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: March 21, 2021

Abstract

The road from a petroleum discovery to production is long, especially in developing countries. On average they take 7 years with a standard deviation of 9 years and a quarter of the fields are yet to reach production. I analyze the drivers of petroleum project timelines using survival analysis and event study methods. Institutions are a key factor. Democracies and state owned firms operating domestically are significantly quicker. My findings suggest earlier research relying on lagged impacts of giant petroleum discoveries provided biased estimates of the effects of oil production shocks.

Keywords: *natural resources, institutions, national oil companies*
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, Gerhard Toews and participants of the World Bank Africa's Resource Future workshop for their comments.
†David Mihalyi: Natural Resource Governance Institute. 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 and and the government now targets oil to first start flowing in 2024. 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. A number of other countries which made important petroleum discoveries subsequently failed to turn these into production (Mihalyi and Scurfield, 2021).

I study the factors affecting petroleum asset extraction timelines. My aim is to untangle geological characteristics of the fields and global time trends from characteristic that are influenced by producer country institutions. My research takes advantage of a unique global dataset with data on project timelines for over 25,000 petroleum fields discovered since 1950 across the globe. I use two different methodologies for the analysis: survival analysis and an event study approach.

Petroleum assets can be slow to be developed. On average, it took oil and gas fields that were developed 7 years to reach production with a standard deviation of 9 years. A quarter of the fields are yet to reach production stage. Giant discoveries take twice the time to turn to production than conventionally assumed in economic literature: the pre-production period is of over 10 years rather than the 5 years used in a number of economic studies.

Asset and country characteristics both matter. For example gas and deep(er) offshore fields are slower to be developed. Those located in countries which are richer, have a longer history of petroleum production or that have stronger institutions are quicker. For example, a similar giant gas discovery is twice as likely to remain underground within a 20 year window if found in an autocracy compared to a democracy.

State ownership also matters. Assets with partial ownership by the domestic national oil company are quicker to be developed once controlling for other factors, but these national oil companies are associated with slower timelines on their projects abroad. I also study how the likelihood of assets starting production changes in the years surrounding the the (partial) nationalization of the industry. Setting up of a national oil company is followed by an about 20 percent increase in likelihood of projects starting up in the subsequent period.

My results call into question the findings from earlier research which treated (giant) petroleum discoveries as causing exogenous shocks to subsequent production by assuming uniform petroleum project timelines. The impacts these earlier studies capture typically five years after giant discoveries underestimate the effects of oil production and are skewed towards measuring production impacts in countries where fields are developed quicker. Alternatively, some of the effects they estimate (e.g. increase in conflict or borrowing) may in fact have happened prior to production. [1]

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 calls 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 key characteristics of project timelines based on summary statistic. Section 5 presents the empirical analysis results based on two distinct empirical strategies: survival analysis and an 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 re-

[1] Cust and Mihalyi (2017) discuss how oil finds may impact a country's development prior to production start, a phenomenon dubbed the 'presource curse'.

source 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, Perez-Sebastian et al. (2021) 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.

The studies above also implicitly or explicitly rely on the assumption that all discoveries are equal in their likelihood and speed to reach production. Arezki et al. (2016), Harding et al. (2020), Abdelwahed (2020) and Perez-Sebastian et al. (2021) use the assumption that production starts five years after discovery, when interpreting subsequent events as being caused by petroleum production. Many of the studies also includes robustness checks, for example Perez-Sebastian et al. (2021) also looks at pre-production periods of 3, 4, 6, 7 and 8 years, with the 5 year being their 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.[2]. 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 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 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

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

countries, which had a first oil discovery but did not reach oil production by the end of the time period reviewed.

Some researchers analyzed the impact of discoveries at the level of a single country or in a region and have more explicitly tackled heterogeneity in project timelines. Toews and Vézina (2018) analyzes the impact of large gas discovery in Mozambique on FDI, while Henstridge (2018) studies the expected benefits of large gas discovery in Tanzania. Both these studies acknowledge the extended delays and uncertain faith of these gas projects, neither of which has reach production 10 years later. A study by Anderson et al. (2018) evaluates the impact of oil prices on extraction decision in Texas. Merrill and Orlando (2020) assesses how violence influences extraction decision in the Middle East.

While these latter studies are notable exceptions, 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. My research provides more reliable estimates of the expected pre-production period based on key country and asset level characteristics.

My research also sheds new light on the role of national oil companies. Previous research by Mahdavi (2014) identifies a number of factors which drive governments to set up such companies, including a desire to extract larger revenues from existing production. Brunnschweiler and Bulte (2008) assesses the impact of such companies on resource exploration and finds that national ownership leads to fewer discoveries.

