The Long Road to First Oil

Mihalyi, David

Natural Resource Governance Institute, Central European University

15 September 2020

The Long Road to First Oil*

David Mihalyi†

First draft: June 1, 2020
This draft: September 15, 2020.

Abstract

This paper analyzes the factors affecting the speed at which newly discovered oil and gas fields are developed. Using data from over 25,000 oil and gas assets globally I demonstrate that both asset and country characteristics are critical in determining which assets reach production stage. I analyze the effects of countries adopting a set of market oriented reforms, to shed light on the impacts of institutional changes on petroleum extraction timeline.

Mitigating climate change will require a large share of the world's already discovered fossil resources to stay underground. The results of this study can help inform how petroleum producers may respond to the energy transition underway. My findings also calls into question the assumption used in earlier research that giant oil and gas discoveries can be considered exogenous in their impacts on subsequent production.

Keywords: *resource curse, natural resources, institutions, liberalization.*
JEL classification: P50, Q33, Q35

*I am grateful to Anna Augusztin for excellent research assistance. I thank Tamas Bartus, Caitlin Brown, Michael Dorsch, Paasha Mahdavi, Rick Van der Ploeg and participants of the World Bank Africa's Resource Future workshop for their comments .

†David Mihalyi: Natural Resource Governance Institute and Central European University. Email: davidmihalyi@gmail.com.

1 Introduction

When a country makes a major oil or gas discovery, policy makers and citizens alike expect it to bring revenues and economic transformation soon. But the reality is that petroleum finds can take over a decade to reach production, if ever. For example, Uganda had a series of large oil discoveries starting in 2006. The government and petroleum companies initially targeted oil production to start in 2009. However, negotiations around taxes and pipeline routes stalled. After repeated revisions of the timeline, the government now targets oil to first start flowing in 2023. In Kazakhstan, the Kashagan field was discovered in 2000, and though companies invested quickly, it took 13 years for them to developed the field after technical set backs and disputes between participants.

Oil and gas projects require large financial investment and the execution of a complex capital investment program. An industry-intelligence study reviewed 365 oil and gas mega-projects finds that 73% of the projects are reporting schedule delays. Another industry study reports on how performance problems are linked to systematic cost overruns reviewing a sample of 200 petroleum projects (Rui et al., 2017). These delays have large financial implications not only for companies but also governments expecting taxes from production. But despite ample reporting on the topic in the industry and financial press, the factors affecting the path from discovery to extraction have received scant attention in the empirical research on the resource curse.

In this study, I provide systematic empirical evidence on the factors affecting petroleum asset timelines. By looking at odds of production starting before and after the adoption of a set of market oriented reforms, I provide some tentative evidence that institutional weaknesses may cause slower timelines.

The paper is structured as follows. Section 2 places this research in the broader literature and highlights selected papers relying on an assumption that my paper puts into question. Section 3 provides some context on petroleum project timelines for the benefit of those less familiar with the industry. Section 4 describes the data I use in the analysis and some stylized facts on project timelines based on summary statistic. Section 5 presents the empirical analysis results based on various empirical strategies: survival analysis, discrete-time event history and event study. Section 6 concludes.

2 Related economic literature

The relationship between economic growth and resource wealth has been subject to extensive study and debate (for recent surveys see Ross (2015); Van der Ploeg (2011)). An emerging consensus agrees that any overall resource curse effect is best understood as mediated by the quality of institutions (Mehlum et al., 2006; Robinson et al., 2006). They argue that countries with strong political institutions are better placed to reap the benefits of resource wealth, in contrast, countries with weak institutions are more susceptible to the various resource curse mechanisms.

One attribute these studies share is the examination of the relationship between resource wealth's contribution to the economy, typically measured via production value, export dependence or government revenue windfalls, and economic performance.

However, as pointed out by Brunnschweiler and Bulte (2008), resource wealth or dependence may be shaped by past economic performance, policy choices and political institutions. For example, exploration efforts by investors, and therefore the observed pattern of geological wealth, are themselves dependent on institutional factors (Arezki et al., 2019; Cust and Harding, 2019). As a consequence any correlations between resource dependence and economic performance do not prove causality on their own - since there may be other factors causing both the observed level of resources in a country, and its economic or political fate.

Hence many recent studies have analyzed the impact of giant oil and gas discoveries instead of the level of petroleum wealth measured by reserves, production or some other contemporaneous measure of its contribution to the economy. For example, research by Arezki et al. (2016) examines the impacts on macroeconomic variables such as employment, savings, investment and the current account, Cust and Mihalyi (2017) the short-term growth responses, Harding et al. (2020) the impact on relative prices and real exchange rates, Abdelwahed (2020) the impact on domestic taxation, der Ploeg et al. (2019) the trade policy responses and Lei and Michaels (2014) studies armed conflicts following giant discoveries. As argued by the authors of above studies, such discoveries are largely unanticipated 'lucky' events where the within-country timing of individual discoveries may be plausibly exogenous once we account for country and year fixed effects. Countries have very little means to influence the timing of such large discoveries.

Some of the studies above also implicitly or explicitly rely on the assumption that all discoveries are equal in their likelihood and speed to reach production. For example, Arezki et al. (2016), assumes that it takes an average 5 year for a giant oil discovery to turn to production in calibrating their models to derive expected economic impacts from the shock. Similarly both and Harding et al. (2020), Abdelwahed (2020) and der Ploeg et al. (2019) explicitly rely on the assumption that production starts five years after discovery, when interpreting subsequent events as being caused by production. The latter study also includes robustness checks for pre-production periods of 3, 4, 6, 7 and 8 years, but the 5 year difference remains the central estimate.

The assumption of an average 5 year pre-production period is originally posited and discussed in most detail in Arezki et al. (2016). It is supported by the following four pieces of evidence. First, there is a graphical illustration of the production profile including pre-production times from two Norwegian oil fields (exact number of years is unclear but approx. 5 years). Second is an expert estimate cited based on US drilling experience which reports an average of 4-6 years between drilling and production.[1]. Third, Mike Horn, a geologist and author of the giant discovery dataset is quoted suggesting it may take an average of 7 years (no citation). Finally the authors' report calculations based on a subset of giant discoveries using data compiled by Global Energy Systems at Uppsala University which contains both discovery and production dates. This dataset consists of 157 giant fields discovered since 1970 where the average pre-production time is of 5.4 years. But as explained by the authors of the dataset in Höök et al. (2009), the "Fields that have not yet reached their decline phase (as of 2005) are excluded". Therefore the dataset is truncated and the estimate is likely to be downward biased given that it excludes fields that failed to reach peak production in time.

The lack of production start date in the complete giant discovery dataset has led to various workarounds. In their study of the impacts of giant discoveries on conflict, Lei and Michaels (2014) try to establish the likely timing of production

[1]source: Why "Drill, Baby, Drill!" is Not a National Energy Policy by Thimothy D Kailing http://www.ellipticalresearch.com/drillingandoilproduction.html

start by looking at the time lag between giant discoveries and total country-level oil output. They find an increase in production 2 years after discovery, which then remains elevated from year 4 post-discovery on-wards. Though their study attributes the increased oil output to the discovery reaching production, a study by Güntner (2019) finds that this is partly driven by increase in production from other oil fields.

