Information Externalities, Free Riding, and Optimal Exploration in the UK Oil Industry

Charles Hodgson*

October 14, 2024

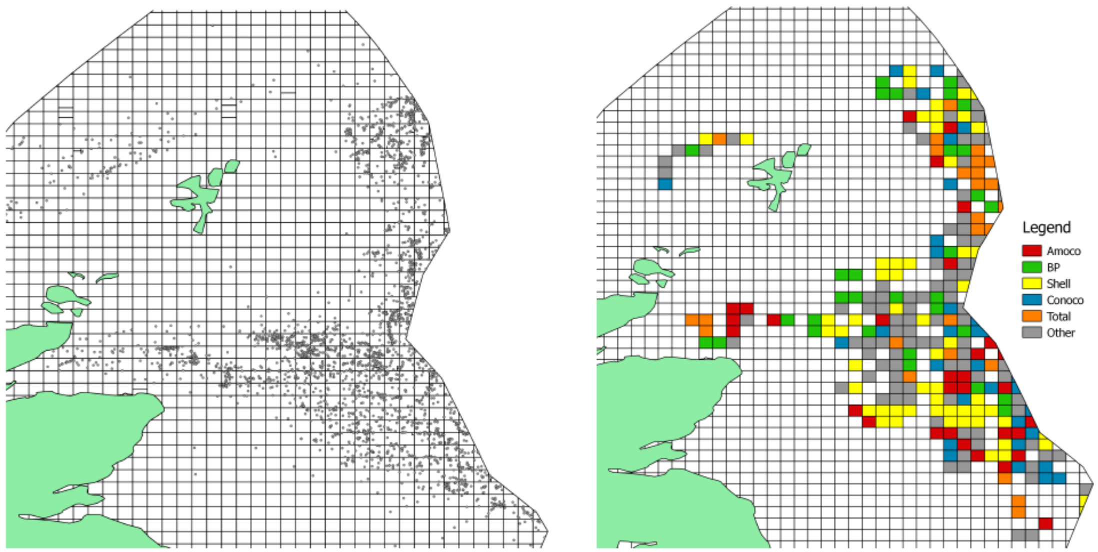
2 UK Oil Exploration: Setting and Data

I use data covering the history of oil drilling in the UK Continental Shelf (UKCS) from 1964 to 1990. Oil exploration and production on the UKCS is carried out by private companies who hold drilling licenses issued by the government. The first such licenses were issued in 1964, and the first successful (oil yielding) well was drilled in 1969. Discoveries of the large Forties and Brent oil fields followed in 1970 and 1971. Drilling activity took off after the oil price shock of 1973, and by the 1980s the North Sea was an important producer of oil and gas. I focus on the region of the UKCS north of $55°N$ and east of $2°W$, mapped in Figure 1, which is bordered on the north and east by the Norwegian and Faroese economic zones. This region contains the main oil producing areas of the North Sea and has few natural gas fields, which are mostly south of $55°N$.[3]

2.1 Technology

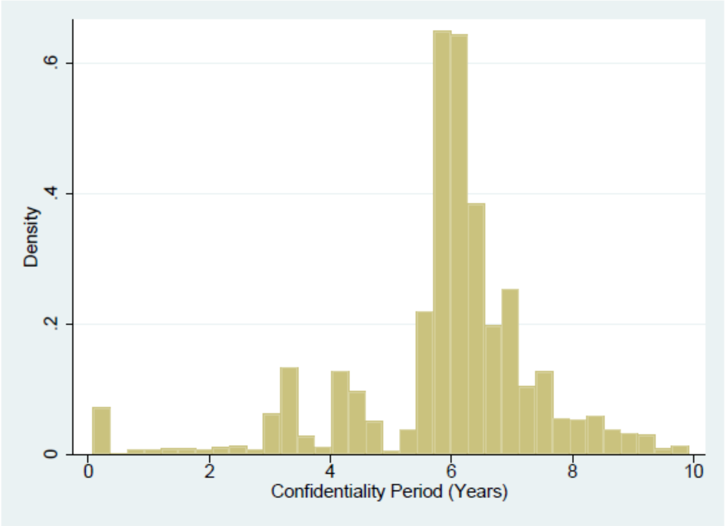
Offshore oil production involves two investment phases. First, *exploration wells* are drilled from mobile rigs or ships to locate oil reservoirs. It is important to note that the results of a single exploration well provide limited information about the size of an oil deposit, and many exploration wells must be drilled to estimate the volume of a reservoir. When a sufficiently large oil field has been located, the field is *developed*. This second phase of investment involves the construction of a production platform - a large static facility typically anchored to the seabed.

Figure 1: Wells and License Blocks



Notes: Grid squares are license blocks. The left panel plots the location of all exploration wells drilled from 1964 to 1990. The right panel records license holders for each block in January 1975. Note that if multiple firms hold licenses on separate sections of a block, only one of those firms (chosen at random) is represented on this map.

Figure A4: Confidentiality Periods



Notes: Histogram of the time from well completion to data release for all exploration wells. 27 observations with confidentiality periods less than or equal to 0 or greater than 10 years are dropped.

Table 1: Summary Statistics: Blocks & Wells

Firm-Blocks	All	Explored	Exp. & Developed	Exp. & Not Dev.	Not Exp.
N	1470	721	160	561	749
Share Explored	.490	1.000	1.000	1.000	0.000
Share Developed	.120	.222	1.000	0.000	.021
First Exp. After Work Date	.	.227	.280	.215	.
Own Share of Nearby Blocks:					
Mean	.199	.178	.181	.177	.219
SD	.217	.199	.206	.197	.231
Exploration Wells per Block	2.002	4.082	10.138	2.355	0.000
Share Successful	.199	.199	.444	.129	.

Notes: Table records statistics on all license-block pairs active between 1964 and 1990. In particular, if a block is licensed to multiple firms it appears multiple times in this Table. Each column records statistics on subsets of license-blocks defined according to whether they are ever explored or developed. Own share of nearby blocks is defined as the share of license-blocks that are at most third degree neighbors that are licensed to the same firm.

