The Marginal Oil Field

Benini, Giacomo and Brandt, Adam and Dotti, Valerio and El-Houjeiri, Hassan

Department of Energy Resource Engineering, Stanford University, Department of Energy Resource Engineering, Stanford University, Department of Economics, Washington University in St. Louis, Climate and Sustainability Group, Aramco Research Center

February 2020

The Marginal Oil Field

Giacomo Benini[1], Adam Brandt[2], Valerio Dotti[3], and Hassan El-Houjeiri[4]

[1,2]Department of Energy Resource Engineering, Stanford University
[3]Department of Economics, Washington University in St. Louis
[4]Climate and Sustainability Group, Aramco Research Center

February 21, 2020

Abstract

The recent diffusion of novel oil technologies has increased the variability of petroleum resources. Today, it is possible to mine oil sands, to extract liquids from tight rocks and to produce high-viscosity oils. Using the Rystad dataset, we examine the sensitivity of 14343 deposits to a marginal change in oil prices or in marginal extraction costs. According to our estimates the variations in the crude properties combined with the value combined with the differences in the marginal extraction costs shift the (median) value of an extra barrel from $29.00 to $64.63 depending upon the type of oil. The range between these two extremes suggests that different oils could respond differently to common as well as specific shocks. Our findings are relevant for the design of Pigouvian taxes affecting the oil sector.

Keywords: Oil Economics, Shadow-Prices, Empirical Analysis of Firm Behaviour, Panel Data, Linear Mixed Models.
JEL Classification: L23, D22, C23, C14.
Corresponding Author: benini@stanford.edu

1 Introduction

The oil market connects tens of thousands of oil fields to billions of consumers via 2 million kilometres of pipelines and 500 millions dead-weight tons of merchant shipping (Cruz & Taylor, 2013). This global web has progressively transformed many regional markets into a worldwide pool of crude where riskless price arbitrages are unattainable (Adelman, 1984; Nordhaus, 2009).

The combined efforts of the logistic and of the financial sector have progressively removed resale opportunities, standardizing the global demand for oil (Milonas & Henker, 2001). Conversely, a set of technological advancements has profoundly diversified the supply of oil, widening the production possibilities of the petroleum sector (Maugeri, 2012). Today, it is possible to extract liquids trapped tightly in impermeable shale rocks, to mine heavy-oil-bearing sands, to explore the far Arctic and to access deep-water deposits located in the oceans depths (Gordon, Brandt, Bergerson, & Koomey, 2015). Consequently, day-to-day expenses as well as exploration investments have become increasingly heterogeneous across oil fields.

In order to account for the diversification of the oil supply, we construct a dynamic model of extraction and of exploration where the marginal extraction costs and the marginal discovery costs are deposit-specific. In our framework, firms simultaneously decide how much output to produce and how much exploration effort to expend. As a result, current and future levels of production and of exploration are function of the past amount of reserves (Pindyck, 1978; Devarajan & Fisher, 1982). Therefore, the shadow-price of discovered and of undiscovered oil is not given, but rather the result of the firms' inter-temporal choices. Consequently, the *hidden value* of an extra barrel is an endogenous variable of the model (Pesaran, 1990).

Using the Rystad Upstream Database (Rystad, 2018), we compute this unobserved variable identifying the shadow-prices of 75.70% of the global oil supply over the time interval 2014-2018. The calculation is a two step process. First, we approximate the price at which fields expect to sell their output. We do it fitting a pricing process which separates demand driven shocks from field-specific characteristics such as the density and the sweetness of the extracted liquids. Then, we estimate the marginal extraction costs. Subtracting these two quantities we identify the amount of money companies are willing to pay in order to manage an extra barrel of oil located in a particular field at a specific point in time. By the same token, we fit the marginal discovery costs and determine how many dollars a firm is willing to pay to discover an extra barrel of oil in an already producing field.

The results show that the value of a barrel is profoundly impacted by the oil's properties, the type of reservoir, the location of the deposit and its depletion rate. For some formations the differences between the expected and the shadow-price is negligible. For others it is significant. For example, in 2018, with an estimated average yearly price of \$67.01 per barrel, 2.49 million barrels (circa 2.51% of the global oil supply), spread across 104 fields, would have become unprofitable if their selling price would have been \$1 per barrel lower or if their marginal extraction costs would have been \$1 per barrel higher. These hundred and four fields represent the extensive margin of the industry. They are not homogeneously distributed across the globe. More precisely, 72.11% of the marginal deposits are Shale & Tight formations located in North America. This geographical concentration suggests that different regions could respond differently to common global trends.

To the best of our knowledge, this article is the first attempt to identify which formations will be the most sensitive to a marginal change in prices or in costs. We achieve this result proposing a theoretically consistent way to monetize how geology, marginal extraction costs, marginal discovery costs and shadow-prices are intertwined.

Our results tie a set of microeconometric estimates to relevant policy questions. For instance, they could be used to evaluate the impact of excise taxes, which are equivalent to a shock on the firm's unit production cost, on oil producers. Thus, our analysis can make an impact on the current debate on the use of Pigouvian taxes as an instrument to reduce carbon emissions (IMF, 2019).

2 Geologic Classes

There is no single, unique and unambiguous way to classify oil resources into a finite number of geologic classes. Each field is, in some sense, a unique geologic deposit and any effort to divide them will result in classes containing at least some ambiguous or confounding characteristics. However, it is still productive to divide oil resources into a broad (small-n) set of general resource classes.

Any classification scheme aiming to link geology and costs generally starts with a bifurcation between "conventional" and "unconventional" oil (George, 1998). For more than a century, oil has been conventionally extracted by vertically drilling into pourous and permeable reservoirs of free-flowing liquids. Unconventional resources can differ from this classical model either in the type of rock reservoir which contains the oil (i.e., tight rock or shale which have very low permeability), or in the type of oil extracted (i.e., heavy oil which does not flow freely).

The first class of unconventional oils – those in impermeable rocks – have been made accessible through technological innovation in horizontal drilling and hydraulic fracturing over the past fifteen years. These new methods generally use horizontal wells. These types of wells drill vertically until the actual deposit is reached. Then, in a second stage, they start drilling horizontally, for hundreds to thousands of meters, through the oil-containing impermeable rock (Gordon, 2012). The horizontal section of the well is often fractured, opening fissures in the rock which enable fluids to flow. These methods increase the per-well cost due to the large quantities of inputs used including water, proppants and chemicals (Torres, Yadav, & Khan, 2016).

The second class of unconventional oils – those with unconventional oil chemistry – were traditionally not producible due to the poor flow properties of the oil. Generally these are "heavy" oils with complex molecular structure and high viscosity (like a thick paste in the most extreme cases). Heavy oils require substantial initial investments as well as intensive use of steam to enable extraction (Tsui, 2010). Therefore, as a general rule of thumb, lighter petroleum exhibits lower marginal extraction costs than heavier oils.

Following a well established geological tradition, we divide fields with conventional rock properties according to their API gravity (Magoon, 1988). Any field containing oil with an API greater or equal than 22.3 is classified as Light & Medium crude. Oils in conventional reservoirs with an API between 20 and 22.3 are considered conventional-heavy. Any field with API below 20 is considered unconventional-heavy or Extra Heavy (Mommessin, Castano, Rankin, & Weiss, 1981). This would include the very heavy oils traditionally produced in California and Venezuela. These oils have high viscosity, high density and high concentrations of nitrogen, oxygen, sulphur, and heavy metals.

The second division is for fields with unconventional rock properties (Dandekar, 2013). Following a geochemical classificaiton, we divide these unconventional deposits into two major categories: sands and tight rock (Cook & Sherwood, 1991). The "sands" geologic class contains oil in loose or partially consolidated sands, with produced material containing a mixture of sand, clay and water. These are generally saturated with extra heavy oil (bitumen). In this case, the production costs become function of the bitumen's refining process. The traditional method is to mine the sands and subsequently upgrade the resulting extra heavy oil in order to make the final product lighter and therefore more valuable (Shah et al., 2010). More recently, it has become possible to heat the sands in-situ and avoid the upgrading.

The "tight rock" geologic class contains shale and tight formations. Within this category there are high degrees of diversity across the various formations. Wells in the tight rock class have a relatively short production life as well as a heterogeneous nature over relatively short distances (Mohr & Evans, 2010). Therefore, instead of looking for many subcategories, we group all of these fields under the label Shale & Tight. Figure (1) synthesizes the classification just described.

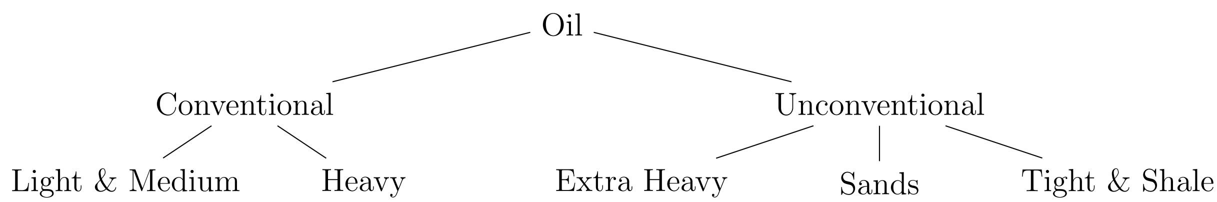

Figure 1: A geological and chemical classification of oil.

Oil fields classified in any of the categories listed in Figure (1) can be located either onshore or offshore, with the exception of those in the "sands" geological class, which are only found onshore. Most of the offshore deposits are located near the cost on the continental shelf. However, some deposits are situated at a considerable distance from the seaboard, beyond the shelf, in mid or deep-water. The location of the field impacts the marginal extraction and exploration process. Generally, both become higher as the location of the field becomes less accessible. Therefore, we combine the taxonomy presented in Figure (1) with four possible placements: onshore, on the shelf, off the shelf in mid-water and off the shelf in deep-water. As a result, we generate a 5×4 categorical variable, labelled *Geo*, which accounts for oil heterogeneity across the two dimensions.

3 Oil Shadow-Prices

In this Section we introduce a micro-econometric model which links the geological properties listed in Section 2 to the extraction and exploration decisions of a price-taking field.

3.1 Inter-Temporal Equilibrium

We assume that every field is managed by a price-taker risk-neutral firm i which exerts no market power. The firm decides in period t its production and investment plan for all periods $t+s$ with $s = 0, 1, 2, ...$[1] Its intra-temporal profits in each period $t + s$,

$$\Pi^i_{t+s} = P^i_{t+s} Q^i_{t+s} - C^i_{t+s}(Q^i_{t+s}, R^i, L^i_{t+s-1}, M^i_{t+s-1}, Geo^i, \epsilon^i_{t+s}) - W^i_{t+s} \ , \qquad (1)$$

are a function of the field's revenues, obtained multiplying the oil price at which field i sells its output P^i_{t+s} by the quantity of output produced Q^i_{t+s}, and of the extraction and exploration costs. The selling price is a function of the chemical characteristics of the crude produced. This assumption can be rationalized as the outcome of Bertrand competition between two or more buyers (typically refineries or intermediaries) whose evaluation of a barrel of oil depends upon its quality[2]. The costs are function of Q^i_{t+s} and of the quantity of reserves available when the production starts. The latter are equivalent to the initial size of the deposit R^i plus the discoveries occurred after the initial assessment of the field $L^i_{t+s-1} = L^i_0 + \sum_{r=1}^{t+s-1} D^i_r$, where D^i_r are the new discoveries in period r, minus the sum of extracted liquids $M^i_{t+s-1} = M^i_0 + \sum_{r=1}^{t+s-1} Q^i_r$. Finally, the costs are function of the peculiar geology of the field Geo^i, as identified in Section 2, and of an idiosyncratic shock ϵ^i_{t+s}. The exploration costs, W^i_{t+s}, are the expenses incurred to discover new oil located in field i.

Every field faces two physical constraints. The first one,

$$L^i_{t+s} \leq L^i_{t+s-1} + D^i_{t+s}(W^i_{t+s}, L^i_{t+s-1}, \xi^i_{t+s}), \qquad (2)$$

restrains the cumulative amount of discoveries at time $t + s$ to be lower or equal to the one obtained till time $t + s - 1$ plus the new ones $D^i_{t+s}(.)$[3]. We assume that the decision maker in period t possesses perfect foresight regarding the realization of the idiosyncratic additive error ξ^i_{t+s} in each period $t + s$.

[1] In other words, the firm is fully commits in period t to its future production and investment plans. While this assumption is admittedly unrealistic, it is often imposed in this class of models because it eases the derivation and interpretation of the results and has negligible consequences on the implications on the analysis (Pesaran, 1990; Favero, 1992; Favero & Pesaran, 1994).

[2] For instance, the output and the production cost of a refinery typically vary with the quality of crude used as input of the productive process.

[3] The inequality captures the implicit assumption that the firm is free to ignore/disregard some newly discovered oil in its assessment of total available reserves.