Another relevant paper, by Smith (2015) using a different dataset constructed by the author, looks at the impact of a country's first oil discovery and its subsequent impact on economic growth. Here the author warns of the possibility that certain countries might be slower to get from discovery to production, but ultimately discards this as a minor confounder with regards to long-term (up to 30 years) economic impacts of oil finds. But his estimation also omits all the countries, which had a first oil discovery but did not reach oil production by the end of the time period reviewed. This may bias the estimated impact of discoveries on GDP.

The studies discussed above all assume production automatically starts some years (usually 5) after discovery, and attribute changes observed after that period to petroleum production. While some studies explore the possibility that there may be variance in the number of pre-production years, they do not systematically analyze and control for potential sources of variations in this respect. This probably stems from the limited availability of field level data with both discovery and production year. The sources cited to estimate discovery to production time period suffer from limited geographical scope (US, Norway) or are in fact a truncated sample of fields that have reached peak production within a certain time frame. This research presents significantly different estimates from a larger global dataset.

Some researchers analyzed the expected economic impacts of the projected onset of production at the level of a single country. For example Mozambique discovered large amounts of gas in 2011. Toews and Vezina (2016) model the expected FDI response and Melina and Xiong (2013) model the optimal investment path. These two studies assume production starts within 5 and 9 years respectively. In reality the project is stalled, and latest estimates are now of 11 years. In contrast, Henstridge (2018) studies the costs associated with the extended delays in the gas projects in Tanzania.

While the latter study is a notable exception, country level research on the expected impact of newly found resource wealth often devotes limited attention as to when (if at all) an oil discovery will be turned to production. This research provides more reliable estimates of the expected pre-production period based on key country and asset level characteristics.

While exploration has been more concentrated in countries with stronger institutions and more openness to trade, Arezki et al. (2019) also documents how there is increased exploration activity across developing countries in recent decades. On the other hand, climate researchers warn that in order to avoid catastrophic climate change, a large share of already discovered oil and gas wealth has to stay underground. For example, McGlade and Ekins (2015) calculates that one third of current oil reserves and half of gas reserves must remain in the ground to meet the 2C target. Similarly, IEA (2015) also forecasts that 50 per cent of oil reserves and 40 per cent of gas reserves need to remain in the ground to stay within the 2C target. Even under some of the slower energy transition scenarios they forecast, many hydrocarbon assets already discovered are likely to be stranded and remain underground.

What factors determine which country's oil has higher likelihood of being developed? Previous analysis, such as Manley et al. (2017) looks at the number of years it would take to deplete reserves based on past recovery rate, while Mercure et al. (2018) and McGlade and Ekins (2015) looks at regional drilling costs associated with extraction. Their analyses are focused on established producers, where both costs and depletion rates are known. My research enables to expand the analysis both in terms of global coverage and by looking at how additional factors such as having state-owned company in charge of extraction, having other assets in the country already producing or adopting market oriented reforms may influence likelihood of project development.

3 Context - The journey from discovery to production

In this section, I provide a description of the steps involved in getting from discovery to production as a background to the subsequent analysis.[2]

Around the world petroleum companies regularly acquire licenses or permit to explore a certain area for oil and gas. Once they have obtained such rights, they may conduct geological and geophysical surveys and carry out exploratory drilling in promising locations. If they do not find anything for a number of years, they are typically required to give up on these rights (relinquish their license) so governments can bring in new companies to carry out exploration. In case of a successful oil find, the company has the right keep the license and develop the asset.

The life of an oil and gas asset, such as those in our database, starts an exploration well strikes oil or gas, hence a new field is *discovered*. After an initial discovery, the companies enter the *appraisal phase*, when further wells labelled appraisal wells or delineation wells are drilled, with the motive of assessing the size and viability of the initial find. Many successive wells may be drilled depending on the results of drilling. The appraisal may take several years to complete.

After appraisal, the next stage is the *feasibility study*. This is the phase in which the initial concept for an oil and gas project is developed. The study identifies the resources, how much (roughly) the project would cost, and where the money to finance it would come from and what the returns may be on the project. If more than one company is developing an oil or gas resource, companies set out the basic structure of a joint venture, including the stakes each company will have and which of them will be the operator, leading the consortium of companies. In many countries, a local company or the state-owned oil gas firm is required to be a joint venture participant. Such negotiations may be protracted.

Next companies need to obtain all the necessary *permits* and file all required doc-

[2] This section draws heavily on Rystad database's handbook and an industry explainer from Oilprice.com https://oilprice.com/Energy/Energy-General/The-Complete-Guide-To-FIDs.html

umentation related to the project, including environmental impact assessments (EIAs) and route permits from authorities. The respective regulators have to approve the project before companies can proceed with any actual construction work. Contentious permitting issues may include the route of pipeline, water use, gas flaring. Permit approval can get delayed or requests may be rejected, requiring change of plans. The Front End Engineering and Design (FEED) stage sets in details the technical and financial options reviewed in the feasibility study. The FEED examines the technical requirements and provides an estimate of the overall project costs and the costs of each phase, with support from engineering contractors. For massive oil and gas projects, FEED contracts typically take around a year to complete.

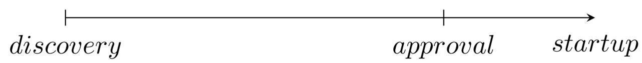
The next big milestone, which we also record in our database, is the *approval*. It designates the when year the asset was approved/sanctioned for development. This is the point in an energy project in which the company or companies owning and/or operating the project approve—or sanction—the project's future development. This is often labelled Final Investment Decision (FID) in the industry press. Typically, it is the board of directors of a company involved in an oil and/or gas project who makes the Final Investment Decision for a project.

After approval, companies start developing the project, a phase labeled *Engineering, procurement, construction (EPC)*. In EPC, engineering includes basic and detailed engineering, planning, construction engineering. Procurement includes procurement, purchasing, invoicing, logistics and transport. Construction includes civil engineering, electrical installation, and mechanical installation. Project development may see unexpected setbacks in any number of these activities.

Finally, the project reaches its *start-up*, the third milestone recorded in the database, when the petroleum recovery begins. This episode is often labelled reaching first-oil or first-gas.

Once production started, production can be halted (labelled shut-in), though this is rarely done due to associated costs. Once most of the oil is extracted from an asset, and any further extraction is no longer commercially viable, then wells are plugged and the asset is *abandoned*. I do not analyze the life of an asset beyond when production starts.

The below graph provides a simple depiction of the stages I analyze using the database. It also highlights that on average, the period from discovery to approval is longer than the period from approval to start.



4 Data description and stylized facts

I rely primarily on a large proprietary database by Rystad, an independent energy research and business intelligence company providing data and related consultancy services to the global energy industry. Their Ucube (Upstream) Database consists of a complete asset-by-asset (field-by-field) database of the world's known oil and gas resources. Though their database includes petroleum fields discovered as far back as 1900 and forecasts for future resources expected to be found (by country) up to 2100, I limit my analysis to the over 25,000 assets discovered between 1960 and 2019 based on the availability of complementary datasets.