3 A Model of Spatially Correlated Beliefs

The effect of information externalities on firms' exploration decisions depends on the spatial arrangement of licenses, the extent to which firms can observe the results of each other's wells, and on the correlation of exploration results at different locations. In Hodgson (2024) I show that in a simple two-firm, two-block model, spatial correlation in well outcomes reduces the equilibrium rate of exploration below the social optimum. The magnitude of this free riding effect is determined by the extent to which well results are correlated over space. In particular, the more correlated are outcomes on neighboring blocks, the lower the equilibrium rate of exploration.

3.1 Statistical Model of the Distribution of Oil

I start by describing a statistical model of the distribution of oil over space. I model the probability that an exploration well at a particular location is successful as a continuous function over space drawn from a Gaussian process. This model assumes that the location of oil is distributed randomly over space but allows spatial correlation - the outcomes of exploration wells close to each other are highly correlated and the degree of correlation declines with distance. A draw from this process is a continuous function that, depending on the parameters of the process, can have many local maxima corresponding to separate clusters of oil fields (see Appendix Figure A5 for a one dimensional example). As I discuss further below, such models are widely used in natural resource exploration to model the spatial distribution of geological features. The estimation of these models is sometimes known as Kriging in the geostatistics literature.[12].

Formally, let $\rho(X) : \boldsymbol{X} \to [0,1]$ be a function that defines the probability of exploration well success at locations $X \in \boldsymbol{X}$. I model $\rho(X)$ as being drawn from a *logistic Gaussian process* $G(\rho)$ over the space $\boldsymbol{X}$.[13] In particular, for any location X,

$$\rho(X) \equiv \rho(\lambda(X)) = \frac{1}{1 + exp(-\lambda(X))}, \tag{1}$$

where $\lambda(X)$ is a continuous function from $\boldsymbol{X}$ to $\mathbb{R}$.

The function $\lambda(X)$ is drawn from a *Gaussian process* with mean function $\mu(X)$ and covariance function $\kappa(X, X')$. This means that for any finite collection of K locations $\{1, ..., K\}$, the vector $(\lambda(X_1), ..., \lambda(X_K))$ is a multivariate normal random variable with mean $(\mu(X_1), ..., \mu(X_K))$ and a covariance matrix with (j,k) element $\kappa(X_j, X_k)$. The prior mean function $\mu : \boldsymbol{X} \to \mathbb{R}$ is assumed to be smooth and the covariance function $\kappa : \boldsymbol{X} \times \boldsymbol{X} \to \mathbb{R}$ must be such that the resulting covariance matrix for any K locations is symmetric and positive semi-definite. One covariance function that satisfies these assumptions is the square exponential covariance function (Rasmussen and Williams, 2006) given by

$$\kappa(X, X') = \omega^2 exp\left(\frac{-\left|X - X'\right|^2}{2\ell^2}\right). \tag{2}$$

The parameter ω controls the variance of the process. In particular, for any X, the marginal distribution of $\lambda(X)$ is given by $\lambda(X) \sim N(\mu(X), \omega)$. The parameter ℓ controls the covariance between $\lambda(X)$ and $\lambda(X')$ for $X \neq X'$. Notice that as the distance $|X - X'|$ between two locations increases, the covariance falls at a rate proportional to ℓ. As $|X - X'|$ goes to 0, the correlation of $\lambda(X)$ and $\lambda(X')$ goes to 1, so draws from this process are continuous functions.

Table 2: Oil Process Parameters

Parameter	Estimate	Implied Statistics	
μ	-1.728	$E(\rho(X))$	0.207
	(0.202)		
ω	1.2664	$SD(\rho(X))$	0.179
	(0.146)		
ℓ	0.862	$Corr(\rho(0), \rho(1))$	0.471
	(0.102)		

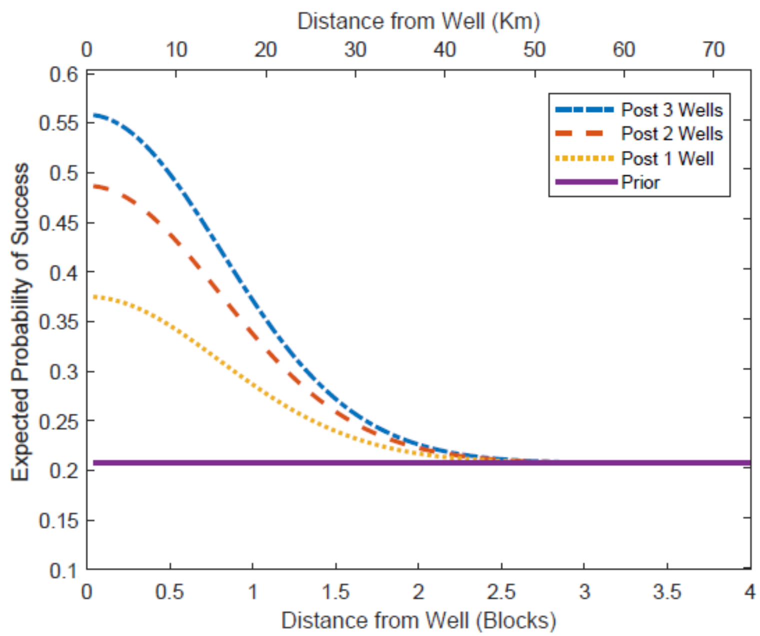
Notes: The first column records parameter estimates from fitting the likelihood function given by equation 3 to data on the outcome of all exploration wells drilled between 1964 and 1990 on the relevant area of the North Sea. Well locations are taken to be the centroids of corresponding blocks. Standard errors computed using the Hessian of the likelihood function in parentheses. The second column records the implied expected probability of success, the standard deviation of the prior beliefs about probability of success, and the correlation of success probability between two locations one block (18 km) away from each other.