The second constraint ensures that the cumulative depletion exerted until $t+s-1$, denoted by M^i_{t+s-1}, plus the production at time $t+s$, equals or exceeds[4] the cumulative depletion at time $t+s$,

$$M^i_{t+s} \geq M^i_{t+s-1} + Q^i_{t+s} \ . \tag{3}$$

Each firm in period t decides the volumes of production, $Q^i_t, Q^i_{t+1}, \ldots$ and the rates of investment in exploratory effort $W^i_t, W^i_{t+1}, \ldots$ by maximizing the expected discounted future stream of profits. The decision is conditioned by the available information set Ω^i_{t-1} which includes previous prices, quantities and shocks,

$$\Omega^i_{t-1} = \left\{ \left[P^i_s\right]^{t-1}_{s=0},\ \left[Q^i_s, W^i_s, M^i_s, L^i_s\right]^{t-1}_{s=0},\ \left[\epsilon^i_s, \xi^i_s\right]^{t-1}_{s=0} \right\} \ .$$

The resulting inter-temporal problem,

$$\max_{\{Q^i_{t+s}, W^i_{t+s}, M^i_{t+s}, L^i_{t+s}\}^{\infty}_{s=0} \in X} \mathbb{E}_{t-1} \left\{ \sum_{s=0}^{\infty} \kappa^s \Pi^i_{t+s}(P^i_{t+s}, Q^i_{t+s}, W^i_{t+s}, M^i_{t+s-1}, L^i_{t+s-1}, Geo^i, \epsilon^i_{t+s}) \middle| \Omega^i_{t-1} \right\}$$

$$s.t. \left\{ \begin{array}{c} -M^i_{t+s} + M^i_{t+s-1} + Q^i_{t+s} \leq 0 \\ L^i_{t+s} - L^i_{t+s-1} - D^i_{t+s} \leq 0 \end{array} \right\}^{\infty}_{s=0} ,$$

where $X = \{[0, +\infty)^4\}^{\infty}_{s=1}$, can be solved using standard methods[5]. Specifically, the Lagrangian is

$$\mathcal{L}^i_t = \mathbb{E}_{t-1} \Bigg\{ \sum_{s=0}^{\infty} \kappa^s \big[\Pi^i_{t+s} + \lambda^i_{t+s}[M^i_{t+s} - M^i_{t+s-1} - Q^i_{t+s}] + \\ + \mu^i_{t+s}\big[L^i_{t+s-1} + D^i_{t+s} - L^i_{t+s}\big]\big] \big| \Omega^i_{t+s-1} \Bigg\} ,$$

where $0 \leq \kappa < 1$ is the inter-temporal discount factor.

The following restrictions on $C^i_{t+s}(.)$ and on $D^i_{t+s}(.)$,

$$\begin{array}{lll}
\frac{\partial^2 C^i_{t+s}(.)}{\partial (Q^i_{t+s})^2} > 0 & \frac{\partial^2 C^i_{t+s}(.)}{\partial (M^i_{t+s-1})^2} > 0 & \frac{\partial^2 C^i_{t+s}(.)}{\partial (Q^i_{t+s})^2} \frac{\partial^2 C^i_{t+s}(.)}{\partial (M^i_{t+s-1})^2} - \left(\frac{\partial^2 C^i_{t+s}(.)}{\partial Q^i_{t+s} \partial M^i_{t+s-1}} \right)^2 \geq 0 \\
\frac{\partial^2 C^i_{t+s}(.)}{\partial (L^i_{t+s-1})^2} > 0 & \frac{\partial^2 C^i_{t+s}(.)}{\partial M^i_{t+s-1} \partial L^i_{t+s-1}} = 0 & \frac{\partial^2 C^i_{t+s}(.)}{\partial (Q^i_{t+s})^2} \frac{\partial^2 C^i_{t+s}(.)}{\partial (L^i_{t+s-1})^2} - \left(\frac{\partial^2 C^i_{t+s}(.)}{\partial Q^i_{t+s} \partial L^i_{t+s-1}} \right)^2 \geq 0 \\
\frac{\partial^2 D^i_{t+s}(.)}{\partial (W^i_{t+s})^2} < 0 & \frac{\partial^2 D^i_{t+s}(.)}{\partial (L^i_{t+s-1})^2} < 0 & \frac{\partial^2 D^i_{t+s}(.)}{\partial (W^i_{t+s})^2} \frac{\partial^2 D^i_{t+s}(.)}{\partial (L^i_{t+s-1})^2} - \left(\frac{\partial^2 D^i_{t+s}(.)}{\partial W^i_{t+s} \partial L^i_{t+s-1}} \right)^2 \geq 0 ,
\end{array}$$

[4]The inequality captures the implicit assumption that the firm can dispose of extracted oil for free.

[5]Note that we are not explicitly accounting the presence of a natural capacity limit for each oil field. This assumption is mostly innocuous if production costs are sufficiently convex, such that marginal production costs become large in the proximity of the capacity limit and the optimal production level is always lower than its natural upper bound.

ensure that each firm solves a convex optimization problem over the compact set X. Thus, the solution of the optimization problem exists and is a global maximum[6]. Lastly, the assumptions on X imply that each choice variable is only bounded from below at zero. Thus, if the observed choices in period t satisfy $(Q_t^i, W_t^i, M_t^i, L_t^i)^* \gg 0$, then the global maximum is an interior solution for $(Q_t^i, W_t^i, M_t^i, L_t^i)^*$. As a result, the first-order necessary conditions of the optimization problem are satisfied with strict equality.

We use the equilibrium conditions to calculate the shadow-prices. The one of already discovered oil,

$$\mathbb{E}_{t-1}[\lambda_t^i|\Omega_{t-1}^i] = \mathbb{E}_{t-1}[P_t^i|\Omega_{t-1}^i] - \mathbb{E}_{t-1}\left[\frac{\partial C_t^i(.)}{\partial Q_t^i}\bigg|\Omega_{t-1}^i\right], \tag{4}$$

is heterogeneous across two dimensions: the value of the output produced and the marginal costs of extracting it. In the same way, the shadow-price of undiscovered oil,

$$\mathbb{E}_{t-1}[\mu_t^i|\Omega_{t-1}^i] = \mathbb{E}_{t-1}\left[\left(\frac{\partial D_t^i(.)}{\partial W_t^i}\right)^{-1}\bigg|\Omega_{t-1}^i\right], \tag{5}$$

varies as long as the marginal impact of W on D differs across fields[7].

As a by-product of the optimization, we obtain the laws of motion of the shadow-price of discovered,

$$\mathbb{E}_{t-1}[\lambda_{t+1}^i|\Omega_{t-1}^i] = \mathbb{E}_{t-1}\left[\frac{\lambda_t^i}{\kappa} - \frac{\partial C_{t+1}^i(.)}{\partial M_t^i}\bigg|\Omega_{t-1}^i\right], \tag{6}$$

and undiscovered oil[8],

$$\mathbb{E}_{t-1}[\mu_{t+1}^i|\Omega_{t-1}^i] = \mathbb{E}_{t-1}\left[\left(\frac{\mu_t^i}{\kappa} + \frac{\partial C_{t+1}^i(.)}{\partial L_t^i}\right)\left(\frac{\partial D_{t+1}^i(.)}{\partial L_t^i} + 1\right)^{-1}\bigg|\Omega_{t-1}^i\right]. \tag{7}$$

These two equations allow to describe how λ and μ evolve over time linking possible shocks occurred in period t, for instance an unexpected change in prices, to future extraction and investment decision of each firm[9].

[6]See the Mathematical Appendix for further details.

[7]Equation (5) is derived assuming that the idiosyncratic error ξ_{t+s} enters the discovery function additively. For further details see Section 3.2.

[8]See footnote 7.

[9]Equations (4) and (5) can be used to derive the firm supply and exploration function. The estimation of these equations goes beyond the scope of the paper, since we are not interested in estimating an extraction-exploration equilibrium resulting from a market made only by price-takers, but rather in the monetary value of discovered and undiscovered oil.

3.2 Heterogeneous Marginal Effects

In order to reverse-engineer the unobserved dependent variables of equations (4)-(5), we need to estimate the cost function and the discovery function. The former is often computed regressing C_t^i on (Q_t^i, R_{t-1}^i) where R_{t-1}^i is defined as the amount of recoverable reserves available when the decision to extract Q_t^i is taken (Uhler, 1976, 1979a, 1979b; Livernois & Uhler, 1987). However, since in our model the magnitude of $R_{t-1}^i = R^i + L_{t-1}^i - M_{t-1}^i$ is function of past extraction and exploration decisions $(Q_s^i, W_s^i)_{s=0}^{t-1}$ regressing C_t^i on R_{t-1}^i would cause an identification problem. More precisely, we would confound the initial stock effect, the discovery effect and the depletion effect. In order to disentangle them, we write a cost function,

$$\begin{aligned} C_t^i =& \theta_0^i + \left(\theta_1^i + \theta_2^{Geo} R^i + \theta_3^{Geo} \frac{L_{t-1}^i}{R^i} + \theta_4^{Geo} \frac{M_{t-1}^i}{R^i}\right) Q_t^i + \theta_5^{Geo} \frac{L_{t-1}^i}{R^i} + \\ &+ \theta_6^{Geo} \frac{M_{t-1}^i}{R^i} + \theta_7 Q_t^{i\ 2} + \theta_8 L_{t-1}^{i\ 2} + \theta_9 M_{t-1}^{i\ 2} + \epsilon_t^i \ , \end{aligned} \tag{8}$$

where their marginal impact is clearly separable. More precisely, in equation (8), the dependent variable C_t^i is the sum of Operating (OPEX) and of Capital Expenditures not linked to exploration (Non Exp CAPEX)[10] measured in Million US Dollars (MM \$) spent per Year. Q_t^i is the amount of output produced measured in Million Barrels of Oil Equivalent (MM BOE) extracted per Year[11]. R^i is the initial volume of recoverable reserves, L_{t-1}^i is the sum of the subsequent discoveries, M_{t-1}^i is the cumulative sum of MM BOE extracted. All three variables are measured in MM BOE. Finally, the idiosyncratic shock is normally distributed with finite homoskedastic variance, $\epsilon_t^i \overset{iid}{\sim} \mathcal{N}(0, \sigma_\epsilon^2)$.

Equation (8) contains two field-level effects, (θ_0^i, θ_1^i), five geology-level effects, $(\theta_2^{Geo}, \theta_3^{Geo}, \theta_4^{Geo}, \theta_5^{Geo}, \theta_6^{Geo})$ and three population-level effects $(\theta_7, \theta_8, \theta_9)$. The first two coefficients (θ_0^i, θ_1^i) identify respectively the fixed costs and the variable costs linked to production. They are both expected to be field-specific since fixed costs like compliance to regulation, geological survey or the price of land as well as variable costs like the amount of drilling activity or the labour costs vary greatly across

[10] OPEX includes expenditures like accounting, license fees, maintenance, repairs, office expenses, utilities and insurance, while the CAPEX expenditures (not linked to the discovery process) comprehend the installation, acquisition, repairing, upgrading and restoring of the physical assets used to extract oil.

[11] The decision to use BOE, rather than the traditional Barrel (B), allows to sum the production of condensate, gas, natural gas liquids (NGL) and oil, so to compare the marginal costs of fields with a different composition of the output. For example, the BOE allows to confront the marginal costs of Sands formations which produce almost only oil with the one of Shale & Tight accumulations which produce considerable quantities of NGL and associated gases.

fields. At the same time, they are both expected to be positive. The coefficients $(\theta_2^{Geo}, \theta_3^{Geo}, \theta_4^{Geo})$ represent the interaction between the current volume of reserves and the amount of output produced. They vary according to the geological variable described in Section 2. θ_2^{Geo} is expected to be negative since fields with more abundant initial reserves face smaller marginal costs (Pearce & Turner, 1990). The same is true for θ_3^{Geo} since more discoveries at time $t-1$ mean a larger reserve pool at time t and hence lower marginal costs. θ_4^{Geo} should be positive since more depleted fields experience higher marginal costs (Heal, 1976). $(\theta_5^{Geo}, \theta_6^{Geo})$ are expected to be respectively positive and negative. Lastly, the three population-level effects identify the rising marginal costs of production and depletion, and the decreasing cost advantage of a marginal increase in reserves. The parameter θ_7 is expected to be positive since extracting the next barrel is usually more expensive than extracting the previous one. By the same token, θ_8 and θ_9 are expected to be positive, in order to ensure that the cost function is convex.

The resulting expected shadow-price of discovered oil,

$$\begin{aligned}\mathbb{E}_{t-1}[\lambda_t^i|\Omega_{t-1}^i] =& \mathbb{E}_{t-1}[P_t^i|\Omega_{t-1}^i] - \theta_1^i - \theta_2^{Geo} R^i + \\ & - \theta_3^{Geo} \frac{L_{t-1}^i}{R^i} - \theta_4^{Geo} \frac{M_{t-1}^i}{R^i} - 2\theta_7 Q_t^i \ ,\end{aligned} \tag{9}$$

is a linear combination of field-specific, geology-specific and population-level returns. In order to see how the different components impact the magnitude of λ_t^i let's consider the following example. Suppose that two fields (i, j), both located onshore, both hosting 120 MM BOE of Heavy oil, decide at time t to extract one BOE more than in $t-1$. Then their marginal costs might differ. This could be the result of heterogeneous labor or energy expenditures $(\theta_1^i \neq \theta_1^j)$. However, if the two fields, which initially hosted the same amount of oil, are equally depleted, then the contribution of the amount of recoverable reserves on $\partial C(.)_t^i / \partial Q_t^i$ should be the same once we control for *Geo*.