For each petroleum asset I retrieve its year of discovery, the year of approval when the asset gets green light for development, and startup when the field reaches production stage, where these stages were reached.[3] A dummy records fields that are yet to reach approval and production stages. I also calculate the number of years the asset has spent without producing, using the year 2020 for the assets that are yet to reach production. This variable takes the minimum value of 0 when production started in same year as the discovery happened and its maximum is 60 years for an asset discovered in 1960 that is yet to reach production as of 2019.[4]

Table 1: Summary statistics for all discoveries

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
Producing	0.690	0.462	0	1	25823
Approved	0.699	0.459	0	1	25823
Start_Disc_Producing	7.377	8.931	0	60	17824
Appr_Disc_Producing	5.787	8.173	0	56	17824
Start_Appr_Producing	1.589	2.093	0	43	17824
Start_Disc_All	10.643	12.743	0	60	25823
Appr_Disc_All	9.512	12.666	0	60	25823
Start_Appr_All	1.608	2.177	0	56	18161

Table 1 provides summary statistics on all assets discovered between 1960 and 2019. First, I show the ratio of assets that reached its start up stage *(Pro-*

[3]For assets not yet granted approval or not yet producing, the Rystad database also provides some forecasts, but I ignore these.

[4]In the survival analysis set up presented below I add one to the year variable to avoid having 0s which are not compatible with the specification.

ducing) and those that passed approval stage *(Approval)*. It shows that 69 percent reached production, while marginally more 70 percent have been approved. Then I show the years between discovery and start up stage *(Start-disc-Producing)*, discovery and approval *(Appr-Disc-Producing)* and approval and start up *(Start-Appr-Producing)* for all assets that have reached production. It takes on average 7.4 years to get from discovery to production among producing assets, of which 5.8 is getting from discovery to approval stage, and another 1.6 from approval to startup. Finally, I show the values for the same variable, but on the full sample but using 2019 for those that have not (yet) started producing *(Start-disc-All)*, *(Appr-Disc-All)*, *(Start-Appr-All)*. The average asset in the full sample has spent about 11 years not producing, and almost 10 years not reaching approval stage. (For assets that are yet to reach approval, *(Start-Appr-All)* does not exist, hence its average value is similar to the producing only sample).

Table 2: Summary statistics for giant discoveries

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
Producing	0.704	0.457	0	1	1158
Approved	0.727	0.446	0	1	1158
Start_Disc_Producing	11.432	11.428	0	54	815
Appr_Disc_Producing	8.774	10.226	0	49	815
Start_Appr_Producing	2.658	2.826	0	39	815
Start_Disc_All	15.929	15.36	0	60	1158
Appr_Disc_All	14.007	15.331	0	60	1158
Start_Appr_All	2.637	2.828	0	39	844

I also provide the same descriptive statistics in Table 2 for the subset of assets (fields) where the estimated volume of petroleum resource discovered exceeds 500 million barrels, the threshold used to denote giant discoveries. It shows that only 70 percent of giants have reached production, a similar ratio to the full sample. Most giant discoveries that reached approval stage have also started production. The pre-production period is over 11 years across the giant discoveries that ultimately reached production stage and nearly 16 years when also considering assets not yet producing. These values are well above the timelines presented on the full sample of discoveries. It takes 2.8 years to get from approval to the start of production, considerably more than the 1.6 for all discoveries, but still a relatively short period within the full timeline from discovery to the start of production.

These figures are relevant and present a stark contrast to the growing literature presented in section 2 on the impacts of giant discoveries. [5] As opposed to the 5 year pre-production period average assumed in multiple studies, this dataset suggests the period is mover 12 years for those that have reached production and a third of the fields are yet to be developed. The large difference in averages is most likely attributed to the fact that earlier studies used evidence of limited geographical scope and truncated data by Höök et al. (2009) only looking at fields which reached peak production within a certain period.

Table 3: Summary stats - all assets - regional breakdown

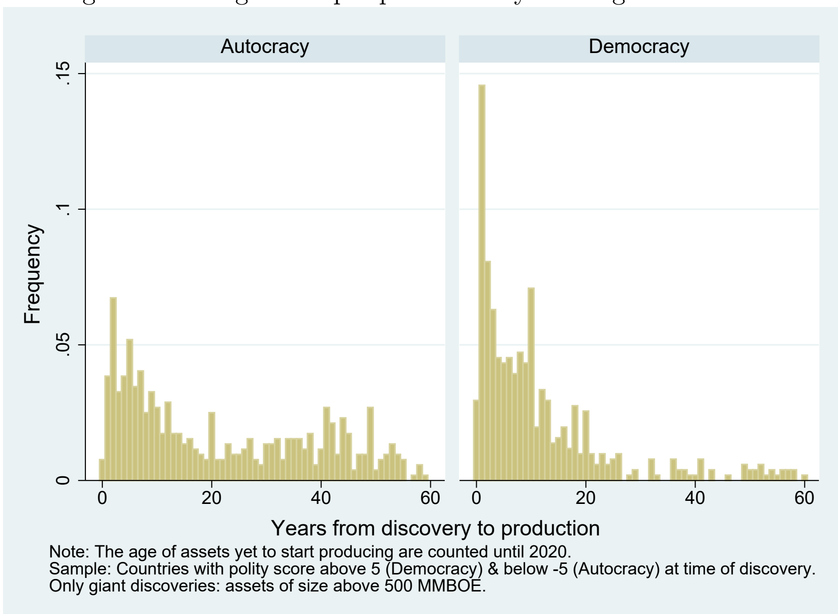
Region (World Bank classif.)			Mean			N
	Prod.	Appr.	Start_Disc_P	Appr_Disc_P	Start_Appr_P	
East Asia and Pacific	0.53	0.54	8.4	6.9	1.5	4,075
Europe and Central Asia	0.74	0.74	9.6	7.7	1.9	6,985
Latin America & Carib.	0.73	0.74	6.0	4.9	1.2	2,895
MENA	0.57	0.58	9.3	7.5	1.8	2,514
North America	0.84	0.84	4.3	3.0	1.4	6,713
South Asia	0.62	0.65	7.7	5.9	1.8	1,023
Sub-Saharan Africa	0.46	0.47	12.1	10.2	1.9	1,618
Total	0.69	0.70	7.4	5.8	1.6	25,823

Table 3 provides summary statistics of the regional breakdown of the ratio of assets that started producing *(Prod.)*, were approved *(Appr.)* the average year between discovery and production *(StartDisc)*, discovery and approval *(ApprDisc)* and approval and production*(StartAppr)* for the assets that have already started producing, as well as the number of assets in each region *(N)*. It shows that there is large variation between regions, with assets in North America on average being developed more than twice as quickly as assets in Sub-Saharan Africa.