3.2 Interpretation as a Bayesian Prior

The estimated parameters, (μ, ω, ℓ), can be thought of as describing primitive geological characteristics that determine the distribution of oil deposits over space. If these parameters are known by firms and the Gaussian process model is a good approximation to the geological process that generates the distribution of oil, then the estimated process $G(\rho|\mu, \omega, \ell)$ describes the rational beliefs that firms should hold about the probability of exploration well success at each location X prior to observing the outcome of any wells. The parameters of this prior also determine how beliefs are updated according to Bayes' rule after well results are observed.

In particular, firms whose prior is described by $G(\rho)$ update their beliefs over the entire space $\boldsymbol{X}$ after observing a success or failure at a particular location X. Posterior beliefs at locations closer to X will be updated more than those at more distant locations. Figure 2 illustrates how posterior beliefs respond to well outcomes at different distances under the estimated parameters. The solid purple line illustrates the firm's constant prior expected probability of success of around 0.2.[14] The dotted yellow line represents the firm's posterior expected probability of success after observing one successful well at 0 on the x-axis. The dashed red and blue lines correspond to posteriors after observing two and three successful

Figure 2: Response of Beliefs to Well Outcomes



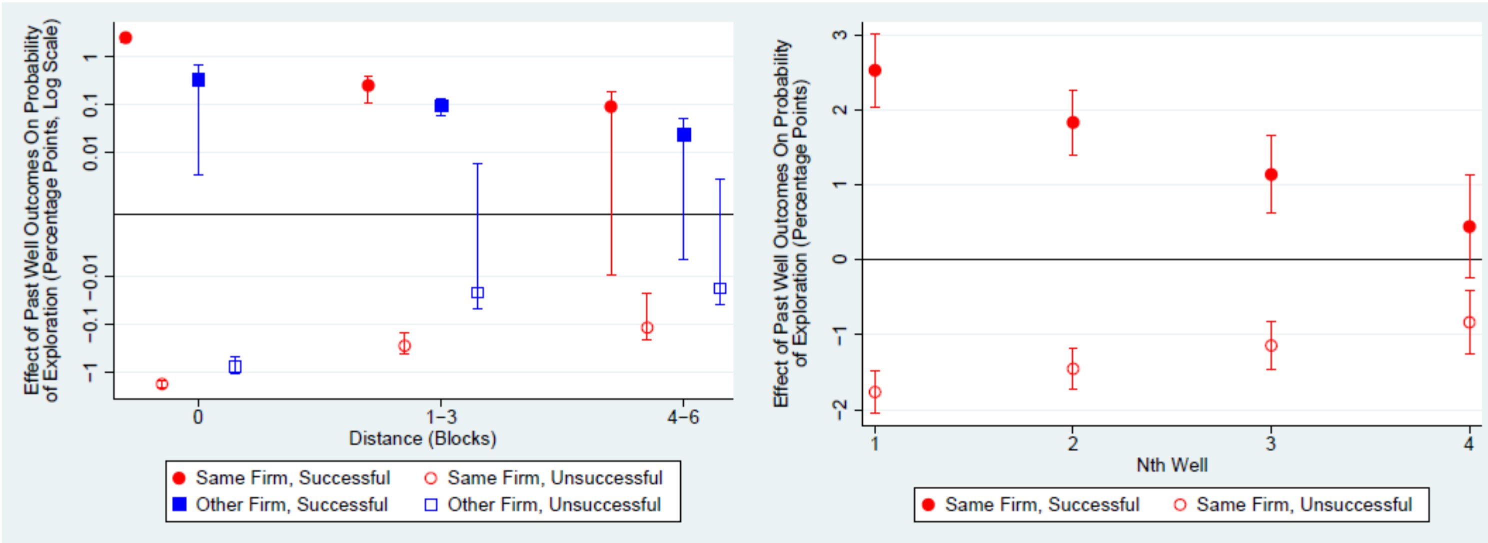
Notes: Figure depicts prior and posterior expected value of $\rho(X)$ in a one dimensional space for posteriors computed after observing one, two, and three successful wells at $X = 0$. The parameters (μ, ω, ℓ) of the logistic Gaussian process prior are set to the estimated values from Table 2.

4.1 Exploration Drilling Patterns

The estimated spatial correlation illustrated by Figure 2 suggests that firms should make inferences across space based on past well results. I test this prediction using data on firm behavior. Let Suc_{jdot} be the cumulative number of successful wells drilled on blocks distance d from block j before date t by firms $o \in \{f, -f\}$, where $-f$ indicates all firms other than firm f (including exploration wells drilled on now-developed blocks). $Fail_{jdot}$ is analogously defined as the cumulative number of past unsuccessful wells. I estimate the following regression specification using OLS:

$$Explore_{fjt} = \alpha_{ft} + \beta_j + \sum_d \sum_{o \in \{f, -f\}} g_{do} \left(Suc_{jdot}, Fail_{jdot} \right)) + \epsilon_{fjt}. \tag{7}$$

Figure 3: Response of Drilling Probability to Cumulative Past Results



Notes: Points are the estimated marginal effect of each type of past well on $Explore_{fjt}$ from the specification given by equation 7 where $g_{do}(\cdot)$ is quadratic in each of the arguments. In the left panel, marginal effects are computed for the first well of each type. In the right panel, the marginal effects of same-firm same-block wells are reported for the nth well of each type. Error bars are 95% confidence intervals computed using standard errors clustered at the firm-month level. Sample includes block-months in the relevant region before 1991. I drop observations from highly explored regions where the number of nearby own wells (those on 1st and 2nd degree neighboring blocks) is above the 80th percentile of the distribution in the data.