We make the volume of discoveries a quadratic function of current exploration expenditures W_t^i and of total past discoveries L_{t-1}^i. The former are the product of the number of exploratory wells and the per-well cost. The latter identify the exploration history of different deposits. The resulting discovery equation,

$$D_t^i = \gamma_0^i + \gamma_1^i W_t^i + \gamma_2^i W_t^{i\ 2} + \gamma_3 L_{t-1}^i + \gamma_4 L_{t-1}^{i\ 2} + \xi_t^i \ , \tag{10}$$

makes D_t^i function of (W_t^i, L_{t-1}^i) and of an idiosyncratic shock ξ_t^i independently distributed from past and present production and exploration decisions[12]. In equation

[12]We assume that the realization of ξ_t is known by the firm – but unobservable to the econo-

(10), D_t^i is measured in MM BOE discovered in one year, W_t^i is the Exploration CAPEX measured in MM $ spent per Year, while L_{t-1}^i equals the sum of past findings measured in MM BOE.

Equation (10) contains three group-level effects $(\gamma_0^i, \gamma_1^i, \gamma_2^i)$ and two population-level effects (γ_3, γ_4). The first two coefficients identify the link between exploration expenditures and amount of discoveries. γ_1^i is expect to be positive. The more a firm invests the more it discovers new oil. To the contrary, γ_2^i is expected to be negative since marginal discoveries are declining in exploration CAPEX. The two population-level coefficients are expected to be negative. γ_3 should be negative since the more oil has been discovered in a field the less likely is to find new one. γ_4 is expected to be negative since marginal discoveries are decreasing in cumulative past levels of discoveries. Lastly, a natural concern is that equation (10) may take negative values as L_{t-1} grows large. While this is admittedly a non-trivial theoretical issue, it happens to be less of a concern with respect to the goals of this paper. Specifically, our analysis makes use of the derivatives of D_t^i with respect to W_t^i and L_{t-1}^i[13]. Thus, concerns about the latter are unlikely to matter for our predictions, and even less likely within the range of empirically relevant values of the covariates. Thus, we maintain the simple functional form of (10) for a matter of convenience[14] As a result, the expected shadow-price of undiscovered oil becomes

$$\mathbb{E}_{t-1}[\mu_t^i|\Omega_{t-1}^i] = \mathbb{E}_{t-1}\left[\frac{1}{\gamma_1^i + 2\gamma_2^i W_t^i}\middle|\Omega_{t-1}^i\right] . \tag{11}$$

Like in the case of λ_t^i, μ_t^i is measured in $/BOE. Its magnitude identifies how many dollars a firm is willing to pay to discover one more barrel of oil in an already producing deposit.

Having derived the closed form of the cost (C) and of the discovery function (D),

---

metrician – in order to ensure that the optimal choice always lies on the boundary of the feasible set. Note that the additive form of the discovery function implies that the realization of ξ_t does not affect the decision maker's optimal choice of investment in new discoveries. Thus, this assumption does not directly affect any key trade-off faced by the firm.

[13]Consistently with this goal, we adopt an empirical strategy which does not identify the parameter γ_0^i. Thus, out-of-sample predictions about the level of new discoveries are not possible. See Section 4.2 for further details.

[14]A possible interpretation of D_t^i is that of the total net resources (effort, money, luck, and amount of residual oil to be discovered in the field) that represent the inputs of the discovery activity. Under this interpretation, discoveries should be a censored variable with a positive probability mass at 0, corresponding to all those observations for which the total net resources do not exceed the threshold $D_t^i = 0$. Because this censored feature is hardly observable in the data, we regard our specification as a reasonable approximation of this more realistic framework.

we are now able to estimate the shadow-price of discovered (λ) and undiscovered oil (μ).

4 Empirical Results

The Rystad UCube Data Tool contains historical data for 58040 fields (Rystad, 2018). Using as a time bracket the last two years of fully available data (2017-2018) we collect information for 13397 fields. Compared with the initial population this sub-sample seems rather small. However, a simple ratio between the two quantities would be misleading. Most of the world's oil is located in a small number of large fields (Sorrell & Speirs, 2009). For example, almost half of the global production comes from 110 fields, while the two thirds of the entire supply comes from 500 giant deposits (Sorrell, Speirs, Bentley, Brandt, & Miller, 2009). The used sub-sample contains virtually all these prominent fields, plus a significant number of smaller, but marginally important, deposits. The resulting snapshot captures an average of 82.88% of world's oil production over a two years time interval, those consituting a micro-panel (large cross-sectional dimension, small time dimension) (IEA, 2017). Tables 1 shows how the 13397 fields are divided across the geological categories listed in Section 2.

Table 1: Absolute Frequency of different Geological Formations

	On Shore	Shelf	Mid-Water	Deep-Water
Light & Medium	8446	2343	383	195
Heavy	571	73	0	0
Extra Heavy	457	55	0	0
Sands	7	0	0	0
Shale & Tight	867	0	0	0

4.1 Shadow-Price of Discovered Oil

In order to reverse-engineer the magnitude of λ_t^i we need to subtract from the firm-level expected price the marginal extraction cost. Both these values are unobserved and need to be estimated.

4.1.1 Firm Expected Prices

In order to calculate the shadow prices $\mathbb{E}_{t-1}[\lambda_t^i|\Omega_{t-1}^i]$, we need a measure of $\mathbb{E}_{t-1}[P_t^i|\Omega_{t-1}^i]$, i.e. the price at which oil producers expect to sell their output. This variable is unobservable. We do, however, observe actual prices and future prices of a variety of oil types (classes), which we exploit in order to construct an estimator of the variable of interest. The procedure is outlined in the next lines.

Firstly, the literature shows that the oil sale price is a function of the gravity (API) and of the sulphur content (S) of the extracted liquids (Lanza, Manera, & Giovannini, 2005; Fattouh, 2010). Thus, in order to quantify how these two chemical properties impact expectations, we write a pricing function,

$$P_t^i = \beta_0 + \beta_{1t} + \beta_2 API^i + \beta_3 S^i + \eta_t^i \ , \tag{12}$$

which explicitly includes them as explanatory variables. The coefficient β_0 identifies the component driven be time-invariant aggregate factors, while β_{1t} isolates common (demand driven) trends. Both coefficients are measured in \$/BOE. Finally, API^i measures the field's density and S^i the percentage of sulphur in the extracted liquids. Since both API^i and S^i are dimensionless quantities (even if API^i is often referred as a measure in degrees) the unit of account of β_2 and of β_3 is still \$/BOE. We assume $\mathbb{E}_{t-1}[\eta_t^i|\Omega_{t-1}^i] = \mathbb{E}_{t-1}[\eta_t^i|\Omega_{t-1}^{pub}] = 0$. As a result,

$$\mathbb{E}_{t-1}[P_t^i|\Omega_{t-1}^i] = \mathbb{E}_{t-1}[P_t^i|\Omega_{t-1}^{pub}] = \beta_0 + \beta_{1t} + \beta_2 API^i + \beta_3 S^i \ , \tag{13}$$

where $\mathbb{E}_{t-1}[P_t^i|\Omega_{t-1}^i]$ is the price expected by the management of field i given the available information Ω_{t-1}^i and $\mathbb{E}_{t-1}[P_t^i|\Omega_{t-1}^{pub}]$ is the expect price given the publicly available information Ω_{t-1}^{pub}.

Secondly, oils with similar characteristics and origin are typically grouped in an oil class. An oil class k is uniquely defined by its average (time-invariant) characteristics $API(k) = \sum_{i\in k} w^i API^i$ and $S(k) = \sum_{i\in k} w^i S^i$ for some time-invariant weights $\{w^i\}_{i=1}^{N(k)}$, where $N(k)$ is the number of members of that particular class and $\sum_{i\in k} w^i = 1$.

Let $P_t^F(k)$ denote the twelve-months future price, measured in US Dollars per Barrel of Oil Equivalent (\$/BOE), of one of these classes. We assume that the future price of oil of type k equals the expected weighted average price of such oil, given the publicly available information Ω_{t-1}^{pub}, such that $P_t^F(k) = \mathbb{E}_{t-1}[\sum_{i\in k} w_t^i P_t^i|\Omega_{t-1}^{pub}]$. The time-variant weights $\{w_t^i\}_{i=1}^{N(k)}$ identify the relative importance of the different fields belonging to class k in a specific period t[15] and are assumed to satisfy $\mathbb{E}_{t-1}[w_t^i|API(k), S(k)] = w^i$ for all $i = 1, 2, ..., N(k)$ and for all t. Using (13), the formula for the future price can be written as

$$P_t^F(k) = \beta_0 + \beta_{1t} + \beta_2 API(k) + \beta_3 S(k) + \varsigma(k) \ , \tag{14}$$

where $\varsigma(k) = \sum_{i\in k}(w_t^i - w^i)(\beta_2 API^i + \beta_3 S^i)$ satisfies $\mathbb{E}_{t-1}[\varsigma(k)|API(k), S(k)] = 0$, ensuring that the error term in (14) satisfies the strict exogeneity assumption[16].

We fit equation (14) collecting historical prices about twenty-three oil products. We obtain the prices using the open-source dataset provided by the Energy Information Administration (EIA, 2019) and the value of API and of S from the PSA Management and Services BV dataset (Management & BV, 2019). The resulting panel tracks the prices of the different products from 1979 till 2018 while recording their chemical characteristics. Table 2 provides the summary statistics of $P_t^F(k)$ and reports their API and S. Figure (2) shows how representative the products are across the $API - S$ spectrum and how they co-moved over the analysed time interval.

Equation (14) contains one group-level effect (β_{1t}) and three population-level effects ($\beta_0, \beta_2, \beta_3$). Among the different options to fit a heterogeneous random variable, Linear Mixed Models (LMM) are an extension of classical regressions able to capture diversified behaviour over time (Hildreth & Houck, 1968; P. Swamy & Mehta, 1979). They are designed to make the true unobserved dependent variable a function of two vector-valued random variables. The first one is the estimate response $\hat{P}_t^F(k)$, the second one is a vector of random coefficients $\mathcal{A}$. In our case the conditional distribution of $P_t^F(k)$, given the realization $\mathcal{A} = a$, is

$$(P|\mathcal{A} = a) \overset{iid}{\sim} \mathcal{N}(aZ_1 + \alpha^T Z_2,\ \sigma_P^2 V_1^{-1}) \ , \tag{15}$$

with $a = [\beta_{1t}]$, $Z_1 = [1]$, $\alpha = [\beta_0, \beta_2, \beta_3]$ and $Z_2 = [1, API, S]$. Its mean moves according to a linear combination of (Z_1, Z_2, a, α) and its homoskedastic variance σ_P^2 multiplies a diagonal matrix V_1 of known prior weights. The random coefficient β_{1t} is normally distributed $\mathcal{A} \overset{iid}{\sim} \mathcal{N}(0, \Sigma_1)$[17]. We estimate equation (15) using the `lme4` optimizer within the `lmer` package of the statistical software `R` (Bates, 2005; De Boeck et al., 2011).

[15]Since the expected production of each field in a given class may vary over time, the weights are assumed to be time-variant random variables.

[16]For instance, this condition is satisfied if $w_t^i = w^i + \upsilon_t^i$, where the random variable υ_t^i satisfies $\mathbb{E}_{t-1}[\upsilon_t^i|\{API^i, S^i\}_{i=1}^{N(k)}] = 0$.

[17]For a detailed description of the relationship between σ_P, V_1 and Σ_1 see Swamy (1970).

Table 2: Summary Statistics of $P_t^F(k)$ for different Oil Types

Oil Type	Country of Origin	Mean	SD	Min	Max	API	S
Arabian Light	Saudi Arabia	40.39	29.36	12.36	109.43	32.8	1.97
Arabian Medium	Saudi Arabia	40.70	29.34	10.86	107.12	30.2	2.59
Basrah Light	Iraq	76.11	25.70	39.90	106.93	30.5	2.90
Berri	Saudi Arabia	78.82	25.79	45.62	110.77	38.5	1.50
Bonny Light	Nigeria	42.21	30.84	13.62	117.70	33.4	0.16
Bow River Heavy	Canada	33.96	22.82	10.41	84.29	24.7	2.10
Brent Crude	United Kingdom	28.10	13.30	13.94	64.60	38.3	0.37
Cabinda	Angola	26.90	13.92	12.69	69.17	32.4	0.13
Forcados Blend	Nigeria	32.34	22.95	14.35	111.07	30.8	0.16
Furrial	Venezuela	18.27	4.26	12.24	28.23	30.0	1.06
Leona	Venezuela	20.98	9.36	9.79	51.55	24.0	1.50
Light Sour Blend	Canada	69.09	20.51	40.04	96.52	64.0	3.00
Lloydminster	Canada	33.88	23.95	10.15	82.50	20.9	3.50
Marlim	Brazil	78.42	27.83	47.77	114.32	19.6	0.67
Mayan	Mexico	36.01	27.09	9.21	100.29	21.8	3.33
Merey	Venezuela	72.31	24.94	38.97	103.28	15.0	2.70
Napo	Ecuador	70.78	25.76	37.46	101.53	19.0	2.00
Olmeca	Mexico	31.82	22.98	13.58	101.14	37.3	0.84
Oriente	Ecuador	39.10	27.57	11.55	105.50	24.1	1.51
Qua Iboe	Nigeria	99.73	22.16	68.26	117.02	36.3	0.14
Rabi-Kouanga	Gabon	33.79	23.38	13.65	95.46	37.7	0.15
Saharan Blend	Algeria	83.16	24.68	49.82	115.82	45.0	0.09
WTI	United States	42.30	27.68	14.34	99.56	39.6	0.24

The great flexibility of $\hat{\beta}_{1t}$, which shifts from a minimum of -27.43 to a maximum of 64.09, see Figure 3, allows to have a good fit of the dependent variable. For example, the adjusted R^2 suggests that equation (15) explains 98% of the prices' variance.

