Supplementary material (SM): Oil carbon intensity impacts of COVID-19 and other short-term demand shocks

Mohammad S. Masnadi[1*], Giacomo Benini[2], Alice Milivinti[3], James E. Anderson[4], Timothy J. Wallington[4], Robert De Kleine[4], Valerio Dotti[5], Patrick Jochem[6], Hassan El-Houjeiri[7], Adam R. Brandt[2*]

[1] *Department of Chemical and Petroleum Engineering, University of Pittsburgh, PA, USA*
[2] *Department of Energy Resources Engineering, Stanford University, CA, USA*
[3] *Department of Medicine, Stanford University, CA, USA*
[4] *Research and Advanced Engineering, Ford Motor Company, MI, USA*
[5] *Department of Economics, Washington University in St. Louis, MO, USA*
[6] *Department of Energy System Analysis, Institute of Engineering Thermodynamics, Germany*
[7] *Climate and Sustainability Group, Aramco Research Center - Detroit, MI, USA*

* m.masnadi@pitt.edu, abrandt@stanford.edu

Contents

1 Theoretical Economic Model

1.1 Firm Problem

We assume that every field is managed by a 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 $s = 0, 1, 2, ...$ In other words, the firm 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 while having negligible consequences on the implications on the analysis (Favero, 1992; Favero & Pesaran, 1994).

The field intra-temporal profits,

$$\Pi^i_{t+s} = P^i_{t+s} Q^i_{t+s} - C^i_{t+s}(Q^i_{t+s}, R^i_{t+s-1}, \epsilon^i_{t+s}) - W^i_{t+s} \ , \tag{1}$$

are the difference between the field's revenues and the field's costs. The former are the product between the price at which field i sells its output P^i_{t+s} and the quantity of the output produced Q^i_{t+s}. The costs are divided into two macro-classes. The first ones are production costs. The second ones are discovery costs. Production costs are function of the quantity of oil extracted and of the amount of reserves available when the production decision 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 $L^i_{t+s-1} = L^i + \sum_{r=1}^{t+s-1} D^i_r$, where D^i_r are the new discoveries in period r, minus the sum of extracted liquids $M^i_{t+s-1} = M^i + \sum_{r=1}^{t+s-1} Q^i_r$, such that $R^i_{t+s-1} = R^i + L^i_{t+s-1} + M^i_{t+s-1}$. Finally, the extraction costs are function of an idiosyncratic shock ϵ^i_{t+s}, which can randomly increase or decrease the extraction costs due to unexpected events. The exploration costs, W^i_{t+s}, are the expenses incurred to discover new oil located in field i. They equal the product between the number of spud wells and the per-well cost.

Every field tries to maximize (1) while facing two physical constraints. The first one,

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

makes the cumulative discoveries obtained till time $t+s$ equal to the cumulative discoveries obtained till time $t+s-1$ plus the ones obtained at time $t+s$, $D^i_{t+s}(.)$. The latter are function of the exploration costs, of the cumulative amount of past discoveries, and of an idiosyncratic shock ξ^i_{t+s}, which can randomly increase or decrease the volumes of discovered reserves.

The second constraint ensures that the reserves available at time $t+s$ equal the ones at time $t+s-1$, plus the new discoveries, minus the quantity of oil extracted, plus eventual idiosyncratic revision-extensions of previously discovered reserves, ν^i_{t+s}, such that

$$R^i_{t+s} = R^i_{t+s-1} + D^i_{t+s} - Q^i_{t+s} + \nu^i_{t+s} \ . \tag{3}$$

The field solves the optimization problem consulting an information set, Ω^i_{t+s-1}, which includes previous prices, quantities and shocks,

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

1.2 Cost and Discovery Function

We make the extraction costs function of the volumes of oil produced and of the volumes of reserves available when the production decision is taken.

The link between extraction costs and volumes of production reflects the convex nature of the field's costs. In other words, fields increase their costs as they increase production. Furthermore, the more the production level is closed to the pick capacity, the more extracting the next barrel becomes costly[1]. The link between extraction costs and volumes of reserves reflects the role played by the reservoir pressure in the production process. If a reservoir contains low viscosity oil, trapped in impermeable rocks (a.k.a. Shale & Tight Oil), the wells fracture the oil-containing rocks to allow the natural pressure of the reservoir to lift above ground a mixture of oil, water and stones. In the same way, if a reservoir contains low (a.k.a. Light & Medium Oil) or high (a.k.a. Heavy & Extra Heavy Oil) viscosity oil, trapped in permeable rocks, wells drill vertically to reach the deposit. Once the wells reach the petroleum liquids, the natural pressure lifts the oil above ground. When the pressure declines, the management needs to inject increasing amounts of water and/or of steam to keep the volumes of production constant. In other words, there is an inverse relation between the costs of production and the reservoir pressure in all types of oil formation, with the exception of oil Sands. Capturing this reality would require to collect information about the volumes of water injected, the

[1]If this assumption is not respected, the optimal production level is the pick capacity of the field. An explanatory analysis of our dataset suggests that this is never the case for the analyzed sample, with a median distance between the actual level of production and the pick production capacity of 11,510 barrels per day and an average distance of 46,849 barrels per day.

volumes of steam injected, the injection pressure, the water-oil-ratio, the steam-oil-ratio, and the water-injection ratio, possibly starting from well-level data. To the best of our knowledge no such data are available on a global scale. Therefore, we follow a long-standing micro-econometric tradition and use the volumes of available reserves as a first-order approximation of all the mentioned variables under the general assumption that more reserves equal more pressure and those lower marginal costs (Uhler, 1976, 1979a, 1979b; Livernois & Uhler, 1987).

We capture the two previous intuitions rewriting the cost function used in Pesaran (1990) in panel data form,

$$C_t^i = \theta_0^{Geo} + \theta_1^{Geo} Q_t^i + \frac{\theta_2^{Geo}}{2} {Q_t^i}^2 + \frac{\theta_3^{Geo}}{2} \frac{{Q_t^i}^2}{R_{t-1}^i} + \epsilon_t^i \ , \tag{4}$$

In equation (4), the dependent variable C_t^i equals the sum of Operating (OPEX) and of Capital Expenditures not linked to exploration (Non Exp CAPEX)[2] measured in Million US Dollars (MM $) spend per Year. The explanatory variable Q_t^i equals the amount of output produced, measured in Million Barrels of Oil Equivalent (MM BOE) extracted per Year[3], while R_{t-1}^i is the volume of recoverable reserves, measured in MM BOE. The idiosyncratic shock is normally distributed with finite homoskedastic variance, $\epsilon_t^i \overset{iid}{\sim} \mathcal{N}(0, \sigma_\epsilon^2)$. Finally, θ_0^{Geo} identifies the fixed costs, while $(\theta_1^{Geo}, \theta_2^{Geo}, \theta_3^{Geo})$ identify the variable production costs, rescaled by the volumes of available reservers. All the coefficients are expected to be positive. θ_0^{Geo} is expected to be positive because fields experience positive costs even when the production level is zero such as license fees and insurances. θ_1^{Geo} is expected to be positive because increasing production increases costs by increasing expenses such as utilities. In the same way, $(\theta_2^{Geo}, \theta_3^{Geo})$ are expected to be positive because increasing production is marginally more expensive, irrespectively of the volumes of reserves.

We make the volumes of discoveries function a quadratic function of current exploration expenditures W_t^i and of cumulated past discoveries L_{t-1}^i, and of the

[2]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.

[3]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.

idiosyncratic shock ξ_t^i,

$$D_t^i = \gamma_1 W_t^i + \gamma_2 W_t^{i\ 2} + \gamma_3 L_{t-1}^i + \gamma_4 L_{t-1}^{i\ 2} + \xi_t^i \ . \tag{5}$$

In equation (5), 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. γ_1 and γ_2 identify the link between exploration expenditures and amount of discoveries. γ_1 is expect to be positive, since the more a field invests in exploration the more it discovers new oil. γ_2 is expected to be negative since marginal discoveries are declining in exploration CAPEX. γ_3 is expected to 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.