As shown in1 the data on the giant discoveries sub-sample also reveals stark differences in pre-production periods in democracies and autocracies. Whereas the mean years between discovery and production (or 2020 for non-producing assets) is 11 years in fields discovered in democracies (polity score above 5 on -10 to 10 scale), it is almost double or 21 years in autocracies (polity score below -5 on -10 to 10 scale).

[5] The giant discovery sub-sample I present is not identical to Horn (2011). Though both datasets measure this using the expected ultimate recovery (EUR) of the fields in barrels of oil equivalent at time of discovery, they rely on different underlying data sources and probably different geological assumptions used in calculations. For the comparable 1960 - 2010 period,

Figure 1: Histogram of pre-production years of giant discoveries



Given that the giant discovery sub-sample is much smaller, the remainder of the analysis looks at the full sample of discoveries.

For each field, I also collect a range of geologically significant characteristic from the Ucube database. These are the size of the field measured in the log of the total barrel of oil and gas resources (*(Asset-Sum-ln)*, the log of the water-depth of the field *ln-waterdepth)*, the ratio of oil vs gas found *(OiltoSum and GastoSum)*, whether the asset is shale or not *(Shale-dummy)*, whether the field is operated by a domestic state-owned company *(OperatorGov)*.

I supplement the dataset with some country level characteristics. These are the polity scores by Polity IV Project on the level of democracy *(polity2)* and the log of the per-capita level of GDP *(ln-gdp-pwt)* from the Penn World Tables. Another characteristic is whether the country is is already producing oil or gas *(pre-prod)*, as additional finds might be quicker to come online if certain infrastructure are in place.

I also add the log of the nominal Brent oil price series from the World Bank com-

there are 756 giant discoveries in Horn (2011), while there are 1059 in Rystad's Ucube dataset.

modity data tables *(lnOilPrice)* and a year variable to capture any technological progress.

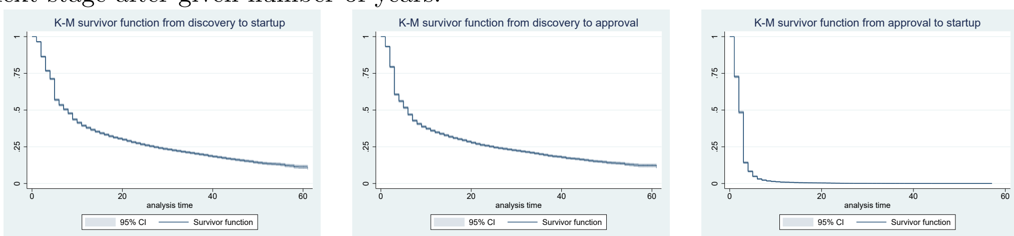
For each asset, the time varying variables can be measured at time of discovery, production start or any other year. See descriptive statistics with time varying variables measured at discovery year in Table 4.

Finally, my main explanatory variable in my empirical estimations will be a country's turn towards market orientation or openness. For this I follow Arezki et al. (2019) in using data on the timing of economic liberalization during the years 1960–2004. This data was originally constructed by Sachs and Warner (1995) and revised and extended by Wacziarg and Welch (2008), then further extended by Arezki et al. (2019). Following Sachs and Warner (1995), the following criteria are used to classify a country as open: (i) the average tariff rate on imports is below 40%; (ii) non-tariff barriers cover less than 40% of imports; (iii) the country is not a socialist economy (according to the definition of Kornai (1992)); (iv) the state does not hold a monopoly of the major exports; and (v) the black market premium is below 20%. As a result they obtain a dichotomous variable, where the country is deemed open in a given year if it satisfies all of these above criteria. Else, if it does not meet either of these criteria, it is characterized as closed. While this indicator was originally designed to capture openness to trade, I follow Arezki et al. (2019) and Buera et al. (2011) by viewing this indicator as a proxy for capturing the timing of a broader set of reforms targeting economic openness and market orientation.

Table 4: Description and summary statistics of additional variables

Variable	Type	Mean	Std.Dev	Description
Asset_Sum_ln	float	2.733	1.864	Log value of the asset's size in barrel of oil equivalent EUR resources
Gas_to_Sum	float	0.464	0.391	Percentage gas resource volume (vs oil) in total asset resource volume
ln_WaterDepth	float	1.478	2.228	Log value of the asset's underwater depth (Onshore is 0)
Shale_dummy	dummy	0.071	0.256	1: Shale asset; 0: Not shale asset
Operator_Gov	dummy	0.318	0.466	1: If the company operating the field is state-owned, 0: if not.
ln_GDP_discovery	float	13.61	1.90	GDP of country at year when asset was discovered. (Source: PWT)
polity2_discovery	double	4.314	7.225	Polity index of country in year when asset was discovered
Region_WB	cat.			Regional variables groups (WB)
Facility type	cat.			1: Fixed 2: Floater 3: Onshore 4: Subsea tie back
Oil_price_discovery	double	33.620	30.483	Oil price at year when asset was discovered
Oil_price_startup	double	41.068	32.234	Oil price at year when asset started its production.
Prod_PreStart	dummy	0.987	0.113	1: There are some already producing assets in that country at start of asset production; 0: None
Open_PreStart	dummy	0.661	0.474	1: Country is open when the asset started production; 0: Country is closed.

Figure 2: Kaplan-Meier estimate on the likelihood of an asset not moving to next stage after given number of years.



5 Empirical strategy and analysis

I carry out econometric analyses regarding the factors that affect the speed and likelihood of a petroleum asset being developed. I use various estimation techniques, including a survival analysis, a discrete-time event-history analysis and an event study. The approach extends on Khan et al. (2016) who analyzes similar issue in the mining sector.

5.1 Survival analysis

Survival analysis is an empirical method used most frequently in epidemiology. It allows to define a failure event, which in the case of epidemiology is often a patient's death, but in this instance it is when the oil asset starts production (which one may consider labeling a success rather than a failure). The survival function provides an estimate on the likelihood of an oil field remaining untapped over the years after discovery.

Survivor function plots using Kaplan–Meier estimator

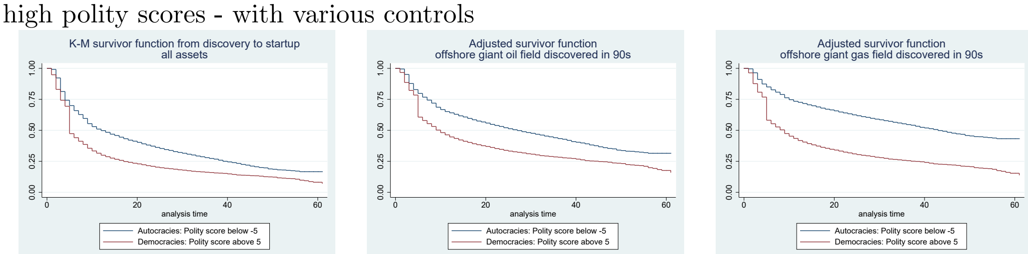
I employ the Kaplan-Meier non-parametric estimator (Kaplan and Meier, 1958) of the survivor function, which provides a simple way to evaluate the fraction of observations, which have remained undeveloped after a number of years. A value of close to 1 means that an average asset of certain age is almost certainly not producing, while close to zero means almost certainly producing. The Kaplan-Meier estimator allows to split the sample into groups and to control for certain characteristics.