Combining the per-period value of $\hat{\beta}_{1t}$ with the estimated unconditional expected price, which is just above \$40/BOE ($\hat{\beta}_0 = 40.04$), allows to identify the impact of API and S on $\mathbb{E}_{t-1}[P_t^{i,k}|\Omega_{t-1}^{i,k}]$. An increase of one degree of API augments the value of the crude by \$0.07/BOE ($\hat{\beta}_2 = 0.07$). In the same way, an increase of 1% in the sulphur content decreases the value of the crude by \$2.14/BOE ($\hat{\beta}_3 = -2.14$). All these results are robust to the introduction of an oil type fixed effect, a country fixed-effect, a time trend or a combination of the three. Furthermore, a series of Analysis of the Variance (ANOVA) tests suggest that the performance of the different models does not change in a statistically significant way. Therefore we use the estimates obtained from (14), $\hat{\beta} = [\hat{\beta}_0, (\hat{\beta}_{1t})_{t=1}^T, \hat{\beta}_2, \hat{\beta}_3]$, to obtain an estimator $\hat{P}_t^i$ of the field level expected prices $\mathbb{E}_{t-1}[P_t^i|\Omega_{t-1}^i]$, which writes

$$\hat{P}_t^i = \hat{\beta}_0 + \hat{\beta}_{1t} + \hat{\beta}_2 API^i + \hat{\beta}_3 S^i \; . \tag{16}$$

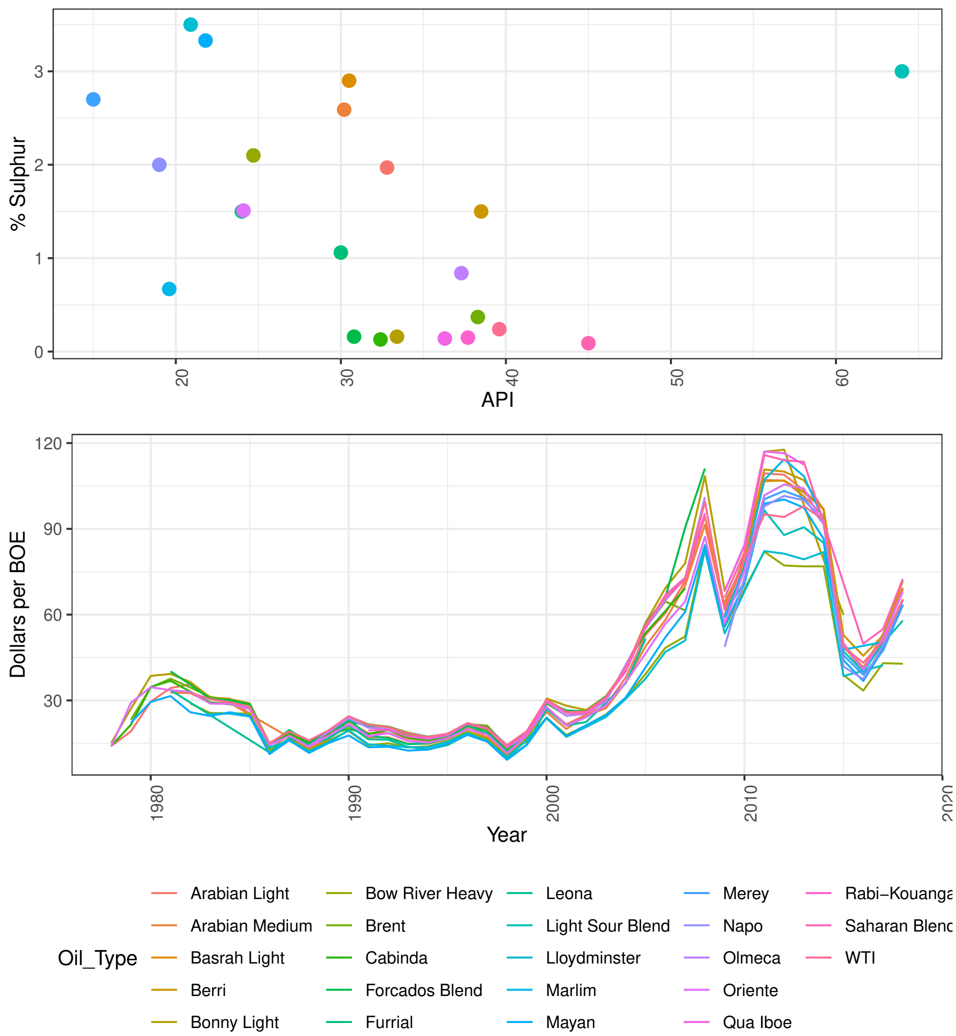

Figure 2: API and Sulphur content of twenty-three Oil Products and their value over time.

Notice that under the assumption stated in this section $\hat{P}_t^i$ is an unbiased estimator of $\mathbb{E}_{t-1}[P_t^i|\Omega_{t-1}^i$ as long as $\hat{\beta}$ is an unbiased estimator of β. Figure (**??**) shows the resulting mapping for the year 2018.

It is interesting to note that in a year where the two most representative oil prices, the West Texas Intermediate and the Brent Crude, where trading above \$60 per barrel ($P_t^F(\text{WTI}) = 64.94, P_t^F(\text{Brent}) = 71.06$), 9922 fields (68.85% of the entire sample) were expected to sell their output below \$60 per barrel. This includes all the seven sands formations reported in Table 1 and 580 out of the 869 Shale & Tight deposits.

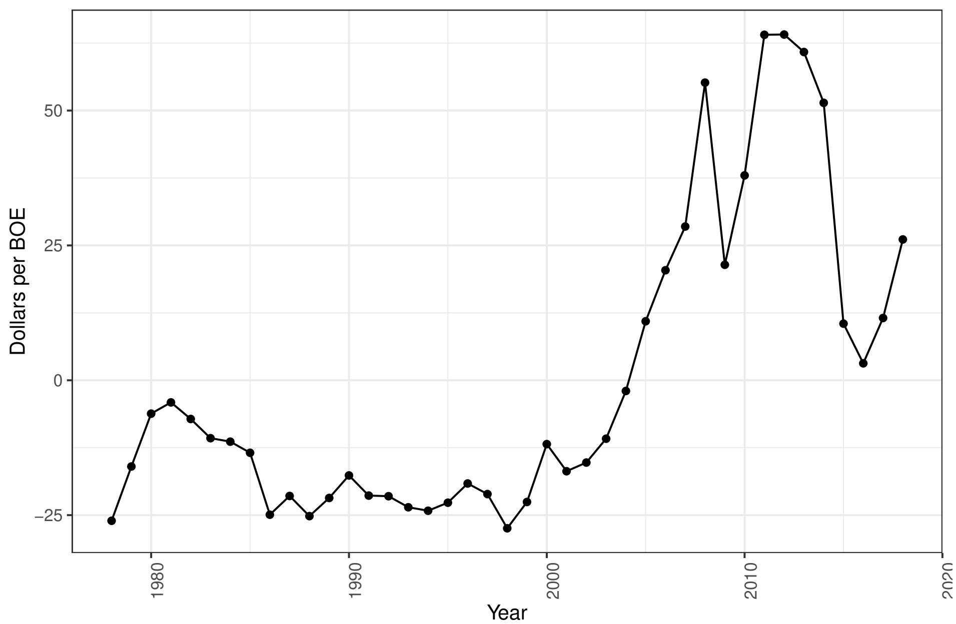

Figure 3: Evolution of $\hat{\beta}_{1t}$ over the analysed time interval.

4.1.2 Marginal Exctraction Costs

Once the first element of (4) has been computed, we need to estimate the cost function. Table 3 provides the summary statistics of the continuous variables required to fit it[18].

Table 3: Summary Statistics of the Continuous Variables

Statistic	Variable	Unit of Account	N	Mean	SD	Min	Max
Costs	C	MM $ Year	25912	48.75	194.54	0.00	6256.68
Volumes	Q	MM BOE Year	25912	2.33	15.30	0.00	988.00
Initial Reserves	R	MM BOE	25912	240.15	1671.13	0.00	94546.05
Development Rate	L/R	Real Number	25912	1.02	1.74	0.00	94.92
Depletion Rate	M/R	Real Number	25912	0.48	0.32	0.00	1.00

An explanatory analysis of the dependent variable C suggests that costs and production co-move. More precisely, they both decline as the field gets older. This

[18]We restrict our attention to active fields ($Q > 0$). Since some fields were either discovered either abandoned during the chosen time interval, we work with an unbalanced dataset. More precisely, a balanced dataset would have had $13397 * 2 = 26794$ data points while our has 25912 data points. Therefore, 3.29% of our sample is not observed at least in one of the two years studied.

is true at a field level as well as at an aggregate (time series) level, see Figure 4. Furthermore, costs and production are strongly autocorrelated. For example, a Covariate Augmented Dickey-Fuller (CADF) test, where $C_t = \sum_i C_t^i$ is regressed on $\Delta Q_t = \sum_i \Delta Q_t^i$, suggests the presence of a unit root (Lupi et al., 2009). The same is true when using a Panel Covariate Augmented Dickey-Fuller test which regresses C_t^i on ΔQ_t^i (Kleiber & Lupi, 2011; Costantini & Lupi, 2013).

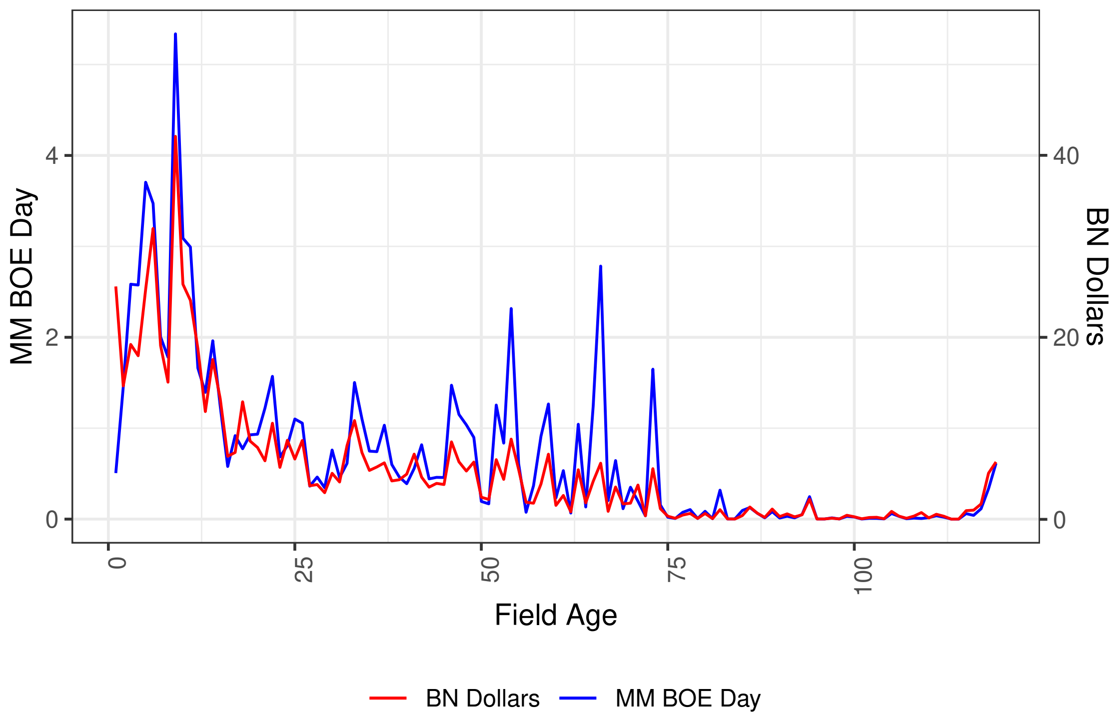

Figure 4: Co-trending behaviour of Costs and Production during the Field Age.