1.3 Optimization

The optimization problem,

$$\max_{\{Q_{t+s}^i, W_{t+s}^i, R_{t+s}^i, L_{t+s}^i\}_{s=0}^{\infty}} \mathbb{E}_{t-1}\left\{ \sum_{s=0}^{\infty} \kappa^s \Pi_{t+s}^i(P_{t+s}^i, Q_{t+s}^i, W_{t+s}^i, R_{t+s-1}^i, L_{t+s-1}^i, \epsilon_{t+s}^i) \big| \Omega_{t-1}^i \right\} ,$$

$$s.t. \left\{ \begin{array}{c} D_{t+s}^i - Q_{t+s}^i + \nu_{t+s}^i - R_{t+s}^i + R_{t+s-1}^i = 0 \\ L_{t+s}^i - L_{t+s-1}^i - D_{t+s}^i = 0 \\ Q_{t+s} \geq 0;\ W_{t+s}^i \geq 0 \\ R_{t+s} \geq 0;\ L_{t+s}^i \geq 0 \end{array} \right\}_{s=0}^{\infty} ,$$

where $0 \leq \kappa < 1$ is the inter-temporal discount factor, can be solved using standard methods[4]. Before ensuring that the solution of the Lagrangian delivers a maximum, we specify the functional form of the cost and of the discovery function.

We need to show that the (typically unique) solution to the FOCs of the firm optimization problem is a global maximum under the restrictions on parameters $\gamma_2 \leq 0$, $\gamma_4 \leq 0$, $\theta_2^{Geo} \geq 0$, $\theta_3^{Geo} \geq 0$. First, we perform a change of variable. That is, we substitute R_{t+s}^i with the previously defined random variable B_{t+s}^i using the formula $B_{t+s}^i = R^i + L_{t+s}^i - R_{t+s}^i$. The cost function becomes

$$C_t^i = \theta_0^{Geo} + \theta_1^{Geo} Q_t^i + \frac{\theta_2^{Geo}}{2} Q_t^i + \frac{\theta_3^{Geo}}{2} \frac{Q_t^i}{R^i - B_{t-1}^i + L_{t-1}^i} + \epsilon_t^i$$

[4]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 as shown in Footnote (1).

and the problem gains the constraints $R^i + L^i_{t+s} - B^i_{t+s} \geq 0$ for $s = 0, 1, 2, ...$ (instead of $R^i_{t+s} \geq 0$). Thus, the firm's optimization problem becomes:

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

$$s.t. \left\{ \begin{array}{c} -B^i_{t+s} + B^i_{t+s-1} + Q^i_{t+s} \leq 0 \\ L^i_{t+s} - L^i_{t+s-1} - D^i_{t+s} \leq 0 \\ Q^i_{t+s} \geq 0 \\ W^i_{t+s} \geq 0 \\ R^i - B^i_{t+s} + L^i_{t+s} \geq 0 \\ L^i_{t+s} \geq 0 \end{array} \right\}_{s=0}^{\infty}.$$

The Lagrangian of the optimization problem is

$$\begin{aligned} \mathcal{L}^i_t = \mathbb{E}_{t-1} \Bigg\{ \sum_{s=0}^{\infty} \kappa^s \Pi^i_{t+s} & - \lambda^i_{t+s} \left[B^i_{t+s-1} - B^i_{t+s} + Q^i_{t+s} - \nu^i_{t+s} \right] + \\ & - \mu^i_{t+s} \left[L^i_{t+s} - D^i_{t+s} \left(W^i_{t+s}, L^i_{t+s-1}, \xi^i_{t+s} \right) - L^i_{t+s-1} \right] + \\ & - \phi^i_{t+s} \left[-Q^i_{t+s} \right] - \chi^i_{t+s} \left[-W^i_{t+s} \right] + \\ & - \psi^i_{t+s} \left[-R^i + B^i_{t+s} - L^i_{t+s} \right] - \zeta^i_{t+s} \left[-L^i_{t+s} \right] \Big| \Omega^i_{t-1} \Bigg\}. \end{aligned}$$

Define the vector of constraints

$$\mathbf{g}^i_t \left(\left\{ Q^i_{t+s}, W^i_{t+s}, L^i_{t+s-1}, B^i_{t+s-1} \right\}_{s=0}^{\infty} \right) = \left(\begin{array}{c} g^i_{t,1}(W^i_t, L^i_t, L^i_{t-1}, \xi^i_t) \\ g^i_{t,2}(Q^i_t, B^i_t, B^i_{t-1}, W^i_t, L^i_{t-1}, \xi^i_t, \nu^i_t) \\ g^i_{t,3}(Q^i_t) \\ g^i_{t,4}(W^i_t) \\ g^i_{t,5}(B^i_t, L_t) \\ g^i_{t,6}(L^i_t) \\ g^i_{t+1,1}(W^i_{t+1}, L^i_{t+1}, L^i_t, \xi^i_{t+1}) \\ g^i_{t+1,2}(Q^i_{t+1}, B^i_{t+1}, B^i_t, W^i_{t+1}, L^i_t, \xi^i_{t+1}, \nu^i_{t+1}) \\ g^i_{t+1,3}(Q^i_{t+1}) \\ g^i_{t+1,4}(W^i_{t+1}) \\ g^i_{t+1,5}(B^i_{t+1}, L_{t+1}) \\ g^i_{t+1,6}(L^i_{t+1}) \\ \vdots \end{array} \right) \leq \mathbf{0}$$

where each element is defined as:

$$\begin{aligned} g^i_{t+s,1}(W^i_{t+s}, L^i_{t+s}, L^i_{t+s-1}, \xi^i_{t+s}) &= -D^i_{t+s}(W_{t+s}, L_{t+s-1}, \xi^i_{t+s}) + L^i_{t+s} - L^i_{t+s-1} \\ g^i_{t+s,2}(Q^i_{t+s}, B^i_{t+s}, B^i_{t+s-1}, \nu^i_{t+s}) &= -B^i_{t+s} + B^i_{t+s-1} + Q^i_{t+s} - \nu^i_{t+s} \end{aligned}$$

for all $s = 0, 1, 2, 3, ...$ The other constraints are $g^i_{t+s,3}(Q^i_{t+s}) = -Q^i_{t+s}$, $g^i_{t+s,4}(W^i_{t+s}) = -W^i_{t+s}$, $g^i_{t+s,5}(B^i_{t+s}, L_{t+s}) = -R^i - L^i_{t+s} + B^i_{t+s}$, $g^i_{t+s,g}(L^i_{t+s}) = -L^i_{t+s}$. The most common approach is that of checking the second-order sufficient (necessary) conditions. This consists in verifying whether the bordered Hessian

$$\begin{bmatrix} 0 & D\mathbf{g}^i_t\left(\left\{Q^i_{t+s}, W^i_{t+s}, L^i_{t+s-1}, B^i_{t+s-1}\right\}_{s=0}^{\infty}\right) \\ D\mathbf{g}^i_t\left(\left\{Q^i_{t+s}, W^i_{t+s}, L^i_{t+s-1}, B^i_{t+s-1}\right\}_{s=0}^{\infty}\right)^T & D^2\mathcal{L}^i_t \end{bmatrix}$$

is positive definite (positive semidefinite). Due to 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. We follow a two steps procedure

1. Prove that the objective function $OF^i_t = \mathbb{E}_{t-1}\left\{ \sum_{s=0}^{\infty} \kappa^s \Pi^i_{t+s} \middle| \Omega^i_{t-1} \right\}$ is a concave
2. Prove that the feasible set $\left\{x \in X \mid \mathbf{g}^i_t\left(\left\{Q^i_{t+s}, W^i_{t+s}, L^i_{t+s-1}, B^i_{t+s-1}\right\}_{s=0}^{\infty}\right) \leq \mathbf{0}\right\}$ is convex.

This approach has the further advantage of ensuring that the solution to the FOCs is globally optimal.