Table 3: Response of Drilling Probability to Posterior Beliefs and License Distribution

Dependent Variable:	KL Divergence	Exploration Well	Develop Block
Posterior Mean	0.414***	0.267***	0.004
	(0.001)	(0.070)	(0.003)
Posterior Mean^2	-0.450***	-0.206**	.
	(0.001)	(0.100)	.
Posterior Variance	0.082***	0.0199***	-0.002***
	(0.000)	(0.006)	(0.001)
Own Share Nearby Blocks	.	0.017**	0.000
	.	(0.008)	(0.001)
R^2	0.970	0.106	0.077
N	86,362	86,362	86,362
Firm-Month and Block FE	No	Yes	Yes

Notes: Standard errors clustered at the firm-month level. Mean, variance, and KL divergence of posterior beliefs computed for each (f, j, t) as if all wells drilled by all firms up to month $t-1$ are observed. Sample is all *undeveloped* firm-block-months in the relevant region,. *** indicates significance at the 99% level. **

5.1 Full Information

I start by specifying a full information game played by a set of firms F. Firms are indexed by f, discrete time periods are indexed by t, and blocks are indexed by j. J is the set of all blocks. $J_{ft} \subset J$ is the set of undeveloped blocks on which firm f holds drilling rights at the beginning of period t. $J_{0t} \subset J$ is the set of undeveloped blocks on which no firm holds drilling rights at the beginning of period t. P_t is the oil price.

Exploration wells are indexed by w, and each well is associated with an outcome $s(w) \in \{0, 1\}$, a block $j(w)$, a firm $f(w)$, and a drill date $t(w)$. The set of all locations and realizations of exploration wells drilled on date t is given by $W_t = \{(j(w), s(w)) : t(w) = t\}$. The set of all blocks that are developed on date t is J_{dt}.

The firm's prior beliefs about the probability of exploration well success on each block are given by the logistic Gaussian process G_0 defined in Section 3.1. G_{ft} is firm f's posterior at the beginning of period t. Under the assumption of full information firms observe the results of all wells, so $G_{ft} = G_t$ for all firms $f \in F$.

The industry state at date t is described by

$$\mathcal{S}_t = \{G_t, \{J_{ft}\}_{f \in F \cup \{0\}}, P_t\}. \tag{8}$$

The timing of the game is as follows: *Exploration Stage*

1. Given state $\mathcal{S}_t$, each firm f observes a vector of private cost shocks $\boldsymbol{\epsilon}_{ft}$.
2. Firm f chooses an exploration action, $a_{ft}^E \in J_{ft} \cup \{0\}$. If $a_{ft}^E \neq 0$, then firm f incurs an exploration cost.

Development Stage

1. Given state $\mathcal{S}'_{ft} = (\mathcal{S}_t, a_{ft}^E)$, each firm f observes a vector of private cost shocks $\boldsymbol{\nu}_{ft}$.
2. Firm f chooses a development action, $a_{ft}^D \in J_{ft} \cup \{0\}$. If $a_{ft}^D \neq 0$, then firm f incurs a development cost.

Well Outcomes Realized

1. Exploration well results W_t are realized.
2. If $a_{ft}^D = j$ then $j \in J_{dt}$, and firm f draws oil revenue π_{jt}.
3. The industry state evolves to $\mathcal{S}_{t+1} = \{G_{t+1}, \{J_{ft+1}\}_{f \in F \cup \{0\}}, P_{t+1}\}$.Exploration well outcomes and development payoffs are not realized until the end of the development phase. Beliefs G_t are then updated based on exploration well results W_t and realized revenues $\{\pi_{jt}\}_{j \in J_{dt}}$according to $G_{t+1} = B(G'_t, W_t, \{\pi_j, P_t\}_{j \in J_{dt}})$, where $B(\cdot)$ is defined in equation 6.

$$V_f^E(\mathcal{S}_t) = E_{\epsilon_{ft}}\left[\max_{a_{ft}^E \in J_{ft}\cup\{0\}} \{V_f^D(\mathcal{S}'_{ft}) - c(a_{ft}^E, \mathcal{S}_t) + \epsilon_{ftj}\}\right] \tag{9}$$

$$V_f^D(\mathcal{S}'_{ft}) = E_{\nu_{ft}}\left[\max_{a_{ft}^D \in J_{ft}\cup\{0\}} \left\{E_{\pi_{a_{ft}^D}, \mathcal{S}_{t+1}}\left[\beta V_f^E(\mathcal{S}_{t+1}) + \pi_{a_{ft}^D} | a_{ft}^D, \mathcal{S}'_{ft}\right] - \kappa(a_{ft}^D, \mathcal{S}'_{ft}) + \nu_{ftj}\right\}\right].$$

Where β is the one period discount rate.The inner expectation in the development Bellman equation is taken over realizations of other firms' exploration and development actions, development revenues $\{\pi_{jt}\}_{j\in J_{dt}}$ and exploration results W_t which affect next period's s beliefs G_{t+1}, as well as the evolution of P_{t+1} and J_{t+1}. There is no expectation inside the exploration Bellman because beliefs are not updated until the end of the development phase.

Define choice-specific ex-ante (before private cost shocks are realized) value functions as,

$$\begin{aligned} v_f^E(a_t^E, \mathcal{S}_t) =& E_{\mathcal{S}'_t}\left[V_f^D(\mathcal{S}'_t) | a_t^E, \mathcal{S}_t\right] - c(a_t^E, \mathcal{S}_t) \\ v_f^D(a_t^D, \mathcal{S}'_t) =& E_{\pi_{a_t^D}, \mathcal{S}_{t+1}}\left[\beta V_f^E(\mathcal{S}_{t+1}) + \pi_{a_t^D} | a_t^D, \mathcal{S}'_t\right] - \kappa(a_t^D, \mathcal{S}'_t). \end{aligned} \tag{10}$$

I assume that ϵ_{ftj} and ν_{ftj} are distributed type I extreme value with standard deviation parameters σ_ϵ and σ_ν, yielding conditional choice probabilities (CCPs):

$$P(a_f^E = j|\mathcal{S}_t) = \frac{exp\left(\frac{1}{\sigma_\epsilon} v_f^E(j, \mathcal{S}_t)\right)}{\sum_{k \in J_{ft} \cup \{0\}} exp\left(\frac{1}{\sigma_\epsilon} v_f^E(k, \mathcal{S}_t)\right)}. \tag{11}$$

With a similar expression for the CCP of development action j, $P(a_f^D = j|\mathcal{S}_t')$.