In order to attenuate the unit root problem, we rewrite the cost function in first differences,

$$\begin{aligned}\Delta C_t^i =& \theta_1^i \Delta Q_t^i + \theta_2^{Geo} \Delta Q_t^i R^i + \theta_3^{Geo}\left(Q_t^i \frac{L_{t-1}^i}{R^i} - Q_{t-1}^i \frac{L_{t-2}^i}{R^i}\right) + \\ &+ \theta_4^{Geo}\left(Q_t^i \frac{M_{t-1}^i}{R^i} - Q_{t-1}^i \frac{M_{t-2}^i}{R^i}\right) + \theta_5^{Geo}\left(\frac{L_{t-1}^i}{R^i} - \frac{L_{t-2}^i}{R^i}\right) + \\ &+ \theta_6^{Geo}\left(\frac{M_{t-1}^i}{R^i} - \frac{M_{t-2}^i}{R^i}\right) + \theta_7 \Delta Q_t^{i\ 2} + \theta_8 \Delta L_{t-1}^{i\ 2} + \theta_9 \Delta M_{t-1}^{i\ 2} + \Delta \epsilon_t^i \ . \end{aligned} \tag{17}$$

The estimation of equation (8) in first differences does not only reduce the co-integration problem, but it also transforms the distribution of the dependent variable. More precisely, the empirical probability density function of C_t^i is always positive, highly over-dispersed ($\mathbb{V}[C] \gg \mathbb{E}[C] > 0$) and, contrary to the one of P, does not resemble the one of a normal distribution. Therefore fitting (8) requires the use of a generalized linear model where the dependent is allowed to be over-dispesed, for example a Quasi-Poisson or a Negative Binomial, and the link function guarantees $\mathbb{E}[C] > 0$ (Nelder & Wedderburn, 1972; Liang & Zeger, 1986). To the contrary, ΔC_t^i is normally distributed. More precisely, a Wilcoxon Signed-Rank test rejects the hypothesis that ΔC and 23962 simulated data points drown from a normal distribution with the same mean and the same variance of ΔC, namely $\Delta C^{\text{sim}} \overset{iid}{\sim} \mathcal{N}(0.94, 61.24)$, are drown from different distributions.

We estimate equation (17) using the same method and software used to fit (**??**). Namely, given the realization $\mathcal{D} = d$,

$$(\Delta C|\mathcal{D} = d) \overset{iid}{\sim} \mathcal{N}(d^T Z_3 + \delta^T Z_4,\ \sigma^2_{\Delta C} V_2^{-1}) \ , \tag{18}$$

with $d = [\theta_1^i, \theta_2^{Geo}, \theta_3^{Geo}, \theta_4^{Geo}, \theta_5^{Geo}, \theta_6^{Geo}]$, $Z_{3t} = [\Delta Q_t, ..., (M_{t-1}/R) - (M_{t-2}/R)]$, $\delta = [\theta_7, \theta_8, \theta_9]$ and $Z_{4t} = [\Delta Q_t^2, \Delta L_{t-1}^2, \Delta M_{t-1}^2]$. Using (18), we fit four regressions. The first one is equation (17). The second one is again equation (17) only this time the error term contains a country fixed effect, $\Delta\epsilon_t^i = \text{Country}^i + \varphi_t^i$ with $\varphi_t^i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2_\varphi)$. In the third one the error contains a time fixed effect, $\Delta\epsilon_t^i = t + \varphi_t^i$, and in the fourth one a country and a time fixed effect, $\Delta\epsilon_t^i = \text{Country}^i + t + \varphi_t^i$. Table 4 shows the obtained results for the two population coefficients.

Table 4: Estimated Population Parameters of the Cost Function

	Dependent Variable: Delta MM $ per Year			
$Q_t^{i\ 2}$	0.26***	0.26***	0.26***	0.26***
	(0.01)	(0.01)	0.01)	(0.01)
$L_{t-1}^{i\ 2}$	$1.75e^{06}$	$1.41e^{06}$	$1.77e^{06}$	$1.49e^{06}$
	(0.00)	(0.00)	(0.00)	(0.00)
$M_{t-1}^{i\ 2}$	$6.65e^{06}$ ***	$6.81e^{06}$ ***	$6.64e^{06}$ ***	$6.81e^{06}$ ***
	(0.00)	(0.00)	(0.00)	(0.00)
Time Fixed-Effects	No	No	Yes	Yes
Country Fixed-Effects	No	Yes	No	Yes
Adjusted R^2	0.57	0.57	0.57	0.57
Note:				*p<0.1; **p<0.05; ***p<0.01.

In all the presented specifications the obtained coefficients $\hat{\delta} = [\hat{\theta}_7, \hat{\theta}_8, \hat{\theta}_9]$ are positive. The sign of $\hat{\theta}_7$ suggests that marginal extraction costs increase in quantities. This is consistent with our assumptions and ensures that the firm's profit maximization problem is convex. $\hat{\theta}_8$ and $\hat{\theta}_9$ are also positive implying that marginal development costs are increasing in quantities discovered and marginal depletion costs are increasing in cumulative extractions. This last finding implies that even if costs tend to decline over time because they co-move with production, the marginal costs are inversely related to the age of the field. In other words, as the field gets depleted the cost of extracting the next barrel increases. Finally, the adjusted R^2 of the four regressions are not particularly impacted by the presence of the fixed effects. This last result suggests that is not the longitudinal nature of the dataset but rather the explanatory power of (Q, R, M, L) which capture circa the 57% of $\mathbb{V}[\Delta C_t^i]$.

The first random coefficient $\hat{\theta}_1$ has an expected value of $\mathbb{E}[\theta_1^i] = 0.25$. Therefore on average increasing production by 1 MM BOE per Year (2740 BOE per Day) translates into an increase in costs of \$250603.40 per Year (\$686.58 per Day).

The second random coefficient $\hat{\theta}_2^{Geo}$ has a negative sign for all formations, with the exception of Shale & Tight deposits, see Table 5. These first findings suggest that the traditional view of natural resource economics on the negative link between the initial stock of reserves and the marginal extraction costs is essentially correct with the exception of oil trapped in tight rocks.

Table 5: Impact of the Initial Reserves Stock on Marginal Costs

	On Shore	Shelf	Mid-Water	Deep-Water
Light & Medium	-0.004	-0.004	-0.025	-0.020
Heavy	-0.012	-0.001	NA	NA
Extra Heavy	-0.002	-0.002	NA	NA
Sands	-0.005	NA	NA	NA
Shale & Tight	0.016	NA	NA	NA

$\hat{\theta}_3^{Geo}$ is negative for all formations, except for Sands and Light & Medium oil located in mid-water, see Table 6. This means that in virtually all cases the addition of new discoveries decreases the marginal extraction costs.

Table 6: Impact of Discoveries on Marginal Costs

	On Shore	Shelf	Mid-Water	Deep-Water
Light & Medium	-1.447	-1.563	-8.594	-14.027
Heavy	1.588	11.050	NA	NA
Extra Heavy	7.382	-10.133	NA	NA
Sands	0.623	NA	NA	NA
Shale & Tight	-0.547	NA	NA	NA

Finally, $\hat{\theta}_4^{Geo}$ is positive for every type of deposit, see Table 7. This means that with no exception it becomes more costly to extract the next barrel as the field gets depleted.

Table 7: Impact of Depletion on Marginal Costs

	On Shore	Shelf	Mid-Water	Deep-Water
Light & Medium	17.373	24.362	34.155	4.068
Heavy	16.548	7.557	NA	NA
Extra Heavy	31.797	33.264	NA	NA
Sands	5.330	NA	NA	NA
Shale & Tight	162.164	NA	NA	NA

It is interesting to notice that the magnitude of $\hat{\theta}_4^{\text{Shale \& Tight}}$ is much higher than any of the other formation. Therefore, not only the initial size of reserves does not lower marginal costs $\hat{\theta}_2^{\text{Shale \& Tight}} = 0.017$, but their depletion increases marginal costs five times more than Extra Heavy and ten times more than Light & Medium or Heavy when located onshore.

Combining these values the the estimates of Section 4.1.1, we are able to compute an estimate $\hat{\lambda}_t^i$ of the shadow-price of discovered oil $\mathbb{E}_{t-1}[\lambda_t^i|\Omega_{t-1}^i]$, which writes:

$$\hat{\lambda}_t^i = \hat{P}_t^i - \hat{\theta}_1^i - \hat{\theta}_2^{Geo} R_0^i - \hat{\theta}_3^{Geo} \frac{L_{t-1}^i}{R^i} - \hat{\theta}_4^{Geo} \frac{M_{t-1}^i}{R^i} - 2\hat{\theta}_7 Q_t^i \ ,$$

and identify those fields for which $\hat{\lambda}_t^i$ is relatively close to zero, namely *the marginal fields*. The plot of the 2018 empirical density function of $\hat{\lambda}_t^i$ divided by geological category suggest than $\mathbb{V}[\hat{\lambda}_t^i]$ varies greatly across different types of formations, see Figure 5. More precisely, Shale & Tight, Sands and all the formations offshore have a significant portion of their production which could be very sensitive to small changes in prices.

4.2 Shadow-Price of Undiscovered Oil

Out of the initial sample, we observe exploration activity in 1137 fields. Tables 8 plots their absolute frequencies, while Table 9 provides the summary statistics of the continuous variables required to fit equation (10).

The discovery function presents some of the problems highlighted by the cost function. More precisely, D is always positive and over-disperse. Furthermore, D and W are co-integrated, see Figure 6.

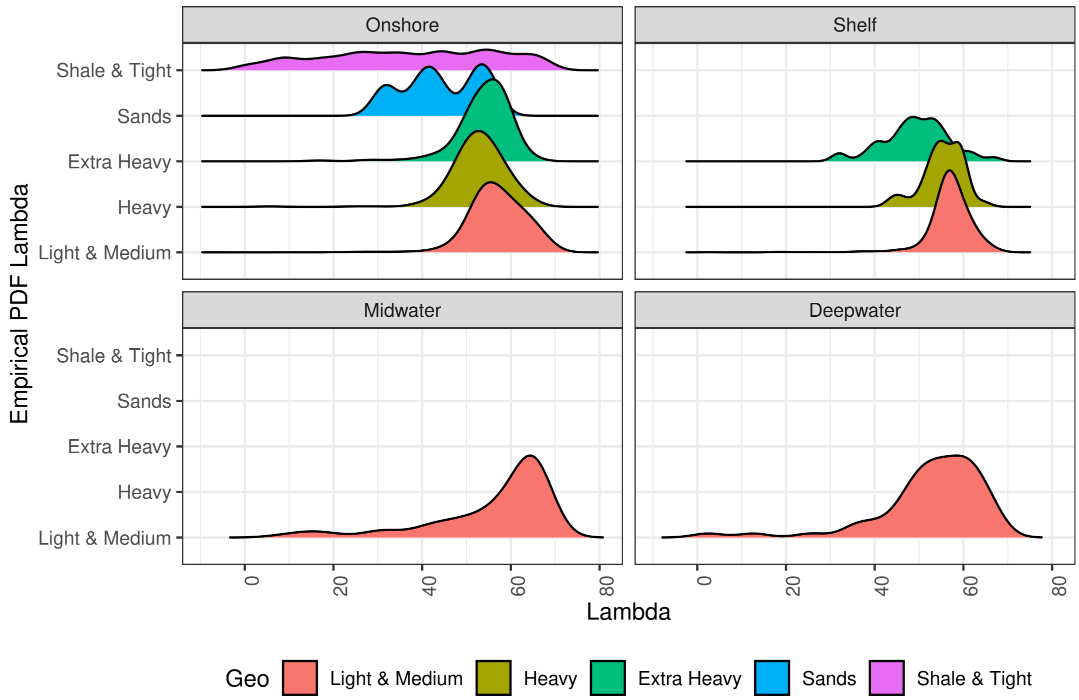

Figure 5: Empirical Probability Density Function of $\hat{\lambda}$.

Table 8: Absolute Frequency of different Geological Formations

	On Shore	Shelf	Mid-Water	Deep-Water
Light & Medium	375	72	0	0
Heavy	23	0	0	0
Extra Heavy	10	0	0	0
Sands	0	0	0	0
Shale & Tight	657	0	0	0

Table 9: Summary Statistics of the Continuous Variables

Statistic	Variable	Unit of Account	N	Mean	SD	Min	Max
Discoveries	D	MM BOE Year	1645	7.21	24.20	0.00	590.00
Expl. CAPEX	W	MM \$ Year	1645	12.30	28.74	0.00	350.90
Cum Discoveries	L	MM BOE	1645	74.18	230.62	0.00	5520.80

In order to solve the first problem and attenuate the second, we rewrite equation

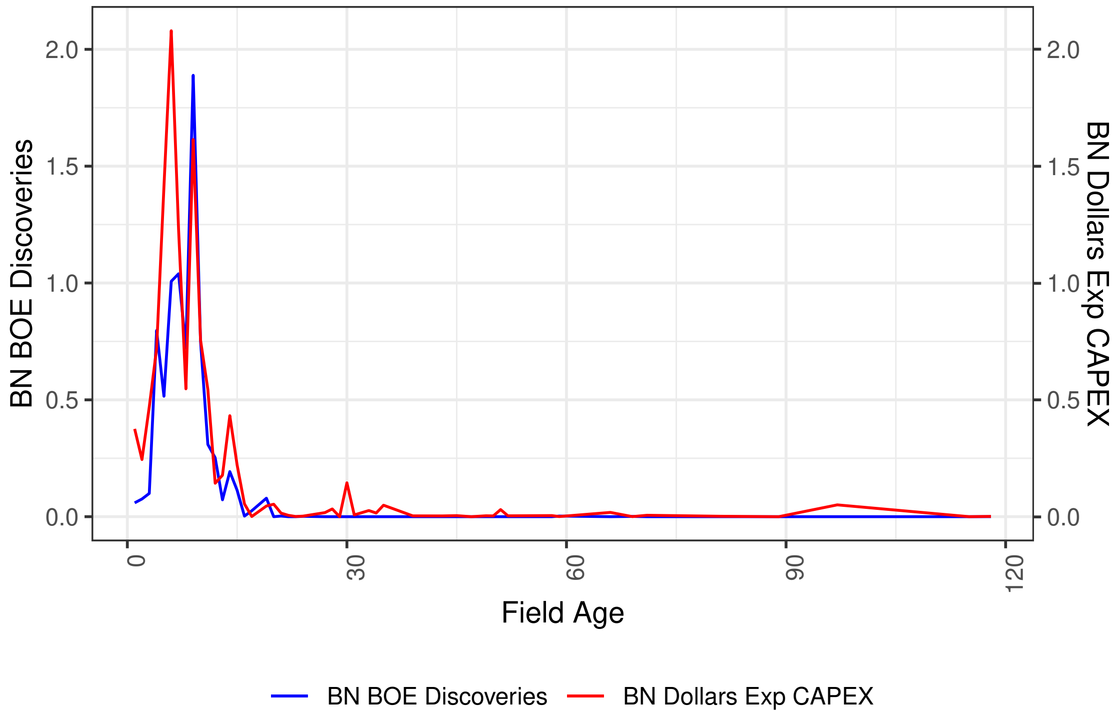

Figure 6: Co-trending behaviour of Discoveries and Exploration CAPEX during the Field Age.