Step 1. Let $d^2(x^i_{t+s}, y^i_{t+r})$ denote the cross derivative of the objective function OF^i_t w.r.t. any two choice variables x^i_{t+s}, y^i_{t+r}. In order to prove that the objective function is concave we must show that the matrix

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

$$\begin{bmatrix}
... & d^2\left(Q_t^i, Q_{t+1}^i\right) & d^2\left(Q_t^i, L_{t+1}^i\right) & d^2\left(Q_t^i, B_{t+1}^i\right) & d^2\left(Q_t^i, W_{t+1}^i\right) & ... \\
... & d^2\left(L_t^i, Q_{t+1}^i\right) & d^2(L_t^i, L_{t+1}^i) & d^2(L_t^i, B_{t+1}^i) & d^2\left(L_t^i, W_{t+1}^i\right) & ... \\
... & d^2\left(B_t^i, Q_{t+1}^i\right) & d^2(B_t^i, L_{t+1}^i) & d^2(B_t^i, B_{t+1}^i) & d^2\left(B_t^i, W_{t+1}^i\right) & ... \\
... & d^2\left(W_t^i, Q_{t+1}^i\right) & d^2(W_t^i, L_{t+1}^i) & d^2(W_t^i, B_{t+1}^i) & d^2\left(W_t^i, W_{t+1}^i\right) & ... \\
... & d^2\left(Q_{t+1}^i, Q_{t+1}^i\right) & d^2(Q_{t+1}^i, L_{t+1}^i) & d^2(Q_{t+1}^i, B_{t+1}^i) & d^2\left(Q_{t+1}^i, W_{t+1}^i\right) & ... \\
... & d^2\left(L_{t+1}^i, Q_{t+1}^i\right) & d^2(L_{t+1}^i, L_{t+1}^i) & d^2(L_{t+1}^i, B_{t+1}^i) & d^2\left(L_{t+1}^i, W_{t+1}^i\right) & ... \\
... & d^2\left(B_{t+1}^i, Q_{t+1}^i\right) & d^2(B_{t+1}^i, L_{t+1}^i) & d^2(B_{t+1}^i, B_{t+1}^i) & d^2\left(B_{t+1}^i, W_{t+1}^i\right) & ... \\
... & d^2\left(W_{t+1}^i, Q_{t+1}^i\right) & d^2(W_{t+1}^i, L_{t+1}^i) & d^2(W_{t+1}^i, B_{t+1}^i) & d^2\left(W_{t+1}^i, W_{t+1}^i\right) & ... \\
... & ... & ... & ... & ... & ...
\end{bmatrix}$$

is negative semidefinite for all values of $\left\{Q_{t+s}^i, W_{t+s}^i, L_{t+s-1}^i, B_{t+s-1}^i\right\}_{s=0}^{\infty}$ in X. Since 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. $\frac{d^2OF_t^i}{dL_{t+s}^i dB_{t+r}^i} = 0$; $\frac{d^2OF_t^i}{dL_{t+s}^i dQ_{t+r}^i} = 0$; $\frac{d^2OF_t^i}{dB_{t+s}^i dW_{t+r}^i} = 0$; $\frac{d^2OF_t^i}{dB_{t+s}^i dL_{t+r}^i} = 0$; $\frac{d^2OF_t^i}{dL_{t+s}^i dW_{t+r}^i} = 0$; $\frac{d^2OF_t^i}{dQ_{t+s}^i dW_{t+r}^i} = 0$ for all $s = 1, 2, 3, ...$ and $r = 1, 2, 3, ...$ except for the following:

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

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

$$d^2\left(B_{t+s-1}^i, L_{t+s-1}^i\right) = \frac{d^2OF_t^i}{dB_{t+s-1}^i dL_{t+s-1}^i} = \kappa^s \frac{\theta_3^{Geo} {Q_{t+s}^i}^2}{{R_{t+s-1}^i}^3} \ ,$$

the matrix of cross derivatives can be rearranged in the form of a diagonal matrix:

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

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 $\widetilde{M_t^i}$ 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

$$d^2(Q_{t+s}^i, Q_{t+s}^i) = -\kappa^s(\theta_2^{Geo} + \frac{\theta_3^{Geo}}{R_{t+s-1}^i}) \qquad d^2(L_{t+s}^i, L_{t+s}^i) = -\kappa^{s+1}\theta_3^{Geo} {Q_{t+s+1}^i}^2 {R_{t+s}^i}^{-3}$$
$$d^2(B_{t+s}^i, B_{t+s}^i) = -\kappa^{s+1}\theta_3^{Geo} {Q_{t+s+1}^i}^2 {R_{t+s}^i}^{-3} \quad d^2(W_{t+s}^i, W_{t+s}^i) = 0 \ ,$$

for all $s = 0, 1, 2, ...$ where $R_{t+s-1}^i = B_{t+s-1}^i + L_{t+s-1}^i$. Since all the own second derivatives satisfy $d^2(x_{t+s}^i, x_{t+s}^i) \leq 0$, 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:

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

$$= \kappa^s \begin{bmatrix} -\theta^{Geo}_2 - \frac{\theta^{Geo}_3}{R^i_{t+s-1}} & -\frac{\theta^{Geo}_3 Q^i_{t+s}}{{R^i_{t+s-1}}^2} & \frac{\theta^{Geo}_3 Q^i_{t+s}}{{R^i_{t+s-1}}^2} & 0 \\ -\frac{\theta^{Geo}_3 Q^i_{t+s}}{{R^i_{t+s-1}}^2} & -\frac{\theta^{Geo}_3 {Q^i_{t+s}}^2}{{R^i_{t+s-1}}^3} & \frac{\theta^{Geo}_3 {Q^i_{t+s}}^2}{{R^i_{t+s-1}}^3} & 0 \\ \frac{\theta^{Geo}_3 Q^i_{t+s}}{{R^i_{t+s-1}}^2} & \frac{\theta^{Geo}_3 {Q^i_{t+s}}^2}{{R^i_{t+s-1}}^3} & -\frac{\theta^{Geo}_3 {Q^i_{t+s}}^2}{{R^i_{t+s-1}}^3} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

for $s = 1, 2, ...$ and is negative semidefinite for $\theta^{Geo}_2 \geq 0$, $\theta^{Geo}_3 \geq 0$. Lastly, notice that for each $s = 1, 2, ...$ if M^i_{t+s} is negative semidefinite for $s = 1, 2, ...$ the objective function is concave. Since Q^i_{t+s} and R^i_{t+s-1} are weakly positive scalars, the matrix M^i_{t+s} is negative semidefinite if the following conditions hold true: $\theta^{Geo}_2 \geq 0$, $\theta^{Geo}_3 \geq 0$.

Step 2. We show that the set defined by $\left\{x \in X \mid \mathbf{g}^i_t\left(\left\{Q^i_{t+s}, W^i_{t+s}, L^i_{t+s-1}, B^i_{t+s-1}\right\}_{s=0}^{\infty}\right) \leq \mathbf{0}\right\}$ 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, recall that 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 $g^i_{t+s,2}(Q^i_{t+s}, B^i_{t+s}, B^i_{t+s-1}, W^i_{t+s}, L^i_{t+s-1}, \xi^i_{t+s}, \nu^i_{t+s})$ for $s = 0, 1, 2, ...$ is convex. Then we have: $g^i_{t+s,2}(Q^i_{t+s}, B^i_{t+s}, B^i_{t+s-1}, \nu^i_{t+s}) = -B^i_{t+s} + B^i_{t+s-1} + Q^i_{t+s} - \nu^i_{t+s}$, which is trivially (weakly) convex because it is a linear function. Secondly, we have

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

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:

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

$$= \begin{bmatrix} -2\gamma_2 & 0 & 0 \\ 0 & -2\gamma_4 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

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 $\gamma_2 \leq 0$ and $\gamma_4 \leq 0$. Lastly, the functions $g^i_{t+s,3}(Q^i_{t+s}) = -Q^i_{t+s}$, $g^i_{t+s,4}(W^i_{t+s}) = -W^i_{t+s}$, $g^i_{t+s,5}(B^i_{t+s}, L_{t+s}) = -R^i - L^i_{t+s} + B^i_{t+s}$, $g^i_{t+s,g}(L^i_{t+s}) = -L^i_{t+s}$ are all linear, which implies that they are trivially (weakly) convex. In turn, these three results imply that the set $\left\{x \in X \mid \mathbf{g}^i_t\left(\left\{Q^i_{t+s}, W^i_{t+s}, L^i_{t+s-1}, B^i_{t+s-1}\right\}_{s=0}^{\infty}\right) \leq \mathbf{0}\right\}$ is a convex set given the restrictions $\gamma_2 \leq 0$, $\gamma_4 \leq 0$.

