt}$, for different assumptions about $\mathbb{P}_{exit}$. Note that I re-estimate all the parameters in each of the robustness exercise scenarios.

robustness exercise for several different values of $\mathbb{P}_{exit}$. For $q_y^{project}(\theta_{yt})$, I treat the equilibrium values from the benchmark model as 'data' and estimate a logit model of the probability of matching each capital type conditional on a second-order polynomial of the state. I then use this model to compute $q_y^{project}(\theta_{yt})$ in the robustness exercise, for different states of the market.

Finally, in all of the robustness exercises I also re-estimate the parameters. This is important to allow the data to discipline the model and ensure that - for example - the model in the robustness exercises continues to replicate the empirical sorting patterns.

Results from the robustness exercises I detail the results of the robustness exercises in Table A-8. Overall, allowing for non-myopic projects - as well as re-estimating the model - appears to increase the value of the sorting effect. The exact change does not appear to be monotone in $\mathbb{P}_{exit}$. The ultimate conclusion that the sorting effect is economically significant still remains. If the parameters are not re-estimated then allowing for non-myopic projects increases the probability of rejecting a match in both booms and busts. This is because the match surplus now becomes:

$$S_{xyt} = W_{xyt} - \beta(1-\mathbb{P}_{exit})q_y^{project}(\theta_{yt})W_{xyt} + V_{xyt} - \beta\mathbb{E}_t U_{y,t+1}$$

and the project's outside option $\beta(1-\mathbb{P}_{exit})q_y^{project}(\theta_{yt})W_{xyt} \geq 0$. However, after re-estimating the parameters, it moderates this effect.[19]

E.4 Robustness to the assumption that projects target capital

Implication of the assumption that project owners direct their search The implications of this assumption enter into the model through the targeting weights expression. In the paper I show the targeting weights can be micro-founded from the individual search decisions of well owners. I now

[19] There are many components to the model that are re-estimated, and so it is not clear before performing the exercise if the net effect of allowing for non-myopic projects will increase or decrease the value of the sorting effect, or if the effects will be monotone in the value of $\mathbb{P}_{exit}$.

derive the equivalent expression for the targeting weights when rig owners make the decision about whom to match with.

To begin this alternative derivation, denote each unit of available capital by j and the corresponding type as y_j. Similarly, denote each searching project by i and its corresponding type by x_i. Using this notation, for a rig of type y_j, the (expected) value of targeting well i is $\pi_{x_i y_j t} = q_{y_j}^{capital}(\theta_{y_j t})(\delta S_{x_i y_j t} + \beta \mathbb{E}_t U_{y_j t+1})$. Here I am using the Nash bargaining result that the surplus of a match will be split so that the rig receives $\delta S_{x_i y_j t} + \beta \mathbb{E}_t U_{y_j t+1}$ and the well receives $(1-\delta) S_{x_i y_j t}$.

Next, as in the micro-foundation of the targeting weights in Appendix B.1, I setup the rig's problem as choosing which well to target based on a *perceived value* $\max_i \hat{\pi}_{x_i y_j t}$. This value is given by:

$$\hat{\pi}_{x_i y_j t} = \pi_{x_i y_j t} - \gamma_1 1\left[x_i \notin A_{y_j t}\right] + \epsilon_{ijt}^{target} \tag{A-41}$$

I assume that ϵ_{ijt}^{target} are drawn from an i.i.d. type-1 extreme value distribution with scale parameter $1/\gamma_0$. The conditional choice probability that a rig j targets well i at time t is then:

$$P_{ijt}^{capital} = \frac{\exp\left(\gamma_0 \left[\delta q_{y_j}^{capital}(\theta_{y_j t}) S_{x_i y_j t} - \gamma_1 1\left[x_i \notin A_{y_j t}\right]\right]\right)}{\sum_k \exp\left(\gamma_0 \left[\delta q_{y_j}^{capital}(\theta_{y_j t}) S_{x_k y_j t} - \gamma_1 1\left[x_k \notin A_{y_j t}\right]\right]\right)} \tag{A-42}$$