(10) in first differences,

$$\Delta D_t^i = \gamma_1^i \Delta W_t^i + \gamma_2^i \Delta W_t^{i\ 2} + \gamma_3 \Delta L_{t-1}^i + \gamma_4 \Delta L_{t-1}^{i\ 2} + \Delta \xi_t^i \ , \tag{19}$$

Using the same distributional hypothesis presented in (15) and in (18), we estimate, for a given realization $\mathcal{E} = e$,

$$(\Delta D | \mathcal{E} = e) \overset{iid}{\sim} \mathcal{N}(i^T Z_5 + \eta^T Z_6,\ \sigma_{\Delta D}^2 V_3^{-1}) \ , \tag{20}$$

with $i = [\gamma_1^i, \gamma_2^{Geo}]$, $Z_{5t} = [\Delta W_t^i, \Delta W_t^{i\ 2}]$, $\eta = [\gamma_3, \gamma_4]$ and $Z_{6t} = [\Delta L_{t-1}^i, \Delta L_{t-1}^{i\ 2}]$. Like in the case of the costs function, we add to this initial estimation other three regressions which combine time and countries fixed effects, see Table 10.

Both population parameters are negative. The first one ensures that the more oil has been extracted over the curse of the field life the less likely is to find new oil. The second one implies that marginal discoveries decrease in past volumes of production. This last finding is consistent with the conditions required to ensure that each firm solves a convex optimization problem over a compact set. Furthermore, like in the case of the cost function, the adjusted R^2 does not change due

Table 10: Estimated Population Parameters of the Discovery Function

	Dependent Variable: Delta MM BOE per Year			
L^i_{t-1}	-0.50***	-0.68***	-0.60***	-0.67***
	(0.03)	(0.04)	(0.03)	(0.04)
$L^{i\ 2}_{t-1}$	-20.02e^{-5} ***	-80.57e^{-6} *	12.67e^{-5} ***	82.75e^{-6} **
	(0.00)	(0.00)	(0.00)	(0.00)
Time Fixed-Effects	No	No	Yes	Yes
Country Fixed-Effects	No	Yes	No	Yes
Adjusted R^2	0.59	0.60	0.58	0.60
Note:				*p<0.1; **p<0.05; ***p<0.01.

to the presence of the fixed effects. In all four cases the four explanatory variables capture circa 60% of the variance of the depended variable.

The random coefficient $\hat{\gamma}_1$ has an expected value of $\mathbb{E}[\gamma^i_1] = 0.01$. Therefore on average increasing production by 1 MM \$ translates into a discovery of 14138.71 BOE. The second random coefficient has an expected value of zero.

Contrary to the estimated shadow-prices of discovered oil, the ones of undiscovered oil,

$$\hat{\mu}^i_t = \frac{1}{\hat{\gamma}^i_1 + 2\hat{\gamma}^i_2 W^i_t} \ ,$$

tend to be on substantially lower than the market price, see Figure 7. Both onshore and shelf oil seem to value the next barrel between 10 and 33\$ per BOE with Extra Heavy being an exception with a mean value of 53.17 \$ per BOE. These numbers are below any of the market prices listed in Table 2. While the shadow-price of undiscovered oil theoretically should be lower than the on of discovered oil, the magnitude of the former is, on average, very small. The discrepancy between the two prices could be the result of the great uncertainty about the future price of oil. Furthermore, the link between D and (W, L) is not always straightforward especially when, link in our case, μ measures the value firms give to a new barrel in an already producing deposit. The discovery of new oil in an already operating field is very different from exploring an undeveloped area. Therefore, even if our estimates are not implausible, we recommend prudence in their use of the estimates shown in Figure 7 to extrapolate the values firms would give to undiscovered oil in an undiscovered field.

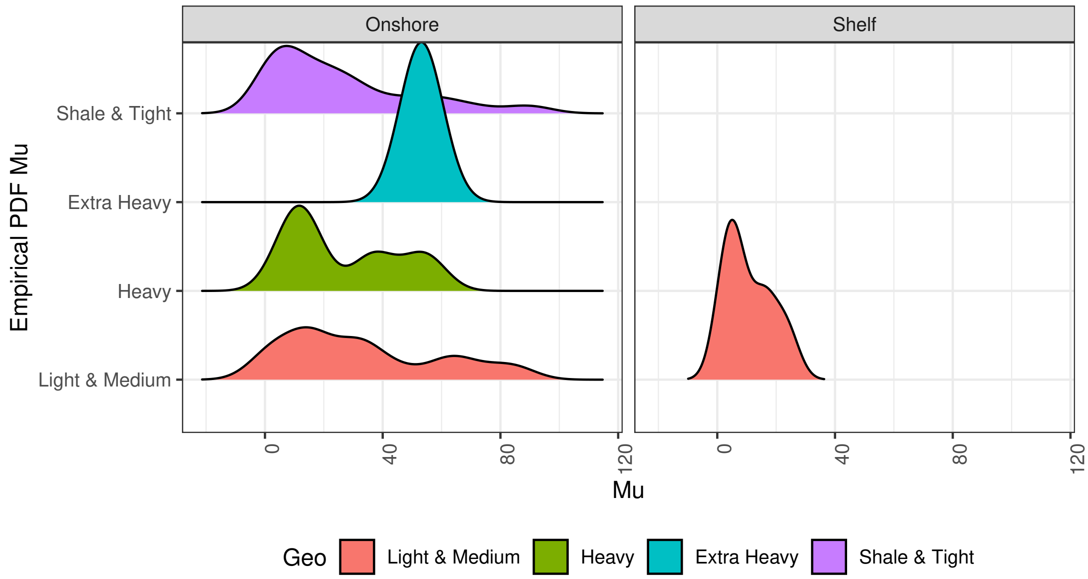

Figure 7: Empirical Probability Density Function of $\hat{\mu}$.

5 Conclusions

The present paper identifies an extraction-exploration equilibrium which singles out the shadow-price of discovered and of undiscovered oil. Using a commercial dataset (Rystad, 2018), we compute the magnitude of these two unknown values.

The resulting estimates suggest that different oils respond very differently to common shocks. For example, in the case of perceived long-run change in the oil price, the shadow-price of discovered and/or of undiscovered oil may become negative for some deposits, inducing the corresponding firms to change their production and discovery choices at the extensive margin and, eventually, to shut down. In turn, the market structure could be affected by the exit of these firms, with potential consequences on the expected prices and the production choices of active fields.

These findings are potentially relevant for policy. Specifically, they imply that the response to targeted taxes and subsidies is likely to be highly heterogeneous across fields with different geological characteristics. In particular, a uniform excise tax on oil production is likely to hit severely the production and investment choices of firms producing heavy and non-conventional oil, while it would have little effect on other fields.

Moreover, since marginal CO_2 emissions are highly heterogeneous across fields producing different types of oil (Masnadi et al., 2018), our findings could have important implication in the design of the optimal pigouvian taxes and/or of a tradable permit schemes aiming to tackle the production of CO_2. For instance, the effect of a pigouvian tax on carbon emissions caused by oil extraction may substantially differ from that of an excise tax on oil production, or that of a sales tax on fossil fuels. In turn, the heterogeneity documented in this paper may also affect the effectiveness of pigouvian taxes in reducing the externality.

Far from being fully exhaustive, the present paper has tried to open a path in the direction of a more all-inclusive approach toward an increasingly diversified oil industry.

References

Adelman, M. A. (1984). International oil agreements. *The Energy Journal*, *5*(3), 1–9.

Bates, D. (2005). Fitting linear mixed models in r. *R news*, *5*(1), 27–30.

Cook, A., & Sherwood, N. (1991). Classification of oil shales, coals and other organic-rich rocks. *Organic Geochemistry*, *17*(2), 211–222.

Costantini, M., & Lupi, C. (2013). A simple panel-cadf test for unit roots. *Oxford Bulletin of Economics and Statistics*, *75*(2), 276–296.

Cruz, J. M., & Taylor, M. S. (2013). *A spatial approach to energy economics* (Tech. Rep.). National Bureau of Economic Research.

Dandekar, A. Y. (2013). *Petroleum reservoir rock and fluid properties*. CRC press.

De Boeck, P., Bakker, M., Zwitser, R., Nivard, M., Hofman, A., Tuerlinckx, F., ... others (2011). The estimation of item response models with the lmer function from the lme4 package in r. *Journal of Statistical Software*, *39*(12), 1–28.

Devarajan, S., & Fisher, A. C. (1982). Exploration and scarcity. *Journal of Political Economy*, *90*(6), 1279–1290.

EIA. (2019). Eia. *EIA OPEN DATA*.

Fattouh, B. (2010). The dynamics of crude oil price differentials. *Energy Economics*, *32*(2), 334–342.

Favero, C. A. (1992). Taxation and the optimization of oil exploration and production: the uk continental shelf. *Oxford Economic Papers*, *44*(2), 187–208.

Favero, C. A., & Pesaran, M. H. (1994). Oil investment in the north sea. *Economic modelling*, *11*(3), 308–329.

George, R. L. (1998). Mining for oil. *Scientific American*, *278*(3), 84–85.

Gordon, D. (2012). *Understanding unconventional oil.* Carnegie Endowment for International Peace.

Gordon, D., Brandt, A. R., Bergerson, J., & Koomey, J. (2015). *Know your oil: Creating a global oil-climate index.* Carnegie Endowment for International Peace Washington, DC.

Heal, G. (1976). The relationship between price and extraction cost for a resource with a backstop technology. *The Bell Journal of Economics*, 371–378.

Hildreth, C., & Houck, J. P. (1968). Some estimators for a linear model with random coefficients. *Journal of the American Statistical Association*, *63*(322), 584–595.

IEA. (2017). World energy statistics. Retrieved from `https://www.oecd-ilibrary.org/content/data/data-00510-en` doi: https://doi.org/https://doi.org/10.1787/data-00510-en

IMF, F. A. D. (2019). *Fiscal policies for paris climate strategies - from principle to practice* (Vol. No. 19/010; Tech. Rep.). International Monetary Fund.

Kleiber, C., & Lupi, C. (2011). Panel unit root testing with r. *URL: https://r-forge. r-project. org/scm/viewvc. php/* checkout*/pkg/inst/doc/panelUnitRootWithR.pdf*.

Lanza, A., Manera, M., & Giovannini, M. (2005). Modeling and forecasting cointegrated relationships among heavy oil and product prices. *Energy Economics*, *27*(6), 831–848.

Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*(1), 13–22.

Livernois, J. R., & Uhler, R. S. (1987). Extraction costs and the economics of nonrenewable resources. *Journal of Political Economy*, *95*(1), 195–203.

Lupi, C., et al. (2009). Unit root cadf testing with r. *Journal of Statistical Software*, *32*(2), 1–19.

Magoon, L. B. (1988). The petroleum systemâ"a classification scheme for research, exploration, and resource assessment. *Petroleum Systems of the United States. US Geological Survey Bulletin*, *1870*, 2–15.

Management, P., & BV, S. (2019). The crude oils and their key characteristics. *https://www.psa-bv.nl/files/CrudeOils.pdf*.

Masnadi, M. S., El-Houjeiri, H. M., Schunack, D., Li, Y., Englander, J. G., Badahdah, A., … others (2018). Global carbon intensity of crude oil production. *Science*, *361*(6405), 851–853.

Maugeri, L. (2012). Oil: the next revolution. *Belfer Center for Science and International Affairs Discussion Paper*(2012-10).

Milonas, N. T., & Henker, T. (2001). Price spread and convenience yield behaviour in the international oil market. *Applied Financial Economics*, *11*(1), 23–36.