Lastly, the results of Step 1 and 2 together imply that the firm maximization is a convex optimization problem given the restrictions on the parameters $\gamma_2 \leq 0$, $\gamma_4 \leq 0$, $\theta_2^{Geo} \geq 0$, $\theta_3^{Geo} \geq 0$. 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:

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

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

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

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

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

plus the standard primal feasibility, dual feasibility and complementary slackness conditions.

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

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:

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

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

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

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

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

Notice that λ_t^i is fully known at time t by the firm – i.e., $\mathbb{E}_{t-1}\left[\lambda_t^i \left| \Omega_{t-1}^i \right.\right] = \hat{\lambda}_t^i$ – but it is a random variable for the econometrician, because the exact realizations of the random coefficients are unknown. All these results are in line with standard natural resource economics.

2 Empirical Economic Model

The empirical analysis associates the shadow price of a field,

$$\mathbb{E}_{t-1}[\lambda_t^i | \Omega_{t-1}^i] = \mathbb{E}_{t-1}[P_t^i | \Omega_{t-1}^i] - \mathbb{E}_{t-1}\left[\frac{\partial C_t^i(.)}{\partial Q_t^i} \middle| \Omega_{t-1}^i \right] , \tag{6}$$

to its Carbon Intensity (CI). Its realization relies on four distinct datasets.

Environmental The OPGEE global CI dataset contains information about the emissions of 8,962 children commercial fields in 2015 (Masnadi et al., 2018).

Price The Energy Information Administration dataset on Landed Costs of Imported Crude for Selected Crude Streams contains the future prices for twenty-three oil classes over the time interval 1979-2018 (EIA, 2019). Complementary information about the relative chemical composition of the Selected Crude Streams is imported from the PSA Management and Services BV database (Management & BV, 2019) .

Costs The WoodMac Upstream Data Tool contains information about costs, production and reserves across two levels: 30,235 children and standalone commercial oil fields and 1,916 parent fields over the time interval 1965-2018 (Mackenzie, 2020).

The granularity of the data sources is heterogeneous. The environmental information is available at children level, the price at an oil class level, while the costs both at the children and standalone and at the parent level. Nevertheless, since the children's level costs presents a considerable number of missing values, we decided to preserve the largest amount of information by focusing on the parent and standalone level. Therefore, we harmonize the granularity of all the other data sources to the standalone and parent level. First, we refer to the oil category prices, enhanced by the information of the chemical composition of the category, to approximate the price at which parent and standalone producers sell their oil (see Section Firm Expected Prices). Then, we aggregate the CI at the parent level by computing a weighted average of the 8,817 fields resulting from the harmonizing the OPGEE and the WoodMac IDs (see Section Data Harmonization and IDs Match).

2.1 Firm Expected Prices

The prices at which fields sell their output respond to the timely interaction between demand and supply. In addition, they depend upon some chemical characteristics of the crude, such as the gravity and the sulphur content (Lanza, Manera, & Giovannini, 2005; Fattouh, 2010)[5]. Consequently, the field-level equation,

$$\mathbb{E}_{t-1}[P_t^i|\Omega_{t-1}^i] = \beta_0 + \alpha_t + \beta_2 API^i + \beta_3 S^i \ , \tag{7}$$

reflects how the price at which field i sells depends on time-invariant aggregate factors (β_0), on the demand driven trend at time t (α_t), on the gravity (API^i) and the percentage of sulphur (S^i). The unit of account of β_0 and α_t is \$ /BOE. The same applies to β_2 and β_3 since API and sulphur are dimensionless quantities.
$\mathbb{E}_{t-1}[P_t^i|\Omega_{t-1}^i]$ is unobservable to the researcher, but it can be proxied by the prices of several mixtures of crudes, which group oils coming from different fields into a

[5]A variation of oil prices due to their quality can be thought as the outcome of a Bertrand competition between two or more buyers (typically refineries or intermediaries), which evaluate the crude depending upon its qualities.

tradable oil class. Using the Energy Information Administration dataset and the PSA Management and Services BV database, we collect, respectively, the yearly future prices and the chemical characteristics of twenty-three oil classes over the time interval 1979-2018[6] (EIA, 2019; Management & BV, 2019). Table 1 provides the summary statistics of the future prices, together with the gravity (API) and sulphur content (S). Figure (1) shows how the future prices (FP) of twenty-three oil classes k homogeneously respond to a general trend (α_t), whereas Figure (2) portrays the positive relation between the future price of a category of crude ($FP(k)$) with its API and the negative relation between $FP(k)$ with its sulphur content in a single year (2015).

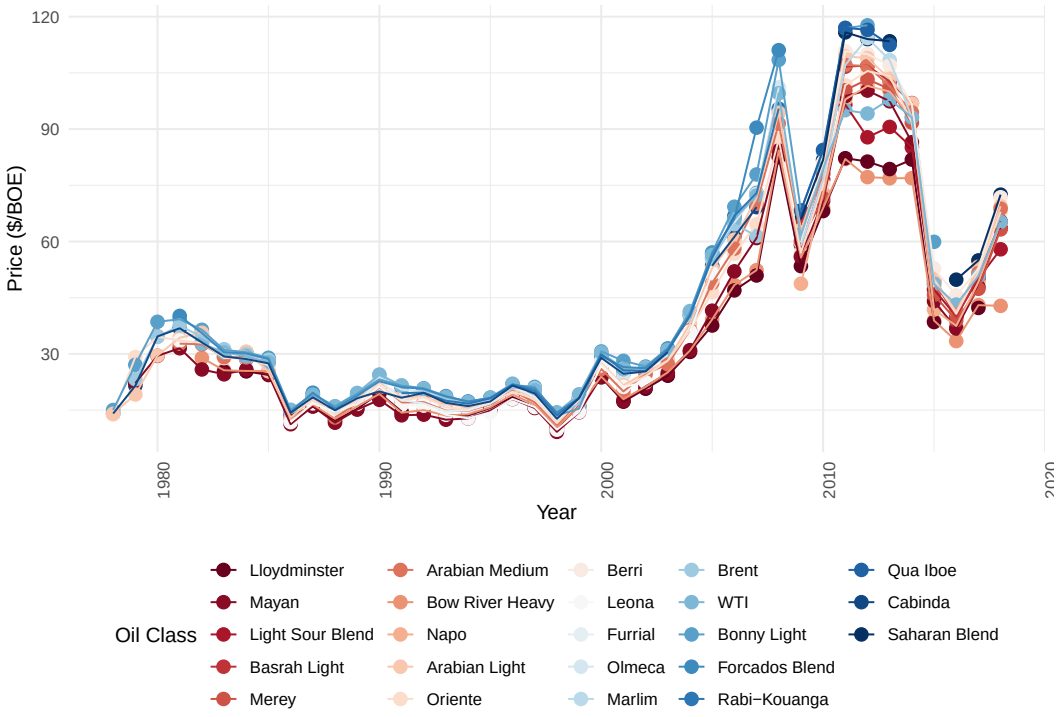

Figure 1: Future prices of twenty-three oil class over the time interval 1979-2018. Colors reflect the Sulphur content, where dark red represents high content whereas dark blue low percentages.

In order to justify the use of such proxy, we assume that the oil class prices equals the weighted average of the field prices belonging to the respective class. In this context, if all the information is publicly available, $\Omega^i_{t-1} = \Omega^{pub}_{t-1}$, the time-varying weights $\{w^i_t\}^{N(k)}_{i=1}$ identify the relative importance of a field belonging to class k in

[6]We take advantage of the full time span of the series to improve the model's fitting.

Table 1: Summary Statistics of $FP_t(k)$ for Twenty-Three Oil Classes.