I present the K-M estimates for the three different periods in Figure 2. First the full period from discovery to the start of production, then followed by discovery until approval and third is the approval to start up phase. The steepest - so quickest and most likely among them - is going from approval to startup stage.

By way of example, I also show the K-M estimates for my main period of analysis, from discovery to start of production comparing assets located in countries with weak versus strong institutional scores. On the one hand one may speculate that weaker institutional settings have less ability to execute complex all petroleum projects. Conversely, it is possible that consolidated autocracies are better able to fast track important infrastructure projects by discarding local resistance to the project.

The first plot shows that assets found in countries with lower polity scores at time of discovery (below -5 on -10 to 10 range) are significantly slower to develop than those with high scores (above 5 on -10 to 10 range). I also present results which controls for certain geological characteristics taking the same values to more closely capture the differences associated with country characteristics rather than geology. As shown in the second plot of Figure 3, there is a large difference in timeline across institutional scores when comparing only offshore giant oil fields discovered in the 90s. That difference increases even further when comparing fields that are mostly gas. The difference between oil and gas may be attributable to the fact that gas finds requires complex auxiliary infrastructure (either to liquefy for transportation or converting it to electricity or heating). The odds of offshore giant gas fields being developed within 20 years is about half when located in countries with weak institutions at time of discovery rather than one with strong institutional score, see plot 3 of Figure 3. Altogether, the above evidence finds that countries with weaker institutions are slower to execute petroleum projects.

Figure 3: Timeline from discovery to startup for assets in countries with low vs high polity scores - with various controls



Survival model- regressions

In order to evaluate the significance of individual variables on project timeline, there are a number of regression types to consider. Within survival analysis set-up, this can take the form a semi-parametric model, such as the Cox regression or a parametric model, such as the Gompertz, Weibull, lognormal, exponential, etc. I first present results from the cox model, which is followed by results from multiple parametric models.

Key results from analysis

I first present results from a Cox regression of the following form.

$$h_i(t) = h_{0i}(t) exp(\beta_1 X_1 + ... + \beta_k X_k), \quad (1)$$

where $h_i(t)$ is the hazard rate for asset i over time (t) following its discovery, in other words the rate at which the asset reaches production and X_1 - X_k are series of explanatory variables.

Results are shown in Table 5. A first specification looks at a set of asset level geological characteristics. It show that shale assets are quicker, while larger, more deep water projects and those producing mostly gas rather than oil are slower (result are of mixed sign and significance on this latter variable).

In a second specification I add country level variables shows that richer countries and those with stronger institutions at the time of discovery are quicker. I also show that when a government entity operates the field, it will be slower. (Most of the variance on this variable is at the country level depending on how the sector is regulated). Finally, I add time variables. The oil price at the time of discovery is negatively associated with speed (which may be because asset development decisions are based on future oil price expectations and not the ones at discovery). On the other hand discoveries in later years are associated with quicker timelines.

I replicate these tables for the two sub-periods I distinguish. Getting from discovery to approval stage (specification 1-3 in Table 6) and then from approval stage to startup (specification 4-6 in Table 6). Results are similar, though there are some differences as well. For example, shale's advantage seems to come from being quicker in getting from approval to start rather than from discovery to approval.

Table 5: Results from Cox regressions

VARIABLES	(1) Disc - Start	(2) Disc - Start	(3) Disc - Start
Shale_dummy	2.548***	1.558***	1.191***
	(0.0692)	(0.0527)	(0.0394)
Asset_Sum_ln	0.943***	0.986***	1.028***
	(0.00387)	(0.00492)	(0.00465)
Gas_to_Sum	1.019	0.827***	0.812***
	(0.0190)	(0.0184)	(0.0173)
ln_WaterDepth	0.919***	0.882***	0.870***
	(0.00348)	(0.00398)	(0.00429)
Operator_Gov		0.883***	
		(0.0189)	
ln_GDP_discovery		1.088***	
		(0.00681)	
ln_Oilprice_discovery		0.959***	
		(0.0143)	
DiscoveryYear		1.006***	
		(0.00130)	
polity2_discovery		1.013***	
		(0.00152)	
Country FE	No	No	Yes
Observations	25,823	17,831	25,823

seEform in parentheses

*** p<0.01, ** p<0.05, * p<0.1

I also ran a number of different forms of parametric models, alongside the Cox model on the timeline from discovery to startup. Results are presented in Table 8 of the Annex. Results are very similar across the 9 specification after taking to account that specification 1-4 presents in Table 8 are results in terms of proportional hazard (meaning a value above 1 is a quicker timeline), while models 5-9 in Table 8 are accelerated failure time models (where a value below 1 is a quicker timeline).

Model selection and limitations

In order for the results from the semi-parametric cox model to hold, they need to satisfy the so-called proportional-hazards assumption. That means that each

Table 6: Results from Cox regressions

VARIABLES	(1) Disc-Appr	(2) Disc-Appr	(3) Disc-Appr	(4) Appr-Start	(5) Appr-Start	(6) Appr-Start
Shale_dummy	2.139***	1.346***	0.970	2.994***	2.214***	2.786***
	(0.0575)	(0.0453)	(0.0319)	(0.0852)	(0.0796)	(0.0976)
Asset_Sum_ln	0.949***	0.990**	1.028***	0.959***	0.974***	0.985***
	(0.00387)	(0.00490)	(0.00461)	(0.00387)	(0.00506)	(0.00462)
Gas_to_Sum	1.040**	0.845***	0.843***	0.900***	0.849***	0.837***
	(0.0192)	(0.0186)	(0.0178)	(0.0172)	(0.0193)	(0.0186)
ln_WaterDepth	0.929***	0.893***	0.886***	0.950***	0.933***	0.918***
	(0.00343)	(0.00393)	(0.00423)	(0.00375)	(0.00442)	(0.00502)
Operator_Gov		0.892***			0.958**	
		(0.0188)			(0.0203)	
ln_GDP_discovery		1.087***			1.020***	
		(0.00674)			(0.00609)	
ln_Oilprice_discovery		0.952***			1.024	
		(0.0140)			(0.0152)	
DiscoveryYear		1.005***			1.002*	
		(0.00127)			(0.00127)	
polity2_discovery		1.013***			1.003	
		(0.00150)			(0.00155)	
Country FE:	No	No	Yes	No	No	Yes
Observations	25,823	17,831	25,823	18,161	13,271	18,161

seEform in parentheses

*** p<0.01, ** p<0.05, * p<0.1

covariate has a multiplicative effect in the hazards function that is constant over time. This assumption does not hold for the time varying controls.(Results not shown in this draft).