A Markov perfect equilibrium of this game is then defined by strategies $a_f^E(\mathcal{S}, \boldsymbol{\epsilon})$ and $a_f^D(\mathcal{S}, \boldsymbol{\nu})$ that maximize the firm's continuation value, conditional on the state variable and the privately observed cost shocks,

$$\begin{aligned} a_f^E(\mathcal{S}, \boldsymbol{\epsilon}) &= \arg \max_{a^E \in J_f \cup \{0\}} \left\{ v_f^E(a^E, \mathcal{S}) + \epsilon_{ta^E} \right\} \qquad (12) \\ a_f^D(\mathcal{S}', \boldsymbol{\nu}) &= \arg \max_{a^D \in J_f \cup \{0\}} \left\{ v_f^D(a^D, \mathcal{S}') + \nu_{ta^D} \right\}, \end{aligned}$$

where the firm forecasts all firms' actions conditional on the industry state using the true conditional choice probabilities is equation 11.

5.2 Asymmetric Information

To allow for imperfect spillovers of information in the model, I make an alternative assumption about when firms observe the results of exploration wells. In particular, when a well w is drilled by firm f, I let each firm $g \neq f$ observe the outcome, $s(w)$, with probability α. $s(w)$ is revealed to all firms τ_w periods after the well is drilled, on expiry of the confidentiality window. The length of the confidentiality window is drawn i.i.d. from the distribution $F_\tau(\tau_w)$.[20]

Formally, let $o_f(w) \in \{0,1\}$ be a random variable drawn independently across firms after the exploration stage of period $t(w)$ where $P(o_f(w) = 1 | f(w) \neq f) = \alpha$ and $P(o_f(w) = 1 | f(w) = f) = 1$. The cumulative set of well results observed by firm f in period t is

$$W^o_{ft} = \{(j(w), s(w)) : (o_f(w) = 1 \text{ and } t(w) \leq t) \text{ or } (o_f(w) = 0 \text{ and } t(w) \leq t - \tau(w))\} . \tag{13}$$

In general, $G_{ft} \neq G_{gt}$ since firms observe different sets of well outcomes, so firm f's state variable can now be written:

$$\mathcal{S}_{ft} = \{G_{ft}, \{J_{ft}\}_{f \in F \cup \{0\}}, P_t, W^u_{ft}\}, \tag{14}$$

where W^u_{ft} is the set of locations and dates of all wells whose results are *not* observed by firm f at date t,

$$W^u_{ft} = \{(j(w), t(w)) : o_f(w) = 0 \text{ and } t(w) \geq t - \tau(w)\} . \tag{15}$$

Assumption 1. *Firm f believes that at every period t the probability of exploration by a firm $g \neq f$ on block $j \in J_{gt}$ is given by $Q^E(g, j; \mathcal{S}_{ft}) \in [0, 1]$. Likewise, beliefs over the probability of development are $Q^D(g, j; \mathcal{S}_{ft}) \in [0, 1]$.*

Assumption 1 says that firms' beliefs are specified by functions Q^E and Q^D of their current state $\mathcal{S}_{ft}$. For example, these functions could be the expected action probabilities based on Bayesian posteriors about other firms' states. Although this assumption allows for Bayesian beliefs, it restricts the set of information that firms can use to make forecasts about other firms' actions to the payoff-relevant state variables in equation 14.

Assumption 2. *Firm f forecasts the outcomes of exploration and development wells using beliefs $Q^G(\rho; G(\rho), W^u_{ft})$.*

Assumption 2 says that firms adjust their beliefs about this distribution of oil based on the number of wells with unobserved results on each block, W^u_{ft}. Thus, beliefs depend on observed actions, not just the observed outcomes which define $G(\rho)$. For instance, it might be that $E(\rho_j|G(\rho)) < E(\rho_j|\tilde{G}(\rho))$ if there are many exploration wells on block j whose results are unobserved. As with Assumption A1, I restrict the observable variables that enter Q^G. For instance I do not allow beliefs to depend on the timing of past exploration.

The following equilibrium definition requires the belief objects Q^E, Q^D, and Q^G to be consistent with firm behavior.

Definition 1. An equilibrium is a set of policies, $a^E(\mathcal{S}_f, \epsilon)$, $a^D(\mathcal{S}_f, \epsilon)$, beliefs about other firms' actions, Q^E, Q^D, and beliefs about the distribution of oil, Q^G such that:

1. $a^E(\mathcal{S}_f, \epsilon)$ and $a^D(\mathcal{S}_f, \epsilon)$ solve the dynamic program described by equation 9, where expectations over other firms actions and well outcomes are described by Assumptions A1 and A2.

2. $Q^E(g, j; \mathcal{S}_{ft})$, $Q^D(g, j; \mathcal{S}_{ft})$ are consistent with equilibrium policies. Let $P(a^E_{f,t} = j|\mathcal{S}_{f,t})$ and $P(a^D_{f,t} = j|\mathcal{S}'_{f,t})$ be firms' equilibrium CCPs.

$$
\begin{aligned}
Q^E(\mathcal{S}_{ft}, g, j) &= E\left[P(a^E_{gt} = j|\mathcal{S}_{gt})|\mathcal{S}_{ft}\right] \\
Q^D(S_{ft}, g, j) &= E\left[P(a^D_{gt} = j|\mathcal{S}_{gt})|\mathcal{S}_{ft}\right].
\end{aligned}
\tag{16}
$$

   Where the expectations are taken over states with respect to the equilibrium distribution of $\mathcal{S}_{gt}$ conditional on $\mathcal{S}_{ft}$.