Oil Class (k)	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

*Sources:*EIA (2019); Management and BV (2019).

period t, such that

$$FP_t(k) = \mathbb{E}_{t-1}\left[\sum_{i\in k} w_t^i P_t^i | \Omega_{t-1}^{pub}\right] = \mathbb{E}_{t-1}\left[\sum_{i\in k} w_t^i P_t^i | \Omega_{t-1}^{i}\right] . \tag{8}$$

Thanks to the assumption entailed in equation 8[7], we can stack the different time series of the oil classes price into a longitudinal dataset and we can estimate regression

$$FP_t(k) = \beta_0 + \alpha_t + \beta_2 API(k) + \beta_3 S(k) + \varsigma_t(k) , \tag{9}$$

where, $\mathbb{E}_t[\varsigma_t(k)|API(k), S(k)] = \mathbb{E}_t[\sum_{i\in k}(w_t^i - w^i)(\beta_2 API^i + \beta_3 S^i)|API(k), S(k)] = 0$ is the random error component. Table 2 reports the results using a Pooled Ordinary Least Square (POLS) ($\alpha_t = 0$) and a time fixed effect (FE) ($\beta_0 = 0$). The

[7]Please notice that equation 8 only serves as a theoretical justification of our next steps and it is not estimated.

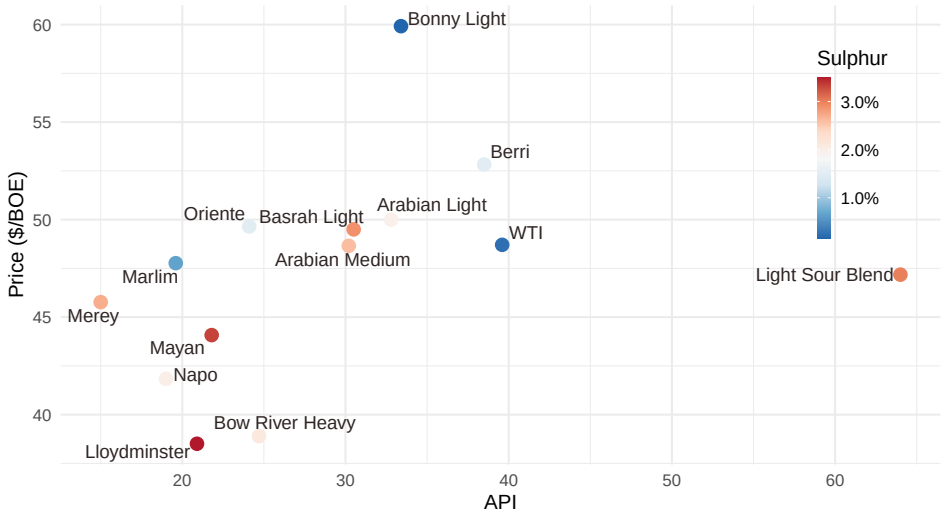

Figure 2: 2015 prices of fourteen oil classes based on their API and Sulphur content.

sole presence of gravity and of sulphur is unable to explain a significant portion of the variance of the future price of the different oil classes since it ignores the demand driven trend. The introduction of the latter (α_t) substantially increases the adjusted R^2 from 2% to 97%, while returning a sign of β_3 in line with economic theory. When controlling for the yearly effect, the time trend spans from a minimum of 12.98 $ / BOE in 1998 to a maximum of 104.70 $ / BOE in 2012. Furthermore, a unit increase in the API augments the value of the crude by 0.07 $ / BOE, while a 1% increase in sulphur content decreases the value of the oil by 2.21 $ / BOE.

Table 2: Estimated Parameters of the Pricing Function

	Dependent Variable: Future Price $/BOE	
	POLS	Time FE
Intercept	20.62**	
	(6.44)	
$API(k)$	0.50**	0.07**
	(0.17)	(0.03)
$S(k)$	3.19**	-2.21***
	(1.21)	(0.19)
Time FE	No	Yes
Adjusted R^2	0.02	0.97
Note:		*p<0.1; **p<0.05; ***p<0.01.

We use the estimates in column 2 to obtain the field-level prices. For example, in 2015, when $\hat{\alpha}_{2015} = 49.80$ \$ / BOE, the estimated price at which field i containing oil with $API = 55$ and $S = 0.03$ would be,

$$\mathbb{E}_{t-1}[\hat{P}_t^i|\Omega_{t-1}^i] = \underset{(1.55)}{49.80} + \underset{(0.03)}{0.07} * 55 - \underset{(0.19)}{2.21} * 0.03 = \frac{53.58\$}{\text{BOE}}. \tag{10}$$

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 $\hat{\beta}, \hat{\alpha}$ are unbiased estimators of β, α. Figure (3) shows the resulting field-level mapping for the year 2015.

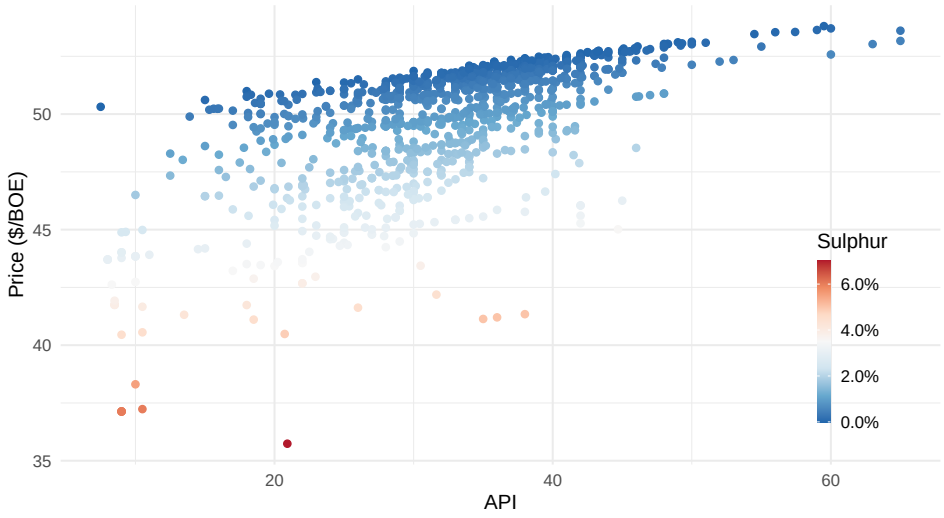

Figure 3: 2015 estimated prices of 1,103 oil fields.

2.2 Data Harmonization and IDs Match

This subsection describes the harmonization of the OPGEE and the WoodMac IDs necessary to link the shadow prices to their emissions. The WoodMac Upstream Data Tool contains historical data for 20,522 children and standalone commercial oil & gas fields (Mackenzie, 2020). The OPGEE global carbon intensity models the emissions of 8,962 commercial children fields. The OPGEE field IDs were harmonized to the Wood Mac ones with the following steps:

- Harmonization of the apostrophes and of the countries' names (`Sudan` vs `South Sudan`, `Republic of Congo` vs `Congo, Brunei` vs `Brunei Darussalam`)
- Partition of the `Sacha` field into `Sacha (pre-2016)` and `Sacha (post-2016)`

- Aggregation of fields `Garzad A`, `Garzan B` and `Garzan C` into a unique field labeled `Garzan`
- Removal of the field `Columba BD` since it was overlapping with two separate fields `Columba B` and `Columba D`
- Aggregation of the fields `G-Ain Dar`, `G-Haradh`, `G-Hawiyah`, `G-Shedgum`, `G-Uthmaniyah` into a unique field named `Ghawar`.

The first attempt to join the two datasets, using only the fields ID, matches 8,632 out of 8,962 fields. Next, we perform a smart string search using `R` commands `str_dectect` and `stringdist_left_joint` to recover 24 out of the 325 unmatched fields. Then, we match by-hand 29 fields checking the field names spelling and their geographical coordinates. Finally, we assign the CI measure for the representative Canadian field `Other heavy (conv. production)` to all the unmatched Heavy and Extra Heavy Canadian fields. In the same way, we assign the CI measure for the representative Canadian field `Other conventional` to all the unmatched Light & Medium and Other Oil Canadian fields. At the end, we were able to match 8,817 fields out of the initial 8,962.

The WoodMac Upstream Data Tool almost exclusively provides balance sheet data at the parent or standalone level. In order to pair the granularity of the two datasets, we compute the volumes weighted average CI, measured in $KgCO_2e$ per BOE,

$$CI^i = \frac{\sum_{j=1}^{J} Q_t^j CI^j}{\sum_{j=1}^{J} Q_t^j} \quad \forall j \in i \ ,$$

of 6,841 children fields[8]. The result is a techno-economic dataset containing the accounting and the environmental characteristics of 995 parent fields and 2,062 standalone fields.