Mohr, S., & Evans, G. (2010). Long term prediction of unconventional oil production. *Energy Policy*, *38*(1), 265–276.

Mommessin, P. R., Castano, J. R., Rankin, J. G., & Weiss, M. L. (1981, February 3). *Process for determining the api gravity of oil by fid.* Google Patents. (US Patent 4,248,599)

Nelder, J. A., & Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, *135*(3), 370–384.

Nordhaus, W. D. (2009). The economics of an integrated world oil market. In *International energy workshop* (pp. 17–19).

Pearce, D. W., & Turner, R. K. (1990). *Economics of natural resources and the environment*. JHU Press.

Pesaran, M. H. (1990). An econometric analysis of exploration and extraction of oil in the uk continental shelf. *The Economic Journal*, *100*(401), 367–390.

Pindyck, R. S. (1978). The optimal exploration and production of nonrenewable resources. *Journal of political economy*, *86*(5), 841–861.

Rystad. (2018). Rystad energy ucube upstream database. *Oslo: Rystad Energy.(http://www. rystadenergy. com/Databases/UCube)*.

Shah, A., Fishwick, R., Wood, J., Leeke, G., Rigby, S., & Greaves, M. (2010). A review of novel techniques for heavy oil and bitumen extraction and upgrading. *Energy & Environmental Science*, *3*(6), 700–714.

Sorrell, S., & Speirs, J. (2009). Methods of estimating ultimately recoverable resources. *Review of Evidence on Global Oil Depletion: Technical Report*, *5*.

Sorrell, S., Speirs, J., Bentley, R., Brandt, A., & Miller, R. (2009). Global oil depletion-an assessment of the evidence for a near-term peak in global oil production.

Swamy. (1970). Efficient inference in a random coefficient regression model. *Econometrica: Journal of the Econometric Society*, 311–323.

Swamy, P., & Mehta, J. (1979). Estimation of common coefficients in two regression equations. *Journal of Econometrics*, *10*(1), 1–14.

Torres, L., Yadav, O. P., & Khan, E. (2016). A review on risk assessment techniques for hydraulic fracturing water and produced water management implemented in onshore unconventional oil and gas production. *Science of the Total Environment*, *539*, 478–493.

Tsui, K. K. (2010). More oil, less democracy: evidence from worldwide crude oil discoveries. *The Economic Journal*, *121*(551), 89–115.

Uhler. (1976). Costs and supply in petroleum exploration: the case of alberta. *Canadian Journal of Economics*, 72–90.

Uhler. (1979a). *Oil and gas finding costs* (No. 7). Canadian Energy Research Institute.
Uhler. (1979b). The rate of petroleum exploration and extraction. advances in the economics of energy resources, vol. 2. *Edited by RS Pindyck. Greenwich, CT: JAI Press Inc., USA.*

Mathematical Appendix

Conditions for a Convex Optimization Problem. We need to show that the (unique) solution to the FOCs is a global maximum under the restrictions on parameters $\gamma_2^i \leq 0$, $\gamma_4^i \leq 0$, $\theta_7 \geq 0$, $\theta_8 \geq 0$, $\theta_9 \geq 0$, $\theta_7\theta_9 - \left(\theta_3^{Geo}/R^i\right)^2 \geq 0$; $\theta_7\theta_9 - \left(\theta_4^{Geo}/R^i\right)^2 \geq 0$, $\theta_8\theta_9 - \left(\theta_4^{Geo}/R^i\right)^2 \geq 0$. Define the vector of constraints

$$\mathbf{g}_t^i\left(\left\{Q_{t+s}^i, W_{t+s}^i, L_{t+s-1}^i, M_{t+s-1}^i\right\}_{s=0}^{\infty}\right) = \begin{pmatrix} g_{t,1}^i(W_t^i, L_t^i, L_{t-1}^i, \xi_t^i) \\ g_{t,2}^i(Q_t^i, M_t^i, M_{t-1}^i) \\ g_{t+1,1}^i(W_{t+1}^i, L_{t+1}^i, L_t^i, \xi_{t+1}^i) \\ g_{t+1,2}^i(Q_{t+1}^i, M_{t+1}^i, M_t^i) \\ \dots \end{pmatrix}$$

where each element is defined as:

$$\begin{gathered} g_{t+s,1}^i(W_{t+s}^i, L_{t+s-1}^i, \xi_{t+s}^i) = \\ = \mathbb{E}_{t-1}\left\{-D_{t+s}^i(W_{t+s}, L_{t+s-1}, \xi_{t+s}^i) + L_{t+s}^i - L_{t+s-1}^i \middle| \Omega_{t-1}^i\right\} \end{gathered}$$

and

$$g_{t+s,2}^i(Q_{t+s}^i, M_{t+s}^i, M_{t+s-1}^i) = \mathbb{E}_{t-1}\left\{-M_{t+s}^i + M_{t+s-1}^i + Q_{t+s}^i \middle| \Omega_{t-1}^i\right\}$$

for all $s = 1, 2, 3, \dots$. The Lagrangian of the optimization problem is:

$$\begin{aligned} \mathcal{L}_t^i = \mathbb{E}_{t-1}\Bigg\{ \sum_{s=0}^{\infty} \kappa^s \Pi_{t+s}^i - \lambda_{t+s}^i[-M_{t+s}^i + M_{t+s-1}^i + Q_{t+s}^i] + \\ - \mu_{t+s}^i\left[L_{t+s}^i - D_{t+s}^i\left(W_{t+s}^i, L_{t+s-1}^i, \xi_{t+s}^i\right) - L_{t+s-1}^i \middle| \Omega_{t-1}^i \Bigg\}. \end{aligned}$$

Checking the second-order sufficient (necessary) conditions is the most common approach. This consists in verifying whether the bordered Hessian,

$$\begin{bmatrix} 0 & D\mathbf{g}_t^i\left(\left\{Q_{t+s}^i, W_{t+s}^i, L_{t+s-1}^i, M_{t+s-1}^i\right\}_{s=0}^{\infty}\right) \\ D\mathbf{g}_t^i\left(\left\{Q_{t+s}^i, W_{t+s}^i, L_{t+s-1}^i, M_{t+s-1}^i\right\}_{s=0}^{\infty}\right)^T & D^2\mathcal{L}_t^i \end{bmatrix},$$

is positive definite (positive semidefinite). Because of the complexity of checking such conditions, we follow a different approach. Namely, we show that this is a convex optimization problem under the stated restrictions on parameters. This proof consists in two steps:

1. We show that the objective function $OF_t^i = \mathbb{E}_{t-1}\left\{ \sum_{s=0}^{\infty} \kappa^s \Pi_{t+s}^i \middle| \Omega_{t-1}^i \right\}$ is a concave function;
2. We show that the feasible set $\left\{ x \in X \mid \mathbf{g}_t^i \left(\left\{ Q_{t+s}^i, W_{t+s}^i, L_{t+s-1}^i, M_{t+s-1}^i \right\}_{s=0}^{\infty} \right) \leq \mathbf{0} \right\}$ is a convex set.

Step 1. Let $d^2(x_{t+s}^i, y_{t+r}^i)$ denote the cross derivative of the objective function OF_t^i w.r.t. any two chosen variables x_{t+s}^i, y_{t+r}^i. In order to prove that the objective function is concave we must show that the matrix

$$\mathbb{M}_t^i = \begin{bmatrix} d^2(Q_t^i, Q_t^i) & d^2(Q_t^i, L_t^i) & d^2(Q_t^i, M_t^i) & d^2(Q_t^i, W_t^i) & ... & d^2(Q_t^i, Q_{t+1}^i) & d^2(Q_t^i, L_{t+1}^i) & d^2(Q_t^i, M_{t+1}^i) & d^2(Q_t^i, W_{t+1}^i) & ... \\ d^2(L_t^i, Q_t^i) & d^2(L_t^i, L_t^i) & d^2(L_t^i, M_t^i) & d^2(L_t^i, W_t^i) & ... & d^2(L_t^i, Q_{t+1}^i) & d^2(L_t^i, L_{t+1}^i) & d^2(L_t^i, M_{t+1}^i) & d^2(L_t^i, W_{t+1}^i) & ... \\ d^2(M_t^i, Q_t^i) & d^2(M_t^i, L_t^i) & d^2(M_t^i, M_t^i) & d^2(M_t^i, W_t^i) & ... & d^2(M_t^i, Q_{t+1}^i) & d^2(M_t^i, L_{t+1}^i) & d^2(M_t^i, M_{t+1}^i) & d^2(M_t^i, W_{t+1}^i) & ... \\ d^2(W_t^i, Q_t^i) & d^2(W_t^i, L_t^i) & d^2(W_t^i, M_t^i) & d^2(W_t^i, W_t^i) & ... & d^2(W_t^i, Q_{t+1}^i) & d^2(W_t^i, L_{t+1}^i) & d^2(W_t^i, M_{t+1}^i) & d^2(W_t^i, W_{t+1}^i) & ... \\ d^2(Q_{t+1}^i, Q_t^i) & d^2(Q_{t+1}^i, L_t^i) & d^2(Q_{t+1}^i, M_t^i) & d^2(Q_{t+1}^i, W_t^i) & ... & d^2(Q_{t+1}^i, Q_{t+1}^i) & d^2(Q_{t+1}^i, L_{t+1}^i) & d^2(Q_{t+1}^i, M_{t+1}^i) & d^2(Q_{t+1}^i, W_{t+1}^i) & ... \\ d^2(L_{t+1}^i, Q_t^i) & d^2(L_{t+1}^i, L_t^i) & d^2(L_{t+1}^i, M_t^i) & d^2(L_{t+1}^i, W_t^i) & ... & d^2(L_{t+1}^i, Q_{t+1}^i) & d^2(L_{t+1}^i, L_{t+1}^i) & d^2(L_{t+1}^i, M_{t+1}^i) & d^2(L_{t+1}^i, W_{t+1}^i) & ... \\ d^2(M_{t+1}^i, Q_t^i) & d^2(M_{t+1}^i, L_t^i) & d^2(M_{t+1}^i, M_t^i) & d^2(M_{t+1}^i, W_t^i) & ... & d^2(M_{t+1}^i, Q_{t+1}^i) & d^2(M_{t+1}^i, L_{t+1}^i) & d^2(M_{t+1}^i, M_{t+1}^i) & d^2(M_{t+1}^i, W_{t+1}^i) & ... \\ d^2(W_{t+1}^i, Q_t^i) & d^2(W_{t+1}^i, L_t^i) & d^2(W_{t+1}^i, M_t^i) & d^2(W_{t+1}^i, W_t^i) & ... & d^2(W_{t+1}^i, Q_{t+1}^i) & d^2(W_{t+1}^i, L_{t+1}^i) & d^2(W_{t+1}^i, M_{t+1}^i) & d^2(W_{t+1}^i, W_{t+1}^i) & ... \\ ... & ... & ... & ... & ... & ... & ... & ... & ... & ... \end{bmatrix}$$

is negative semidefinite. This is an hard task given the size of M_t^i. Nevertheless, notice that the cross-derivatives $d^2(x_{t+s}^i, y_{t+r}^i)$ for $x_{t+s}^i \neq y_{t+r}^i$ are all equal to zero, i.e. $\frac{d^2OF_t^i}{dL_{t+s}^i dM_{t+r}^i} = 0$; $\frac{d^2OF_t^i}{dL_{t+s}^i dQ_{t+r}^i} = 0$; $\frac{d^2OF_t^i}{dM_{t+s}^i dW_{t+r}^i} = 0$; $\frac{d^2OF_t^i}{dM_{t+s}^i dL_{t+r}^i} = 0$; $\frac{d^2OF_t^i}{dL_{t+s}^i dW_{t+r}^i} = 0$; $\frac{d^2OF_t^i}{dQ_{t+s}^i dW_{t+r}^i} = 0$; $\frac{d^2OF_t^i}{dW_{t+s}^i dL_{t+r}^i} = 0$ for all $s = 1, 2, 3, ...$ and $r = 1, 2, 3, ...$ except for:

$$d^2\left(Q_{t+s}^i, L_{t+s-1}^i\right) = \frac{d^2OF_t^i}{dQ_{t+s}^i dL_{t+s-1}^i} = -\kappa^s \frac{\theta_3^{Geo}}{R_0^i}$$

$$d^2\left(Q^i_{t+s}, M^i_{t+s-1}\right) = \frac{d^2 OF^i_t}{dQ^i_{t+s} dM^i_{t+s-1}} = -\kappa^s \frac{\theta_4^{Geo}}{R^i_0} \ .$$