3. $Q^G(\rho; G(\rho), W^u_{ft})$ is the posterior distribution of ρ conditional on $(G(\rho), W^u_{ft})$ consistent with the equilibrium distribution of $(G(\rho), W^u_{ft})$ conditional on ρ and the prior distribution of ρ, $G_0(\rho)$.

Background for Hodgson paper
QJE 2012

DYNAMIC GAMES WITH ASYMMETRIC INFORMATION: A FRAMEWORK FOR EMPIRICAL WORK*

CHAIM FERSHTMAN AND ARIEL PAKES

We develop a framework for the analysis of dynamic oligopolies with persistant sources of asymmetric information that enables applied analysis of situations of empirical importance that have been difficult to deal with. The framework generates policies that are "relatively" easy for agents to use while still being optimal in a meaningful sense, and is amenable to empirical research in that its equilibrium conditions can be tested and equilibrium policies are relatively easy to compute. We conclude with an example that endogenizes the maintenance decisions of electricity generators when the costs states of the generators are private information. *JEL* Codes: L13, C73, D82.

strong complimentarities.

To accomplish these tasks we extend the framework in Ericson and Pakes (1995) to allow for asymmetric information.[1] Each agent's returns in a given period are determined by all agents' "payoff relevant" state variables and their actions. The payoff relevant random variables of producers would typically include indexes of their cost function, qualities of the goods they market, and so on, whereas in a durable good market those of consumers would include their current holdings of various goods and the household's demographic characteristics. Neither a player's "payoff relevant" state variables nor its actions are necessarily observed by other agents. Thus producers might not know either the cost positions or the details of supplier contacts of their competitors, and in the durable goods example neither consumers nor producers need know the entire distribution of holdings crossed with household characteristics (even though this will determine the distribution of future demand and prices).

The fact that not all state variables are observed by all agents and that the unobserved states may be correlated over time implies that variables that are not currently payoff relevant but are related to the unobserved past states of other agents will help predict other agent's behavior. Consequently, they will

More formally we define a state of the game to be the information sets of all of the players (each information set contains both public and private information). An experience-based equilibrium (EBE) for our game is a triple which satisfies three conditions. The triple consists of (1) a subset of the set of possible states, (2) a vector of strategies defined for every possible information set of each agent, and (3) a vector of values for every state that provides each agent's expected discounted value of net cash flow conditional on the possible actions that agent can take at that state. The conditions we impose on this triple are as follows. The first condition is that the equilibrium policies ensure that once we visit a state in our subset we stay within that subset in all future periods, visiting each point in that subset repeatedly. That is, the subset of states is a recurrent class of the Markov process generated by the equilibrium strategies. The second condition is that the strategies are optimal given the evaluations of outcomes. The final condition is that optimal behavior given these evaluations actually generates expected discounted value of future net cash flows that are consistent with these evaluations in the recurrent subset of states.

6 Estimation & Identification

6.1 Parameterization

I measure the oil price using the monthly West Texas Intermediate oil price inflated to 2011 dollars using the UK Producer Price Index converted to dollars using the UK/US exchange rate. The WTI price is highly correlated with the Brent crude price over the time period studied, and data on Brent crude prices is not available before 1980. For years before 1980 where the Brent price is unavailable I use projected values from a regression of Brent on the West Texas Intermediate price. I let a period be one month.[22] I set the one month discount rate to $\beta = 0.992$, which corresponds to a 10% annual discount.

I specify exploration and development costs according to equation 17.

$$\begin{aligned} c(j, \mathcal{S}_{ft}) &= c_0 \\ \kappa(j, \mathcal{S}_{ft}) &= \kappa_0 + \kappa_1 1(Var_{ft}(\rho_j) < \underline{v}). \end{aligned} \tag{17}$$

6.3 Identification

Identification of CCPs The first step of the estimation procedure recovers the parameter α and conditional choice probabilities $\hat{P}(a = j|\mathcal{S})$ at each state $\mathcal{S}$ from data in which each observation is consistent with a set of states $\tilde{\mathcal{S}}$. The model's information structure means these objects are separately identified despite the fact that the econometrician does not observe the full state. In particular, I claim that the list of choice probabilities $P(a = j|\tilde{\mathcal{S}})$ for each set of states $\tilde{\mathcal{S}}$ that it is possible to *observe* in the data can be inverted to uniquely identify choice probabilities conditioned on the *unobserved* states $P(a = j|\mathcal{S})$ *and* the information spillover parameter α.

To illustrate identification, consider the following simplified example. Suppose that a state is described by a triple, $\mathcal{S} = (suc, fail, unobs)$, where suc is the number of successful wells observed, $fail$ is the number of unsuccessful wells observed, and $unobs$ is the number of wells with unobserved outcomes. Consider data that contains observations consistent with the following sets of states:

$$\tilde{\mathcal{S}}_A = \{(1,0,0)\},\ \tilde{\mathcal{S}}_B = \{(0,1,0)\},\ \tilde{\mathcal{S}}_C = \{(1,0,0),(0,0,1)\},\ \tilde{\mathcal{S}}_D = \{(0,1,0),(0,0,1)\} \quad (20)$$

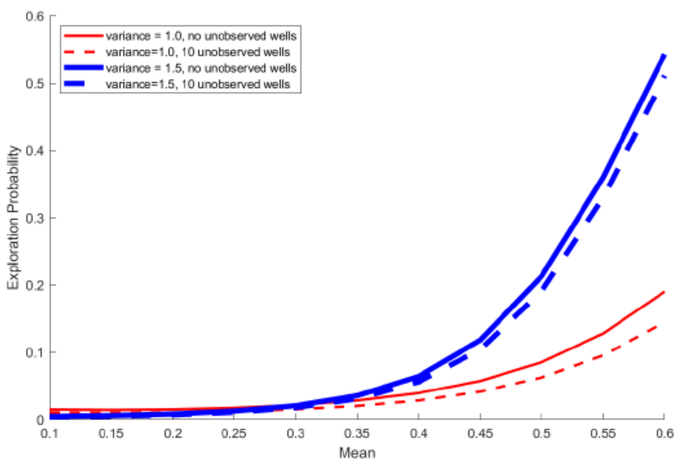
$\tilde{\mathcal{S}}_A$ and $\tilde{\mathcal{S}}_B$ are observed by the econometrician when there is one own-firm well outcome. The econometrician then knows the state with certainty since the firm always observes their own well outcome. $\tilde{\mathcal{S}}_C$ and $\tilde{\mathcal{S}}_D$ are observed by the econometrician when there is one other-firm well outcome. In this case, the econometrician knows whether the well was successful or unsuccessful, but not whether the firm observed the outcome or not. Given a value of the parameter α, choice probabilities conditional on the observed set of states can be written as:

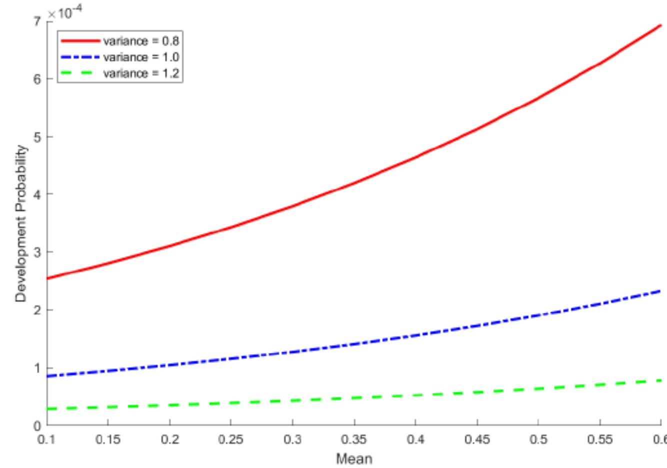
$$\begin{aligned}
P(a=j|\tilde{\mathcal{S}_A}) &= P(a=j|\mathcal{S}=(1,0,0)) \\
P(a=j|\tilde{\mathcal{S}_B}) &= P(a=j|\mathcal{S}=(0,1,0)) \\
P(a=j|\tilde{\mathcal{S}_C}) &= \alpha P(a=j|\mathcal{S}=(1,0,0)) + (1-\alpha)P(a=j|\mathcal{S}=(0,0,1)) \\
P(a=j|\tilde{\mathcal{S}_D}) &= \alpha P(a=j|\mathcal{S}=(0,1,0)) + (1-\alpha)P(a=j|\mathcal{S}=(0,0,1)).
\end{aligned} \tag{21}$$

The left hand side of each equation is a probability that is observable in the data. Notice that there are four equations and four unknowns - three conditional choice probabilities and the parameter α. The first two equations yield estimates of $P(a=j|\mathcal{S}=(1,0,0))$ and $P(a=j|\mathcal{S}=(0,1,0))$ directly. Rearranging the third and fourth equations yields:

$$\alpha = \frac{P(a=j|\tilde{\mathcal{S}_C}) - P(a=j|\tilde{\mathcal{S}_D})}{P(a=j|\tilde{\mathcal{S}_A}) - P(a=j|\tilde{\mathcal{S}_B})}. \tag{22}$$

Figure 4: Estimated CCPs





Notes: Simulated exploration and development probabilities using the estimated CCPs. Probabilities are simulated for a firm with a single block and one neighboring block held by another firm. Oil price is set to the average value in the data.

Table 4: Parameter Estimates

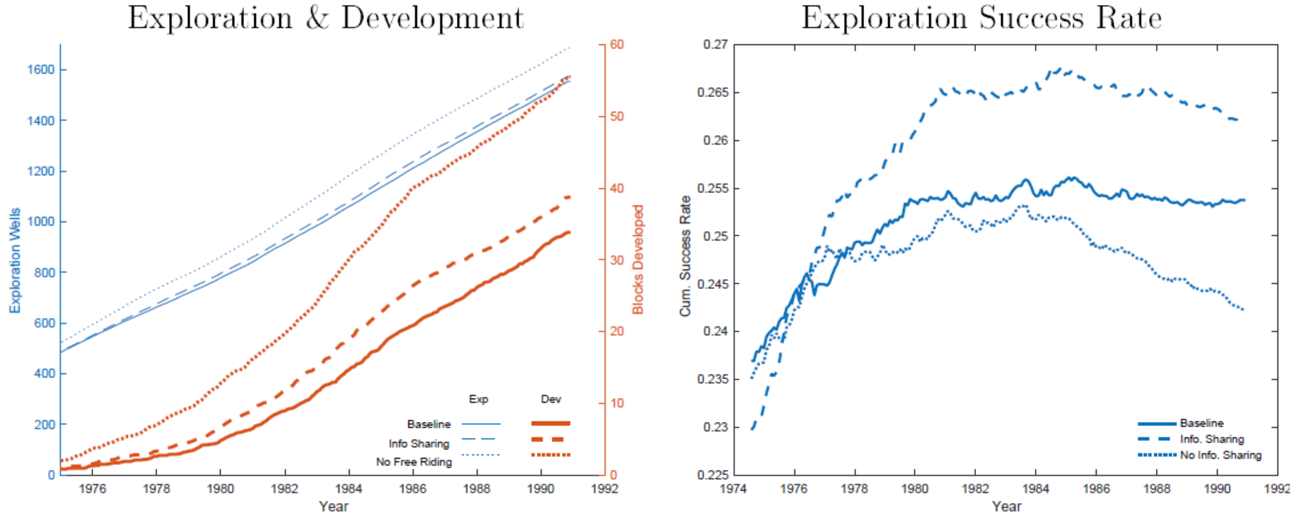
Parameter	Estimate	SE	Parameter	Estimate	SE
α	0.64	0.017	Oil Price Process		
Exploration Cost	0.141	0.054	Intercept	0.302	0.208
σ_c	0.036	0.017	P_{t-1}	0.988	0.008
Development Cost			Variance of Shock	3.323	1.147
High Variance Locations	2.022	0.651			
Low Variance Locations	1.043	0.4148			
σ_κ	0.252	0.111			
Average Realized Costs					
$E(c(j,\mathcal{S}_{ft})\|a_t^E = j)$	0.036		$E(\kappa(j,\mathcal{S}_{ft})\|a_t^E = j)$	0.870	
Comparison to $\alpha = 1$ Model					
λ_{LR}	53.054		$C_{0.95}$	3.841	