The various parametric functions I presented are more flexible in this regard, they do not require such assumption to hold. But the parametric functions need selection to ascertain best fit. This can be done using the Akaike Information Criterion (AIC). As reported in the last row of Table 8 in the Annex, the AIC test suggests that the best fitting model is the one relying on a gamma distribution (column 9 which has the lowest AIC number).

This approach has shown that various geological, country-related and time-related factors are associated with significant differences in production timeline. Assets located in countries with higher institutional scores at time of discovery are quicker to be developed. I obtain quantitatively similar results using a number of specification of survival models.

5.2 Discrete-time event-history

I also analyze the data using a discrete-time event-history model setup. In this approach all years when the asset is not producing are considered a separate observation with an additional observation for the year the asset starts up production. I create a panel consisting of each asset across the years observed until startup. A dummy variable codes for whether the asset started producing in a given year or not yet *(Start)*. Using the startup event as my dependent variable, I run a random-effects panel regression model. This approach allows to include time-varying explanatory variables for every year of the asset's pre-production life instead of having to pick a single year for each asset (e.g. the discovery year, as done in the survival analysis).

I estimate the linear model using an asset-year panel regression presented in Equation 2. I use robust standard errors clustered at the country-level for experimental design reasons: the level of treatment (liberalization) is at the country-year level, while observations are at asset-year level (Abadie et al., 2017). In various specifications I include country-level fixed effects, year fixed effects.

$$Start_{c,i,t} = \beta_0 + \beta_1 Open_{c,t-1} + \beta_2 age + \beta_3 age^2 + \beta_4 Z_{c,i,t} + \alpha_c + \delta_t + \epsilon_{c,i,t} \quad (2)$$

where $Start_{i,c,t}$ represents a dummy, which takes the value of 1 if asset i in country c is opening in year t. The main variable of interest is $Open_{t-1,c}$, taking a value of 1 if the country c opened up in the preceding year. I also include an asset age variable age and age squared age^2 variable to capture the fact that the oil field has a decreasing likelihood of opening as years progress. A series of control variables are denoted Z. α_c denotes country time-invariant characteristics, while δ_t captures common time varying effect.

I use this approach to test for the significance of including the adoption of market oriented reforms as an explanatory variable in a way that may include assets that have spent some years in a closed economy and some years in an open economy. This is the variable constructed by Wacziarg and Welch (2008) described in the data section.

Key results from analysis

The results displayed in Table 7 show the effects of various variables on the likelihood of an oil asset reaching start up stage in any given year. This model set up uses an additional age variable *(t)* and age squared *(tsq)* variable to capture the fact that the oil field has a decreasing likelihood of opening as years progress. While the likelihood of opening drops sharply in the initial years it later decelerates. Additional controls used in earlier regressions are also included.

The new insight comes from the inclusion of a dummy variable on whether the country is open or closed at any point in time *(Open-state)*. I first run a logistic panel regression with random effects (1). I then replicate the regression using a linear panel model (2). This followes on (Angrist and Pischke, 2008) who suggest that a linear model is more straightforward to analyze than a logistic model especially when dealing with small changes in likelihoods. Specification (3) adds year fixed effects and specification (4) also adds country fixed effects to the regression.

The switch from closed to open is associated with a substantial increase in chances of opening up in the first three specifications presented but disappear when adding country fixed effects. The latter null result may be a result of not having enough within-country variation in openness, as most countries do not switch at all, and other countries once.

Table 7: Results from discrete-time event-history model

	(1)	(2)	(3)	(4)
	xtlogit	xtreg	xtreg	xtreg
VARIABLES	Start	Start	Start	Start
_t	-0.0588***	0.00557*	0.00382	0.00434
	(0.00446)	(0.00301)	(0.00273)	(0.00277)
tsq	0.000454***	-0.000120*	-9.21e-05	-0.000108*
	(0.000107)	(7.14e-05)	(6.09e-05)	(6.09e-05)
Shale_dummy	0.359***	0.0911***	0.0789***	0.0281
	(0.0380)	(0.0312)	(0.0265)	(0.0495)
Asset_Sum_ln	-0.0593***	-0.00923**	-0.00804**	-0.00149
	(0.00572)	(0.00378)	(0.00376)	(0.00467)
Gas_to_Sum	-0.138***	-0.0306	-0.0263	-0.0268
	(0.0259)	(0.0248)	(0.0223)	(0.0264)
ln_WaterDepth	-0.128***	-0.0153***	-0.0143***	-0.0197***
	(0.00575)	(0.00318)	(0.00286)	(0.00271)
ln_GDP_PWT	0.169***	0.0152**	0.0141*	-0.00358
	(0.00722)	(0.00667)	(0.00722)	(0.00858)
polity2	-0.00273	-0.000403	-0.000297	0.000548
	(0.00239)	(0.00117)	(0.00109)	(0.000679)
Open_state	0.462***	0.0544***	0.0548***	-0.00691
	(0.0351)	(0.0201)	(0.0167)	(0.0100)
Constant	-4.286***	-0.0671	-0.145	0.208**
	(0.111)	(0.109)	(0.132)	(0.0929)
lnsig2u - Constant	-3.542***			
	(0.897)			
Observations	154,045	154,045	154,045	154,045
Number of assetid	16,111	16,111	16,111	16,111
Random effects	Yes	Yes	Yes	Yes
Year FE	No	No	Yes	Yes
Country FE	No	No	No	Yes
Robust SE	NA	Yes	Yes	Yes
Cluster	NA	Country	Country	Country

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

The association between country openness and increased likelihood of asset startup is not necessarily causal. A third factor may both contribute to openness and quicker timelines, but these initial results do show that assets are quicker once such liberalization event took place.

5.3 Event study

The event study approach allows to estimate changes in likelihood of an asset reaching production in the years surrounding a particular event. In this case, I present results from analyzing likelihoods of production start in the years before and after the opening up event. This followed on the earlier section. I use a random effect panel model with robust standard errors clustered by country and with year fixed effects alongside controls for asset characteristics (t, tsq, $Shale\_dummy$, $Asset\_Sum\_ln$, Gas_to_sum, $ln\_WaterDepth$) but not country characteristics. On top of that, I add a dummy for all possible lags and leads to the liberalization events. I use a limited sample of assets discovered at a time when the country was closed (9,222 assets).

Key results from analysis

The results presented in Figure 4 provide a clear indication that there is a jump in likelihood of assets turning to production in the years following a country opens up.

The figure depicts how chances of an asset starting up varies in the 5 years prior to and up to 10 years after a country opens up. The reference year used, where the coefficient is manually set to zero, is the year prior to opening up and the results presented for all other years are in comparison to this one. While there are no strong trends in the 5 years prior to opening up, there is an immediate jump in the year of liberalization which stays positive in the105 years after, though its statistical significance is mixed (see bars showing 90 percent confidence intervals).

The results appear immediately in the year of opening up and although fluctuate, remain strong in the 10 year window after opening. This makes the results even more visible when looking at cumulative impact over 10 years (bottom plot). Having included year fixed effects should capture spurious correlations in case years with more liberalization events globally coincided with more project start ups. The robust standard errors clustered at the country level should ensure that the results are not overly driven by very few liberalizing countries with many assets.