Using this findings, we can write the matrix of cross derivatives in form of a diagonal matrix

$$\widetilde{\mathbb{M}}^i_t = \begin{bmatrix} A_1 & \mathbf{0}_{12} & \mathbf{0}_{13} & ... \\ \mathbf{0}_{21} & A_2 & \mathbf{0}_{23} & ... \\ \mathbf{0}_{31} & \mathbf{0}_{32} & A_3 & ... \\ ... & ... & ... & ... \end{bmatrix}$$

where each $(n_j \times n_j)$ submatrix A_j is either diagonal, or it contains all the cross derivatives that are non-zero with respect to a specific choice variable, and where $\mathbf{0}_{jk}$ is a $(n_k \times n_j)$ null matrix. Then, the matrix $\widetilde{\mathbb{M}}^i_t$ is negative semidefinite if all the non-zero submatrices A_1, A_2, A_3, ... are negative semidefinite. First, notice that the own second derivatives are

$$\begin{array}{rr} d^2\left(Q^i_{t+s}, Q^i_{t+s}\right) = -\kappa^s \theta_7 & d^2\left(L^i_{t+s}, L^i_{t+s}\right) = -\kappa^{s+1}\theta_8 \\ d^2\left(M^i_{t+s}, M^i_{t+s}\right) = -\kappa^{s+1}\theta_9 & d^2\left(W^i_{t+s}, W^i_{t+s}\right) = 0 \end{array}$$

for all $s = 0, 1, 2, ...$ Since all the own second derivatives satisfy $d^2(x^i_{t+s}, x^i_{t+s}) \leq 0$, then any diagonal submatrix A_j is negative semidefinite. Thus, for the purposes of this prove it is sufficient to show that each non-diagonal submatrix is also negative semidefinite. Each of such submatrices has form:

$$\mathbb{M}^i_{t+s} = \begin{bmatrix} d^2\left(Q^i_{t+s}, Q^i_{t+s}\right) & d^2\left(Q^i_{t+s}, L^i_{t+s-1}\right) & d^2\left(Q^i_{t+s}, M^i_{t+s-1}\right) & d^2\left(Q^i_{t+s}, W^i_{t+s}\right) \\ d^2\left(L^i_{t+s-1}, Q^i_{t+s}\right) & d^2\left(L^i_{t+s-1}, L^i_{t+s-1}\right) & d^2\left(L^i_{t+s-1}, M^i_{t+s-1}\right) & d^2\left(L^i_{t+s-1}, W^i_{t+s}\right) \\ d^2\left(M^i_{t+s-1}, Q^i_{t+s}\right) & d^2\left(M^i_{t+s-1}, L^i_{t+s-1}\right) & d^2\left(M^i_{t+s-1}, M^i_{t+s-1}\right) & d^2\left(M^i_{t+s-1}, W^i_{t+s}\right) \\ d^2\left(W^i_{t+s}, Q^i_{t+s}\right) & d^2\left(W^i_{t+s}, L^i_{t+s-1}\right) & d^2\left(W^i_{t+s}, M^i_{t+s-1}\right) & d^2\left(W^i_{t+s}, W^i_{t+s}\right) \end{bmatrix} =$$

$$= \kappa^s \begin{bmatrix} -\theta_7 & -\frac{\theta_3^{Geo}}{R^i} & -\frac{\theta_4^{Geo}}{R^i} & 0 \\ -\frac{\theta_3^{Geo}}{R^i} & -\theta_8 & 0 & 0 \\ -\frac{\theta_4^{Geo}}{R^i} & 0 & -\theta_9 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

for $s = 1, 2, ...$ is negative semidefinite. Lastly, notice that for each $s = 1, 2, ...$ one gets $\mathbb{M}^i_{t+s+1} = \kappa \mathbb{M}^i_{t+s}$, which implies that if $\mathbb{M}^i_t$ is positive semidefinite, then all $\mathbb{M}^i_{t+s}$ for $s = 1, 2, ...$ also are, and therefore the objective function is concave. The matrix M^i_{t+s} is negative semidefinite if the following conditions hold true: $\theta_7 \geq 0$, $\theta_8 \geq 0$, $\theta_9 \geq 0$, $\theta_7\theta_8 - \left(\theta_3^{Geo}/R^i\right)^2 \geq 0$; $\theta_7\theta_9 - \left(\theta_4^{Geo}/R^i_0\right)^2 \geq 0$ $\theta_8\theta_9 - \left(\theta_4^{Geo}/R^i\right)^2 \geq 0$.

Step 2. We show that the set defined by $\left\{x \in X \mid \mathbf{g}^i_t\left(\left\{Q^i_{t+s}, W^i_{t+s}, L^i_{t+s-1}, M^i_{t+s-1}\right\}_{s=0}^{\infty}\right) \leq \mathbf{0}\right\}$ is a convex set. First, recall that – because X is a convex subset of $\mathbb{R}_+$ – the set $\left\{x \in X \mid g^i_{t+s,j}(...) \leq 0\right\}$ is convex if $g^i_{t+s,j}$ is a convex function of the choice variables. Second, the intersections of convex sets are convex sets. Thus, in order to prove the result it is sufficient to show that each set $\left\{x \in X \mid g^i_{t+s,j}(...) \leq 0\right\}$ for $s = 1, 2, ...$ and $j = 1, 2$ is a convex set. In conclusion, it is sufficient to show that each function $g^i_{t+s,1}(W^i_{t+s}, L^i_{t+s}, L^i_{t+s-1}, \xi^i_{t+s})$ and $g^i_{t+s,2}(Q^i_{t+s}, M^i_{t+s}, M^i_{t+s-1})$ for $s = 0, 1, 2, ...$ is convex. Then we have

$$g^i_{t+s,2}(Q^i_{t+s}, M^i_{t+s}, M^i_{t+s-1}) = -M^i_{t+s} + M^i_{t+s-1} + Q^i_{t+s}$$

which is (weakly) convex in all choice variables because it is linear. Secondly, we have

$$g^i_{t+s,1}(W^i_{t+s}, L^i_{t+s}, L^i_{t+s-1}, \xi^i_{t+s}) = L^i_{t+s} - D^i_{t+s}\left(W^i_{t+s}, L^i_{t+s-1}, \xi^i_t\right) - L^i_{t+s-1}$$

Let $c^2_{t+s}\left(x^i_{t+s}, y^i_{t+r}\right) \equiv \frac{\partial^2 g^i_{t+s,1}}{\partial x^i_{t+s} \partial y^i_{t+r}}$. In order to prove that $g^i_{t+s,1}$ is convex, we need to show that the matrix:

$$\begin{bmatrix} c^2_{t+s}\left(W^i_{t+s}, W^i_{t+s}\right) & c^2_{t+s}\left(W^i_{t+s}, L^i_{t+s-1}\right) & c^2_{t+s}\left(W^i_{t+s}, L^i_{t+s}\right) \\ c^2_{t+s}\left(L^i_{t+s-1}, W^i_{t+s}\right) & c^2_{t+s}\left(L^i_{t+s-1}, L^i_{t+s-1}\right) & c^2_{t+s}\left(L^i_{t+s-1}, L^i_{t+s}\right) \\ c^2_{t+s}\left(L^i_{t+s}, W^i_{t+s}\right) & c^2_{t+s}\left(L^i_{t+s}, L^i_{t+s-1}\right) & c^2_{t+s}\left(L^i_{t+s}, L^i_{t+s}\right) \end{bmatrix} =$$

$$= \begin{bmatrix} -2\gamma^i_2 & 0 & 0 \\ 0 & -2\gamma^i_4 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

is positive semidefinite. Thus, $g^i_{t+s,1}(W^i_{t+s}, L^i_{t+s}, L^i_{t+s-1}, \xi^i_{t+s})$ is convex if $\gamma^i_2 \leq 0$ and $\gamma^i_4 \leq 0$. In turn, this implies that the set $\left\{x \in X \mid \mathbf{g}^i_t\left(\left\{Q^i_{t+s}, W^i_{t+s}, L^i_{t+s-1}, M^i_{t+s-1}\right\}^{\infty}_{s=0}\right) \leq \mathbf{0}\right\}$ is a convex set given the restrictions $\gamma^i_2 \leq 0$, $\gamma^i_4 \leq 0$.

Lastly, the results in Step 1 and Step 2 together imply that the maximization presented in Section 3.1 is a convex optimization problem given the restrictions on the parameters $\gamma^i_2 \leq 0$, $\gamma^i_4 \leq 0$, $\theta_7 \geq 0$, $\theta_8 \geq 0$, $\theta_9 \geq 0$, $\theta_7\theta_8 - \left(\theta^{Geo}_3/R^i\right)^2 \geq 0$; $\theta_7\theta_9 - \left(\theta^{Geo}_4/R^i\right)^2 \geq 0$ $\theta_8\theta_9 - \left(\theta^{Geo}_4/R^i\right)^2 \geq 0$. This implies that the solution to the FOCs is a global maximum. Q.E.D.

First-Order Necessary Conditions for a Global Maximum. The FOCs for each $s = 0, 1, 2, 3, ..$ write:

$$[Q^i_{t+s}] : \mathbb{E}_{t-1}\left\{P^i_{t+s} - \frac{\partial C^i_{t+s}(.)}{\partial Q^i_{t+s}} - \lambda^i_{t+s} \middle| \Omega^i_{t-1}\right\} = 0$$

$$[W^i_{t+s}] : \mathbb{E}_{t-1}\left\{-1 + \mu^i_{t+s}\frac{\partial D^i_{t+s}(.)}{\partial W^i_{t+s}} \middle| \Omega^i_{t-1}\right\} = 0$$

$$[L^i_{t+s}] : \mathbb{E}_{t-1}\left\{-\mu^i_{t+s} - \kappa\frac{\partial C^i_{t+s+1}(.)}{\partial L^i_{t+s}} + \kappa\mu^i_{t+s+1}\frac{\partial D^i_{t+s+1}(.)}{\partial L^i_{t+s}} + \kappa\mu^i_{t+s+1} \middle| \Omega^i_{t-1}\right\} = 0$$

$$[M^i_{t+s}] : \mathbb{E}_{t-1}\left\{ \lambda^i_{t+s} - \kappa \frac{\partial C^i_{t+s+1}(.)}{\partial M^i_{t+s}} - \kappa \lambda^i_{t+s+1} \middle| \, \Omega^i_{t-1} \right\} = 0$$

$$[\mu^i_{t+s}] : \mathbb{E}_{t-1}\left\{ -D^i_{t+s}(W_{t+s}, L_{t+s-1}, \xi^i_{t+s}) + L^i_{t+s} - L^i_{t+s-1} \middle| \, \Omega^i_{t-1} \right\} \leq 0$$

$$[\lambda^i_{t+s}] : \mathbb{E}_{t-1}\left\{ M^i_{t+s} - M^i_{t+s-1} - Q^i_{t+s} \middle| \, \Omega^i_{t-1} \right\} \leq 0$$

Interior Solution. From the first part we know that the solution is a global maximum. The solution must be interior for $(Q^i_{t+s}, W^i_{t+s}, L^i_{t+s}, M^i_{t+s})$ if $(Q^i_{t+s}, W^i_{t+s}, L^i_{t+s}, M^i_{t+s})^* \gg 0$. Thus, whenever $(Q^i_{t+s}, W^i_{t+s}, L^i_{t+s}, M^i_{t+s})^* \gg 0$ the FOCs w.r.t. $Q^i_{t+s}, W^i_{t+s}, L^i_{t+s}, M^i_{t+s}, \lambda_{t+s}, \mu_{t+s}$ must be binding, and therefore they can be used to derive the following equilibrium conditions.

Equilibrium Conditions (Shadow-Prices). The results in the previous sections lead to the following equilibrium conditions in period t:

1. Shadow-price of discovered oil:

$$\mathbb{E}_{t-1}\left[\lambda^i_t \middle| \Omega^i_{t-1}\right] = \mathbb{E}_{t-1}\left[P^i_t \middle| \, \Omega^i_{t-1}\right] - \mathbb{E}_{t-1}\left[\frac{\partial C^i_t(.)}{\partial Q^i_t} \middle| \, \Omega^i_{t-1}\right]$$

2. Expected shadow-price of undiscovered oil:

$$\mathbb{E}_{t-1}\left[\mu^i_t \middle| \Omega^i_{t-1}\right] = \mathbb{E}_{t-1}\left[\left(\frac{\partial D^i_t(.)}{\partial W^i_t}\right)^{-1} \middle| \, \Omega^i_{t-1}\right]$$

3. Law of motion of the shadow-price of discovered oil:

$$\mathbb{E}_{t-1}\left[\lambda^i_{t+1} \middle| \Omega^i_{t-1}\right] = \mathbb{E}_{t-1}\left[\frac{\lambda^i_t}{\kappa} - \frac{\partial C^i_{t+1}(.)}{\partial M^i_t} \middle| \, \Omega^i_{t-1}\right]$$

4. Law of motion of the expected shadow-price of undiscovered oil:

$$\mathbb{E}_{t-1}\left[\mu^i_{t+1} \middle| \Omega^i_{t-1}\right] = \mathbb{E}_{t-1}\left[\left(\frac{\mu^i_t}{\kappa} + \frac{\partial C^i_{t+1}(.)}{\partial L^i_t}\right)\left(\frac{\partial D^i_{t+1}(.)}{\partial L^i_t} + 1\right)^{-1} \middle| \, \Omega^i_{t-1}\right] ,$$

which correspond to those in the main body of the paper. Q.E.D.

Note that λ^i_t is fully known at time t by the firm – i.e., $\mathbb{E}_{t-1}\left[\lambda^i_t \middle| \Omega^i_{t-1}\right] = \hat{\lambda}^i_t$ – but it is a random variable for the econometrician, because the exact realizations of the random coefficients are unknown. Thus, the estimates provided in Section 4 correspond to the expectations of these random variables taken over the distribution of the estimated coefficients.