After restricting our sample to fields which have production bigger than zero and provide information about their OPEX and CAPEX expenditures, the final dataset follows 2,062 fields, across 77 countries, over the decade 2009-2018. Our panel contains 17,494 data points[9] and covers an average of 72.57 percent of the global oil demand as estimated by the Organization for Economic Co-operation and Development (OECD). The dataset does not only cover a large fraction of the global

[8]Please notice that no action as been taken for the remaining standalone fields.

[9]Note that a balanced dataset would have had $N \times T = 2,062 * 10 = 20,620$ data points. The discrepancy originates because the 15.16 percent of the fields are not observed at every point in time because they either stop or start operations during the studied period.

oil demand, but it is also representative of the supply in terms of field location (On Shore, Shallow Water, Deepwater, Ultra-Deepwater) and of the chemical and geological peculiarities. Namely, it contains information about low viscosity oil trapped in impermeable rocks (Shale & Tight Oil), low viscosity oil trapped in permeable rocks (Light & Medium Oil), high viscosity oil trapped in permeable rocks (a.k.a. Heavy & Extra Heavy Oil), and oil sands, see Table 3.

Table 3: Absolute Frequency of different Geological Formations

	On Shore	Shallow Water	Deepwater	Ultra-Deepwater
Light & Medium	682	480	118	31
Heavy	102	47	10	5
Shale & Tight	339	7	0	0
Sands	27	0	0	0
Extra Heavy	9	0	0	0
Other Oil	195	6	4	1

2.3 Marginal Extraction Cost

Extraction costs are the sum of all the expenditures faced to get the oil out from the ground. Exploration costs are the sum of all the expenditures faced to find new oil. WoodMac classifies costs into twenty-three categories, see Table 4 . We sum the first twenty-one of them to obtain the extraction costs,

$$C_t^i = \text{Abandonment Costs}_t^i + \text{Capital Receipts}_t^i + ... + \text{Terminal}_t^i \ ,$$

and the last two to obtain the exploration costs,

$$W_t^i = \text{Development Drilling}_t^i + \text{Exploration and Appraisal}_t^i \ .$$

Figure 4 shows the relative importance of the different cost categories, where fixed costs ($\sim 30\%$) and variable costs ($\sim 20\%$) already account for a half of the total costs a field faces. Figure 4 reflects Equation (4), where θ_0^{Geo} captures the field fixed costs and the joint behavior of $\theta_1^{Geo}, \theta_2^{Geo}$ and θ_3^{Geo} the sum of all the other extraction costs. An explanatory analysis of C_t^i confirms the positive relation of the costs with both the production (see Figure 5) and the reserves (see Figure 6). However, the relations vary across different geological formations. Light & Medium, Heavy and Shale & Tight Oil depict a concave pattern as the quantity

Table 4: Summary Statistics of the Twenty-Three Types of Cost listed in WoodMac.

Cost Type	Number of Observations	Mean	SD	Min	Max
Extraction Costs					
Abandonment Costs	14,196	1.155	9.767	-8.870	378.720
Capital Receipts	401	6.218	59.260	0.000	1,044.790
Country Specific CAPEX	3,859	7.824	60.812	0.000	1,586.930
Country Specific OPEX	1,772	3.957	12.661	0.000	317.390
Field Fixed Costs	18,056	73.583	237.959	0.000	5,175.100
Field Variable Costs	17,701	45.779	117.941	0.000	2,928.380
General and Administrative	2,549	6.388	12.904	0.000	186.450
Insurance	40	0.066	0.414	0.000	2.620
Non Tariff Transport	2,055	20.051	70.269	0.000	931.030
Offshore Loading	854	3.934	20.699	0.000	264.490
Other CAPEX	10,461	19.141	77.250	-235.080	1,805.820
Other Costs	776	52.394	94.010	0.000	1,246.240
Other OPEX	726	9.080	31.826	0.000	212.170
Pipeline	14,491	6.295	32.496	-17.140	1,060.070
Processing Equipment	13,287	30.313	138.371	-32.600	3,191.380
Production Facilities	16,670	40.375	188.193	-40.140	6,260.000
Subsea	3,701	25.614	84.894	0.000	1,378.670
Tariff Gas	6,987	11.527	60.259	0.000	1,678.200
Tariff Oil	11,731	30.491	134.681	0.000	3,378.910
Tariff Receipts	1,604	11.474	25.170	0.000	239.460
Terminal	759	3.400	17.454	0.000	269.040
Exploration Costs					
Development Drilling	17,898	82.588	206.935	0.000	4,495.880
Exploration and Appraisal	123	3.458	12.332	0.000	82.870

*Sources:*Mackenzie (2020).

produced increases, whereas Extra Heavy, Sands and Other Oil depict a convex one. The scenario is a little different when visualizing the relation between the costs and the reserves. Heavy, Shale & Tight and Extra Heavy report a concave pattern, Light & Medium a linear pattern, Other Oil a convex pattern, while Sands a non-linear one.

In equation 4 the dependent variable is always positive defined. In order to avoid the use of generalized linear model, we estimate it in first differences. More precisely, we estimate five versions of equation 4. We start using a Linear Mixed Model, column (1) in Table 5, where all the random coefficients are normally distributed. Then, we add a time effect, such that $\epsilon_t^i = \theta_4^{Geo} Q_t^i \; t + \varepsilon_t^i$, to isolate the

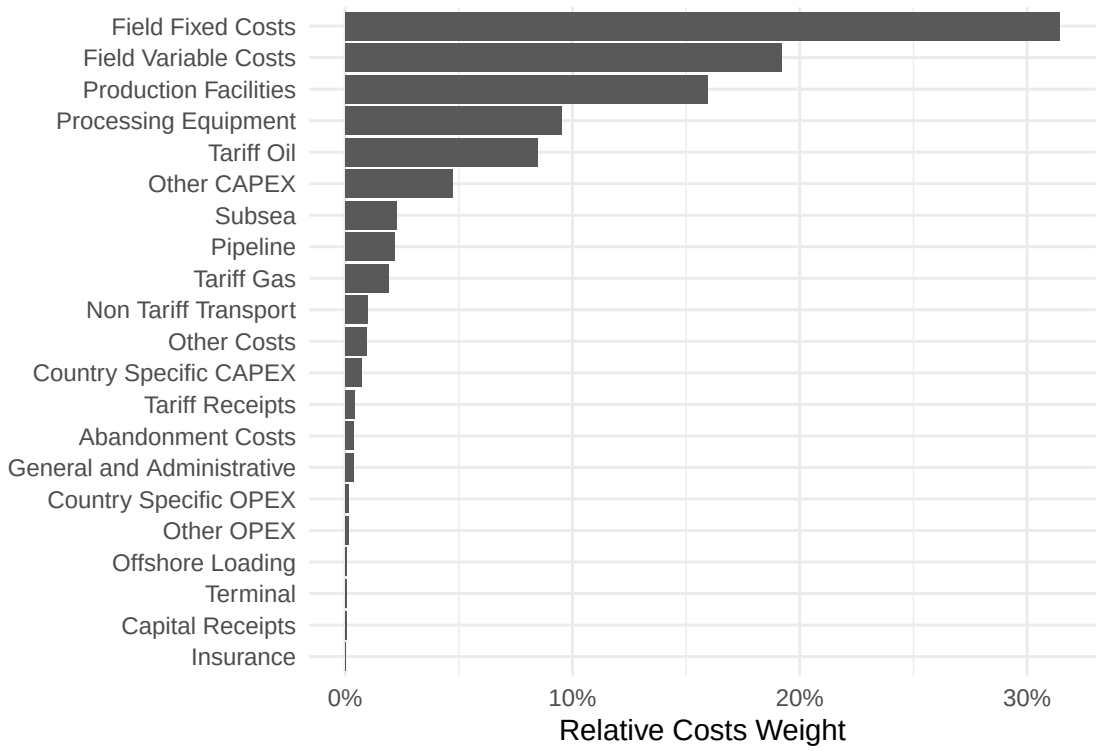

Figure 4: Relative weight of the different cost categories.

impact of technological change, column (2), and a quadratic time effect, such that $\epsilon_t^i = \theta_4^{Geo} Q_t^i \ t + \theta_5^{Geo} Q_t^i \ t^2 + \varepsilon_t^i$[10], to allow the technological improvment to have a (linear) trend. Finally, column (5) combines the three fixed effects, such that $\epsilon_t^i = \varphi_t + Geo^i + \eta^i + \varepsilon_t^i$. In all four cases, we assume $\varepsilon_t^i \overset{iid}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$.