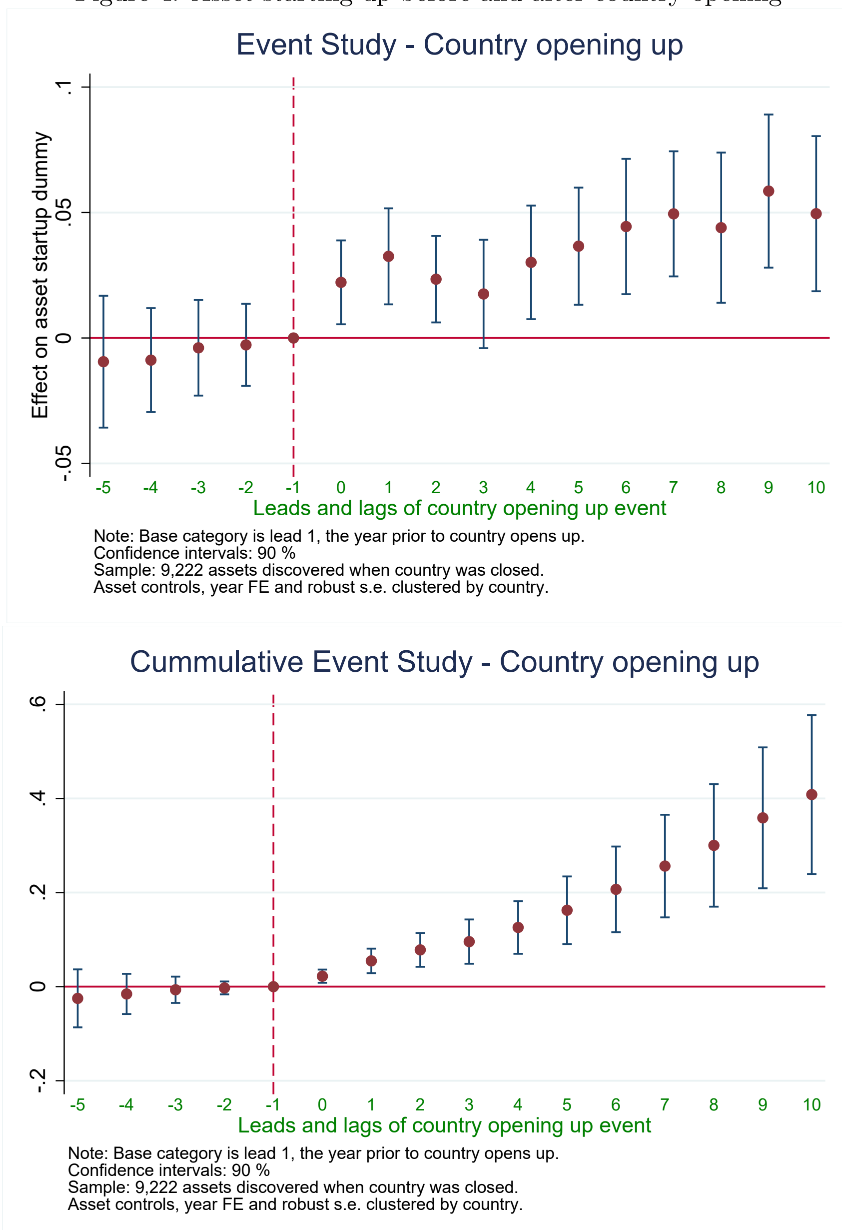
I have presented evidence showing that a range of factors influence the speed at which oil assets are being developed. Assets located in countries with stronger institutions are developed quicker. Assets also experience an increased likelihood once the country adopts major institutional reforms. This phenomenon can be observed markedly across assets before such reforms: there is a jump in likelihood of the asset opening up in the 10 years following such events.

6 Conclusion

In this paper I make two distinct contributions. First, I presented a detailed analysis on the geological, institutional factors and time trends that influence the speed at which petroleum assets are being developed globally. I provide evidence on institutional reforms being followed by (though not necessarily directly causing) an increase in likelihood of oil field development. These results may help inform analysis of future petroleum exploitation, which is especially critical given the energy transition underway.

Secondly, my analysis calls into question some results from earlier economic research using giant oil discoveries as exogenous shocks to a country's subsequent oil production and revenues. I find that oil discoveries on average take over twice as long to be developed than earlier economic research typically assumed and with large variation depending on institutional factors. As a result, this earlier research underestimated the importance of pre-production impacts of oil discoveries especially in countries with weaker institutions.

Figure 4: Asset starting up before and after country opening



References

Abadie, A., S. Athey, G. W. Imbens, and J. Wooldridge (2017). When should you adjust standard errors for clustering? Technical report, National Bureau of Economic Research.

Abdelwahed, L. (2020). More oil, more or less taxes? new evidence on the impact of resource revenue on domestic tax revenue. *Resources Policy 68*, 101747.

Angrist, J. D. and J.-S. Pischke (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Arezki, R., V. A. Ramey, and L. Sheng (2016). News shocks in open economies: Evidence from giant oil discoveries. *The Quarterly Journal of Economics*.

Arezki, R., F. van der Ploeg, and F. Toscani (2019). The shifting natural wealth of nations: The role of market orientation. *Journal of Development Economics 138*(C), 228–245.

Brunnschweiler, C. N. and E. H. Bulte (2008). The resource curse revisited and revised: A tale of paradoxes and red herrings. *Journal of environmental economics and management 55*(3), 248–264.

Buera, F. J., A. Monge-Naranjo, and G. E. Primiceri (2011). Learning the wealth of nations. *Econometrica 79*(1), 1–45.

Cust, J. and T. Harding (2019, 06). Institutions and the Location of Oil Exploration. *Journal of the European Economic Association 18*(3), 1321–1350.

Cust, J. and D. Mihalyi (2017). Evidence for a presource curse? oil discoveries, elevated expectations, and growth disappointments. *World Bank Policy Research Working Paper 8140.*

der Ploeg, R. V., F. Perez-Sebastian, and O. Raveh (2019, December). Oil Discoveries and Protectionism. Economics Series Working Papers 895, University of Oxford, Department of Economics.

Güntner, J. H. (2019). How do oil producers respond to giant oil field discoveries? *Energy Economics 80*, 59–74.

Harding, T., R. Stefanski, and G. Toews (2020). Boom Goes the Price: Giant Resource Discoveries and Real Exchange Rate Appreciation. *The Economic Journal 130*(630), 1715–1728.

Henstridge, M. (2018). Understanding the boom: Country study—tanzania. *World Institute for Development Economic Research (UNU-WIDER) Working Paper*.

Höök, M., B. Söderbergh, K. Jakobsson, and K. Aleklett (2009). The evolution of giant oil field production behavior. *Natural Resources Research 18*(1), 39–56.

Horn, M. M. K. (2011). Giant oil and gas fields of the world.

IEA (2015). World energy outlook 2015.

Kaplan, E. L. and P. Meier (1958). Nonparametric estimation from incomplete observations. *Journal of the American statistical association 53*(282), 457–481.