Here the $q_{y_j}^{capital}(\theta_{y_j t})\beta \mathbb{E}_t U_{y_j t+1}$ terms in $\pi_{x_i y_j t}$ enter additively and are the same for all alternatives and so they cancel. Finally, aggregating $P_{ijt}^{capital}$ over all j, the targeting weight is:

$$\omega'_{xyt} = \frac{n_{yt} \exp\left(\gamma_0 \left[\delta q_{y}^{capital}(\theta_{yt}) S_{xyt} - \gamma_1 1\left[x \notin A_{yt}\right]\right]\right)}{\sum_{k \in Y} n_{kt} \exp\left(\gamma_0 \left[\delta q_{k}^{capital}(\theta_{kt}) S_{xkt} - \gamma_1 1\left[x \notin A_{kt}\right]\right]\right)} \tag{A-43}$$

This can be compared to the targeting weight in the model which - writing in terms of the total surplus of a match - is:

$$\omega_{xyt} = \frac{n_{yt} \exp\left(\gamma_0 \left[(1-\delta) q_{y}^{project}(\theta_{yt}) S_{xyt} - \gamma_1 1\left[x \notin A_{yt}\right]\right]\right)}{\sum_{k \in Y} n_{kt} \exp\left(\gamma_0 \left[(1-\delta) q_{k}^{project}(\theta_{kt}) S_{xkt} - \gamma_1 1\left[x \notin A_{kt}\right]\right]\right)} \tag{A-44}$$

I now discuss under what conditions the targeting weights are the same, regardless of which side of the market is assumed to choose who to match with. That is, I discuss under what conditions $\omega'_{xyt} = \omega_{xyt}$. If search is random ($\gamma_0 = 0$) then the targeting weights are the same. Similarly, if both sides of the market have approximately equal bargaining power ($\delta = 1/2$) and the probability of matching is the same for capital and projects ($q_y^{project}(\theta_{yt}) = q_y^{capital}(\theta_{yt})$), then the targeting weights are then same. Therefore, the extent to which the assumption matters hinges on whether these conditions hold in the estimated model.

Robustness exercise Finally, I re-run the model using the targeting weights derived in Equation A-43. (That is, I re-run the model assuming the other extreme that it is rigs - rather than project owners

Table A-9: Robustness test of rigs vs project owners (wells) targeting their search.

	Wells target rigs	Rigs target wells
Sorting Effect	12.0%	11.0%

- who target their search). I report the results for the sorting effect counterfactual in Table A-9. Overall, Table A-9 shows that the results do not hinge on which side of the market is assumed to direct search, with qualitatively similar results under either assumption.

E.5 Robustness to changes in the period length

Table A-10: Results from the period length robustness check

	One-month period	Two-week period
Welfare (billions USD)	\$5.01	\$5.55

Note: This table reports robustness of the model to different period lengths. Specifically, I compare the results for the sorting effect with a one-month period length (which is assumed in the model) versus a fortnightly period length. In order to compute the fortnightly period length model I take the following steps. First, I recompute Step 1 of the estimation with a two-week period length; this includes re-estimating the empirical objects that underpin the value functions, and re-simulating the value functions. Next, I re-compute the model at the baseline parameters (i.e. the parameters corresponding to a one-month period length) except for the potential well draw parameters d_0, d_1, which I halve. Therefore, each two-week period has approximately half the number of potential wells as the corresponding one-month period. I also halve the match value terms, $m_{0,y}, m_{1,y}$ for $y \in \{low, mid, high\}$ and m_2, since they now correspond to the match value for half a month.

References

Brancaccio, G., Kalouptsidi, M. and Papageorgiou, T. (2020), 'Geography, Transportation, and Endogenous Trade Costs', *Econometrica* **88**, 657–691.

Kaiser, M. (2007), 'A Survey of Drilling Cost and Complexity Estimation Models', *International Journal of Petroleum Science and Technology* **1**(1), 1–22.

Kellogg, R. (2011), 'Learning by Drilling: Interfirm Learning and Relationship Persistence in the Texas Oilpatch', *The Quarterly Journal of Economics* **126**(4), 1961–2004.

Vreugdenhil, N. (2021), 'Incomplete Environmental Regulation and Spatial Equilibrium in the Offshore Oil and Gas Industry', *Working Paper* .