In all cases the coefficients have the expected sign. Namely, both $\hat{\theta}_1^{Geo}$ and $\hat{\theta}_2^{Geo} + \hat{\theta}_3^{geo}/R_{t-1}^i$ are bigger than zero. The introduction of a constant technological effect does not substantially change the magnitude of the coefficients suggesting the absence of a homogenous shock across fields belonging to a same geological class. The introduction of a linear technological trend diminishes the magnitude of $\hat{\theta}_1^{Geo}$, while increasing the overall convexity of the problem while rising the adjusted R^2 from 17% to 23%. Note that, since the regression is estimated in first differences, these goodness-of-fit measures are in line with standard econometric results[11]

[10] Geo^i is a categorical variable which takes values Light & Medium, Heavy, Extra Heavy, Shale & Tight, Sands and Other Oil.

[11] Further diagnostic analysis are avilable upon request of the reader.

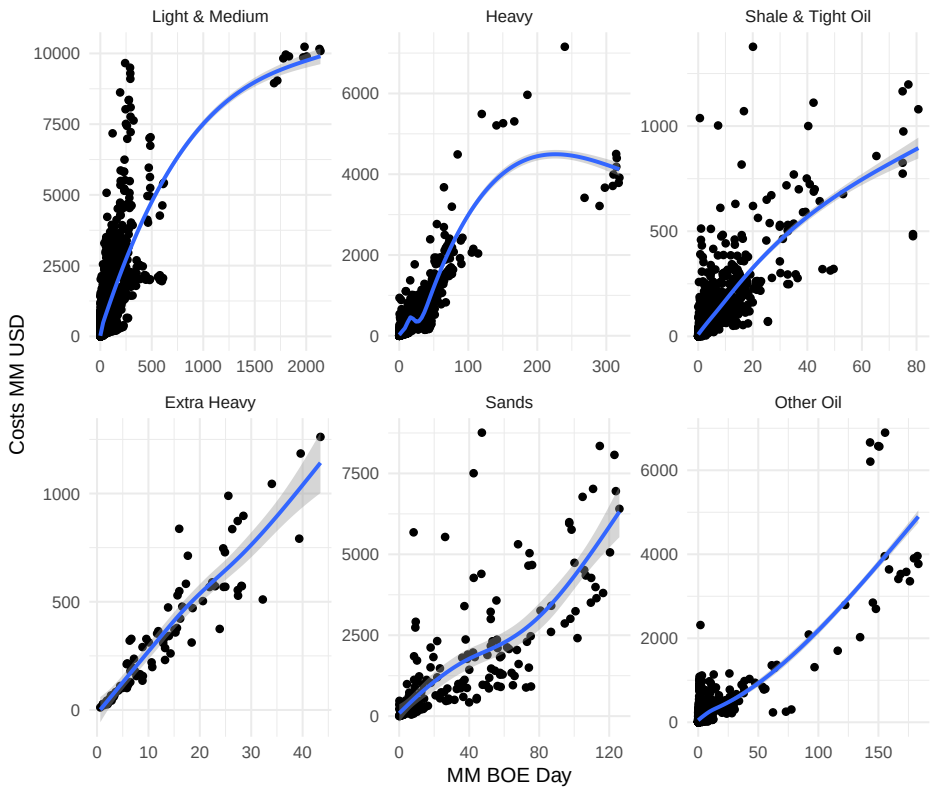

Figure 5: Relation between costs and production for different geological formations.

3 Shadow Prices

In this section we bridge the theoretical with the empirical model. Namely, we conjugate the results obtained in section Firm Expected Prices and Marginal Extraction Costs with equation 6 to obtain

$$\hat{\lambda}_t^i = \underbrace{\hat{\alpha}_t + \hat{\beta}_2 API^i + \hat{\beta}_3 S^i}_{\text{Expected Field Future Price}} - \underbrace{\left(\hat{\theta}_1^{Geo} + \hat{\theta}_2^{Geo} Q_t^i + \hat{\theta}_3^{Geo} \frac{Q_t^i}{R_{t-1}^i} + \hat{\theta}_4^{Geo} t + \hat{\theta}_5^{Geo} t^2\right)}_{\text{Marginal Extraction Cost}} . \tag{11}$$

Equation (11) identifies the profitability of a field at a point in time. In other words, when $\hat{\lambda}_t^i$ is strictly positive the field is lucrative, when it is equal to zero the field breaks-even, and when it is negative the field looses money. Its magnitude is impacted by the demand driven trend at time t (α_t), the chemical qualities of the oil ($\hat{\beta}_2, \hat{\beta}_3$), the volumes of production ($\hat{\theta}_2$) and the volumes of production rescaled by the volumes of recoverable reserves ($\hat{\theta}_3$) and the technological trends with a particular geological class of fields ($\hat{\theta}_4, \hat{\theta}_5$).

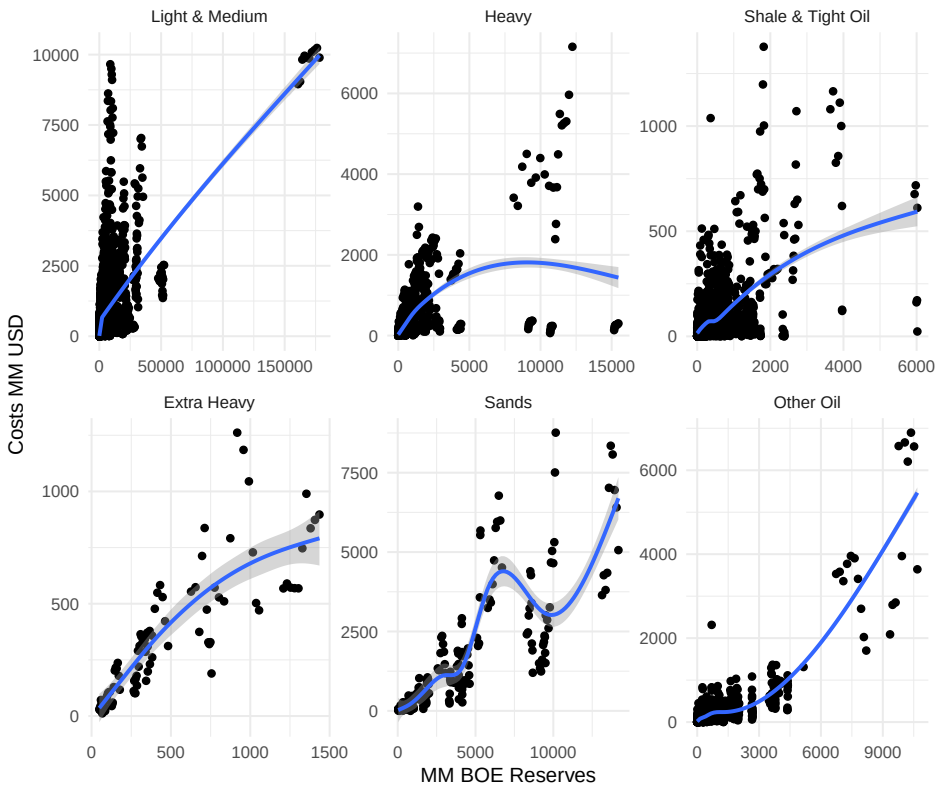

Figure 6: Relation between costs and reserves for different geological formations.