Khan, T., T. Nguyen, F. Ohnsorge, and R. Schodde (2016). *From commodity discovery to production*. The World Bank.

Kornai, J. (1992). *The socialist system: The political economy of communism*. Oxford University Press.

Lei, Y.-H. and G. Michaels (2014). Do giant oilfield discoveries fuel internal armed conflicts? *Journal of Development Economics 110*, 139–157.

Manley, D., J. F. Cust, and G. Cecchinato (2017). Stranded nations? the climate policy implications for fossil fuel-rich developing countries. *OxCarre Policy Paper 34*.

McGlade, C. and P. Ekins (2015). The geographical distribution of fossil fuels unused when limiting global warming to 2 c. *Nature 517*(7533), 187–190.

Mehlum, H., K. Moene, and R. Torvik (2006). Institutions and the resource curse. *The economic journal 116*(508), 1–20.

Melina, M. G. and Y. Xiong (2013). *Natural gas, public investment and debt sustainability in Mozambique*. Number 13-261. International Monetary Fund.

Mercure, J.-F., H. Pollitt, J. E. Viñuales, N. R. Edwards, P. B. Holden, U. Chewpreecha, P. Salas, I. Sognnaes, A. Lam, and F. Knobloch (2018). Macroeconomic impact of stranded fossil fuel assets. *Nature Climate Change 8*(7), 588–593.

Robinson, J. A., R. Torvik, and T. Verdier (2006). Political foundations of the resource curse. *Journal of development Economics 79*(2), 447–468.

Ross, M. L. (2015). What have we learned about the resource curse? *Annual Review of Political Science 18*, 239–259.

Rui, Z., F. Peng, K. Ling, H. Chang, G. Chen, and X. Zhou (2017). Investigation into the performance of oil and gas projects. *Journal of Natural Gas Science and Engineering 38*, 12–20.

Sachs, J. D. and A. M. Warner (1995). Natural resource abundance and economic growth. Technical report, National Bureau of Economic Research.

Smith, B. (2015). The resource curse exorcised: Evidence from a panel of countries. *Journal of Development Economics 116*, 57–73.

Toews, G. and P.-L. Vezina (2016). Resource discoveries and fdi bonanzas. *OxCarre Working Papers 199*.

Van der Ploeg, F. (2011). Natural resources: Curse or blessing? *Journal of Economic Literature 49*(2), 366–420.

Wacziarg, R. and K. H. Welch (2008). Trade liberalization and growth: New evidence. *The World Bank Economic Review 22*(2), 187–231.

7 Annex

Table 8: Results from various parametric regressions w AIC test

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
VARIABLES	Cox	exp	gom	wei	logl	llog	logn	ln	ggammma
Shale_dummy	1.271***	1.233***	1.264***	1.225***	0.846***	0.846***	0.801***	0.801***	0.814***
	(0.0519)	(0.0507)	(0.0518)	(0.0502)	(0.0408)	(0.0408)	(0.0380)	(0.0380)	(0.0331)
DiscoveryYear	1.001	1.011***	1.001	1.008***	0.999	0.999	0.997**	0.997**	0.999
	(0.00105)	(0.00103)	(0.00104)	(0.00104)	(0.00123)	(0.00123)	(0.00118)	(0.00118)	(0.00103)
Asset_Sum_ln	0.998	1.003	0.999	1.001	1.015**	1.015**	1.013**	1.013**	1.027***
	(0.00514)	(0.00510)	(0.00512)	(0.00511)	(0.00659)	(0.00659)	(0.00631)	(0.00631)	(0.00586)
Gas_to_Sum	0.876***	0.886***	0.877***	0.884***	1.293***	1.293***	1.274***	1.274***	1.284***
	(0.0204)	(0.0205)	(0.0204)	(0.0205)	(0.0365)	(0.0365)	(0.0346)	(0.0346)	(0.0311)
ln_WaterDepth	0.944***	0.940***	0.940***	0.942***	1.074***	1.074***	1.081***	1.081***	1.103***
	(0.0157)	(0.0157)	(0.0156)	(0.0157)	(0.0195)	(0.0195)	(0.0196)	(0.0196)	(0.0172)
Prod_PreStart	1.428***	1.454***	1.432***	1.446***	0.533***	0.533***	0.558***	0.558***	0.556***
	(0.131)	(0.134)	(0.132)	(0.133)	(0.0557)	(0.0557)	(0.0567)	(0.0567)	(0.0479)
Facility: Floater	1.068	1.076	1.063	1.073	1.040	1.040	1.021	1.021	1.110
	(0.0754)	(0.0758)	(0.0748)	(0.0756)	(0.0812)	(0.0812)	(0.0808)	(0.0808)	(0.0766)
Facility: Onshore	1.237***	1.285***	1.235***	1.269***	0.691***	0.691***	0.717***	0.717***	0.779***
	(0.0811)	(0.0846)	(0.0809)	(0.0835)	(0.0503)	(0.0503)	(0.0516)	(0.0516)	(0.0480)
Facilty: Subsea tie back	0.650***	0.597***	0.630***	0.609***	1.732***	1.732***	1.640***	1.640***	1.278***
	(0.0335)	(0.0307)	(0.0324)	(0.0313)	(0.0962)	(0.0962)	(0.0896)	(0.0896)	(0.0611)
Operator_Gov	0.940***	0.937***	0.938***	0.939***	1.094***	1.094***	1.089***	1.089***	1.083***
	(0.0208)	(0.0207)	(0.0208)	(0.0208)	(0.0283)	(0.0283)	(0.0269)	(0.0269)	(0.0230)
PWT_GDP_discovery	1.000***	1.000***	1.000***	1.000***	1.000***	1.000***	1.000***	1.000***	1.000***
	(2.51e-09)	(2.51e-09)	(2.51e-09)	(2.52e-09)	(2.77e-09)	(2.77e-09)	(2.76e-09)	(2.76e-09)	(2.43e-09)
Oil_price_discovery	0.998***	0.999*	0.999	0.999**	1.001	1.001	1.001	1.001	1.000
	(0.000503)	(0.000502)	(0.000501)	(0.000502)	(0.000574)	(0.000574)	(0.000560)	(0.000560)	(0.000474)
polity2_discovery	1.017***	1.022***	1.018***	1.020***	0.979***	0.979***	0.981***	0.981***	0.993***
	(0.00149)	(0.00149)	(0.00149)	(0.00149)	(0.00173)	(0.00173)	(0.00164)	(0.00164)	(0.00149)
Constant		0***	0.0165**	4.43e-09***	255.9**	255.9**	6,051***	6,051***	54.09**
		(0)	(0.0339)	(9.04e-09)	(622.7)	(622.7)	(14,087)	(14,087)	(109.6)
AIC		51571.78	50209.17	51324.52	49522.54	49522.54	48946.71	48946.71	47829.56
Observations	17,013	17,013	17,013	17,013	17,013	17,013	17,013	17,013	17,013

seEform in parentheses

** p<0.01, ** p<0.05, * p<0.1