Table 5: Outcomes of Counterfactual Simulations

	Baseline	No Free Riding	Info Sharing	No Info Sharing	Info Sharing (Equilibrium)	No Price Shocks	Monopoly	Clustered
Exp. Wells	1554.35	1674.65	1574.45	1539.85	1556.60	1553.95	2202.30	1553.95
	[6.46]	[14.68]	[7.98]	[6.92]	[9.65]	[6.50]	[10.17]	[6.50]
Blocks Dev.	33.70	56.15	39.25	26.50	36.85	38.55	89.40	38.55
	[1.19]	[2.27]	[1.99]	[1.69]	[1.74]	[2.09]	[2.89]	[2.09]
Exp. Wells/Dev.	47.42	30.81	42.12	63.89	44.73	42.94	25.26	42.94
	[1.91]	[1.28]	[2.06]	[4.95]	[2.82]	[2.57]	[0.99]	[2.57]
Total Surplus								
Discounted	5.35	9.27	6.96	3.23	4.72	5.57	13.02	5.57
	[0.37]	[0.77]	[0.59]	[0.35]	[0.44]	[0.55]	[0.86]	[0.55]
Not Discounted	42.41	64.52	52.27	30.16	39.79	44.87	81.56	44.87
	[2.44]	[4.66]	[3.75]	[2.60]	[2.90]	[3.86]	[4.70]	[3.86]

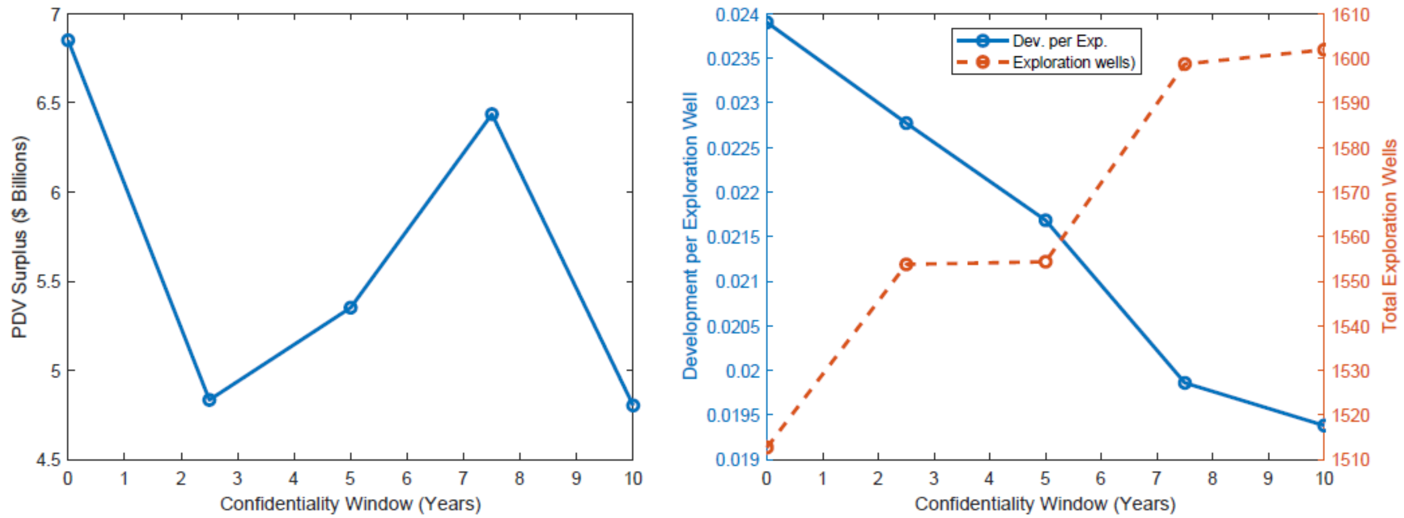
Notes: Results are averages over 40 simulations that cover 1964-1990. The assignment of blocks to firms and the oil price are set at their realized values. Well outcomes and development revenue are drawn from the posterior of the estimated Gaussian process using all observed wells. Revenue and profits are in billions of 2015 dollars. Total surplus is firm and government revenue less costs, including the value of cost shocks for every period. PDV revenue and profit are 1964 values where the annual discount factor is 0.9. Simulation standard errors are in square brackets.

Figure 5: Exploration Paths



Notes: The left panel plots the cumulative number of exploration wells drilled and blocks developed for each month from 1975 to 1990 for three simulations. Thick red lines plot the number of blocks developed and correspond to the right axis. Thin blue lines plot the number of exploration wells and correspond to the left axis. The solid lines are the average of 40 simulations using the baseline equilibrium choice probabilities. The dashed lines are the average of 40 simulations under the information sharing counterfactual. The dotted lines are the average of 40 simulations under the no free riding counterfactual. The right panel plots the cumulative success rate of exploration wells under the baseline,information sharing, and no information sharing counterfactuals.

Figure 6: Confidentiality Window



Notes: The left panel records the 1964 present discounted value of 1964-1990 industry surplus in counterfactual simulations with different confidentiality window lengths. In the right panel, the blue line, corresponding to the left y-axis, records the average number of developed blocks per exploration well evaluated using baseline choice probabilities under different confidentiality window lengths. The dashed red line, corresponding to the right y-axis, records the average number of exploration wells evaluated at counterfactual equilibrium choice probabilities under baseline information flow. Surplus is in billions of 2015 dollars. All figures are average over 40 simulations.