The magnitude of $\hat{\lambda}$ is a versatile measure of the extensive margin of the industry. On a disaggregated level it shows the least profitable fields. For example, according to our estimates, three fields had a negative shadow price (Kucavo-Arreza (Albania), Sabanero (Colombia) and Bozhong Area (China)), while all the others were profitable. Among the profitable ones the least profitable one was West Esh El Mallaha A-1 (Egypt), an On Shore Heavy deposit hosting petroleum with an API fo 20.92 and 7% sulphur content. According to our estimates its shadow price is

$$\hat{\lambda}_t^i = \underset{(1.55)}{49.80} + \underset{(0.03)}{0.07} * 20.92 - \underset{(0.19)}{2.21} * 7 - 34.65 = \frac{1.08\$}{\text{BOE}} \ .$$

Either a decline in selling price or a rise in marginal extraction costs of 1 \$ / BOE would make the field almost unprofitable. In other words this field identifies the extensive margin of the industry in 2015. On a more aggregate level $\hat{\lambda}$ shows which regions of the planet are the least profitable. For example, a volume weighted average of the shadow prices suggests that three countries with the least valuable oil are Albania (23.19 \$ / BOE), Colombia (27.00 \$ / BOE) and Suriname (28.76 \$ / BOE), while among the top ten producers Mexico (35.11 \$ / BOE), Canada

Table 5: Cost Function Regression Results

	Dependent variable: Delta Total Costs C_t^i MM USD		
	(1)	(2)	(3)
Quantity	11.090***	11.062***	10.106***
	(0.162)	(0.161)	(0.157)
Quantity squared	−0.004***	−0.004***	−0.004***
	(0.0001)	(0.0001)	(0.0001)
Reserves squared	127.168***	128.797***	167.830***
	(7.157)	(7.135)	(6.946)
Constant Technology	No	Yes	Yes
Linear Technological Trend	No	No	Yes
Observations	17,503	17,503	17,503
Adjusted R^2	0.16	0.17	0.23
Note:			*p<0.1; **p<0.05; ***p<0.01

(39.34 $ / BOE) and the United States (42.30 $ / BOE) are the ones closest to the margin.

4 Additional Treatments for North American Fields

Additional treatments are conducted on two important global producers based on the available data:

- Canada: generic OPGEE CI estimated are assigned to unmatched Canadian fields using crude type and density (i.e., conventional and heavy conventional).
- United States: many U.S. oil fields (shale tight oils in particular) are not named consistently and systematically. These fields are typically labeled based on their basin, the producer company name, or a combination of both. Therefore, it is very difficult to match these fields between different datasets. In order to improve the U.S. coverage in this work, after a rigorous field-by-field manual matching, two generic oil fields - *US tight oil_generic* and *US heavy_generic* - are created, and volume-weighted-average CI and MC of production of other U.S. matched tight oil and heavy fields are attributed

to these two generic fields. The volumetric production of these two generic fields are estimated based on the missing production volume from the total production of the corresponding crude types using EIA U.S. tight oil production EIA (2020b) and crude oil and lease condensate production by API gravity EIA (2020a) statistics in 2015, respectively.

5 Data coverage

The production coverage summary of the top 20 largest global crude oil producers in 2015 (shown in Table 6) confirms that the dataset used in this study covers different types of crudes from different countries and fairly represents the global oil production market. See the supplementary Excel sheet for list of all countries coverage.

Table 6: Production coverage summary of top 20 largest global crude oil producers in year 2015.

Country	Count of parent fields	2015 total production*, $mmbbl/d$	Coverage in this work, %
Russian Federation	151	10.3	96
Saudi Arabia	15	10.2	100
United States	490	9.4	80
China	65	4.3	97
Iraq	27	4.0	70
Canada	151	3.7	90
Iran	23	3.3	92
United Arab Emirates	8	3.1	44
Kuwait	4	2.8	100
Venezuela	44	2.5	69
Brazil	34	2.4	94
Mexico	19	2.3	99
Nigeria	73	2.2	91
Angola	32	1.8	95
Kazakhstan	35	1.7	84
Norway	56	1.6	100
Qatar	9	1.5	51
Algeria	20	1.4	83
Colombia	59	1.0	91
Oman	10	1.0	100

**Crude oil including lease condensate. Source:*EIA (2016).

6 Additional maps

The geographical location and the qualitative volumetric production magnitude of covered global oil fields are presented in Figure (7). The 2015 carbon intensity map of global oil fields is also illustrated in Figure (8).

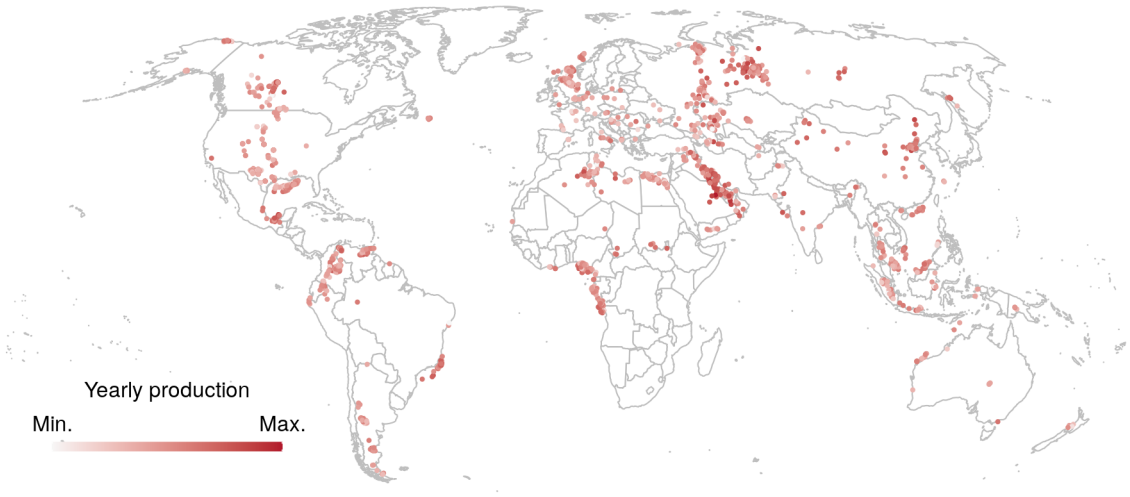

Figure 7: Geographical location and the qualitative volumetric production magnitude of covered global oil fields in 2015

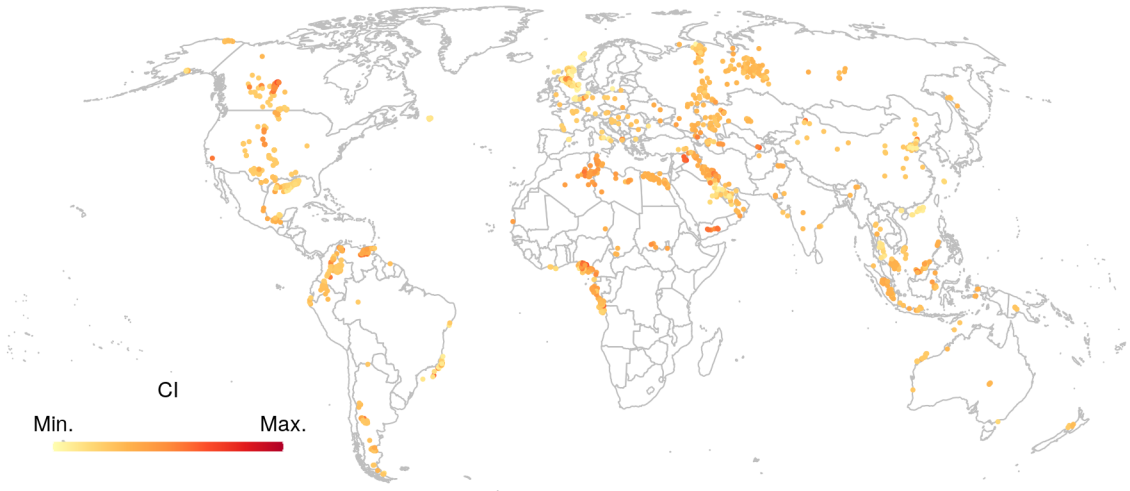

Figure 8: Geographical location and the carbon intensity of covered global oil fields in 2015

References

EIA. (2016). Total petroleum and other liquids production-2016. Retrieved from https://www.eia.gov/beta/international/

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

EIA. (2020a). Crude oil and lease condensate production by api gravity. Retrieved from https://www.eia.gov/dnav/pet/pet_crd_api_adc_mbblpd_m.htm

EIA. (2020b). Tight oil production estimates by play. Retrieved from https://www.eia.gov/petroleum/data.php#crude

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.

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

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

Mackenzie, W. (2020). Wood mackenzie upstream data tool. *WoodMackenzie Upstream Oil & Gas Analysis.(https://www.woodmac.com/our-expertise/capabilities/upstream-oil-and-gas/)*.

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.

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.

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*.