My results could be applied to study the energy transition. Over the past 35 years, for every barrel of oil extracted globally, approximately two have been added to estimates of proved oil reserves. (Dale and Fattouh, 2018). On the other hand, climate researchers warn that in order to mitigate 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 or 'stranded' to meet the 2C target. When studying which country's reserves are most likely to be stranded, earlier research relied on estimated drilling costs associated with extraction, e.g Mercure et al. (2018), McGlade and Ekins (2015) and Manley and Heller (2021), while Manley et al. (2017) relied on past recovery rates.

Institutional factors may influence which country's hydrocarbon reserves become stranded. My research enables to study future energy transition scenarios assuming various geological and institutional factors continue to excerpt similar influence as they did in the past. Economists have warned that there may be a green paradox, where profit-maximizing oil companies decide to accelerate fuel extraction in anticipation of a shift to renewable energy (Sinn, 2008), (Van der Ploeg and Withagen, 2012). My findings would suggest that in fact, state-owned companies may extract even quicker.

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.[3]

Around the world petroleum companies regularly acquire licenses or permit to explore a certain area for oil and gas. These licenses/permits provide the fiscal and regulatory terms for their operations. 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.

[3]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

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 and gas firm is required to be a joint venture participant. The oil companies may request the revision of initial terms from the government in order to make the project commercially more viable. Such negotiations may be protracted.

Next companies need to obtain all the necessary *permits* and file all required documentation 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.

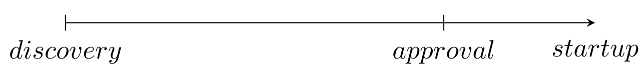
The next big milestone, which I also record in the database, is the *approval.* It designates the when year the asset was approved/sanctioned for development. This is the point in an energy project in which the company or companies owning and/or operating the project approve—or sanction—the project's 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 Descriptive statistics on project timelines

I analyze the production timelines of petroleum projects using a global dataset of oil and gas fields. I rely primarily on a large proprietary database by Rystad Energy [4] Their Ucube (Upstream) Database consists of a complete asset-by-asset database of the world's known oil and gas resources. Though their database includes petroleum fields discovered as far back as 1900, I limit my analysis to the 27,690 assets discovered between 1950 and 2020 based on the availability of complementary datasets. Of these I also drop 729 observations which are labelled extensions, expansions or consecutive phases of existing assets.[5] In the remaining cases, Rystad's definition of an asset is generally equivalent to a petroleum field. This results in a total dataset size of 26,961 petroleum assets.

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.[6] 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 70 years for an asset discovered in 1950 that is yet to reach production as of 2020.[7]

Table 1: Summary statistics for all discoveries

Variable	Mean	Std. Dev.	Min.	Max.	N
Producing	0.758	0.428	0	1	26961
Approved	0.77	0.421	0	1	26961
Start_Disc_Producing	6.984	9.234	0	70	20445
Appr_Disc_Producing	5.46	8.33	0	63	20445
Start_Appr_Producing	1.524	2.404	0	53	20445
Start_Disc_All	10.765	13.018	0	70	26961
Appr_Disc_All	9.575	12.88	0	70	26961
Start_Appr_All	1.547	2.472	0	56	20751

[4] Rystad is an independent energy research and business intelligence company providing data and related consultancy services to the global energy industry.

[5] I drop them as their timeline are not indicative of first oil or gas from a given discovery.

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

[7] In the survival analysis set up presented below I add one to the number of years between dates to avoid having 0s which are not compatible with the specification.

Table 1 provides summary statistics on all assets discovered between 1950 and 2020. First, I show the ratio of assets that reached its start up stage *(Producing)* and those that passed approval stage *(Approval)*. It shows that 76 percent reached production, while marginally more 78 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 years to get from discovery to production among producing assets, of which 5.5 is getting from discovery to approval stage, and another 1.5 from approval to startup. Finally, I show the values for the same variable, but on the full sample but using 2020 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. The average value for *(Start-Appr-All)* is similar to the producing only sample, as few of the assets have reached approval but not yet producing.

Table 2: Summary statistics for giant discoveries

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
Producing	0.775	0.418	0	1	1284
Approved	0.798	0.401	0	1	1284
Start_Disc_Producing	10.367	11.244	0	63	995
Appr_Disc_Producing	7.984	10.114	0	59	995
Start_Appr_Producing	2.383	2.646	0	39	995
Start_Disc_All	13.73	14.533	0	70	1284
Appr_Disc_All	11.828	14.325	0	70	1284
Start_Appr_All	2.382	2.644	0	39	1025

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 78 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 10 years across the giant discoveries that ultimately reached production stage and close to 14 years when also considering assets not yet producing. These values are well above the timelines presented on the full sample of discoveries. It takes 2.4 years to get from approval to the start of production, considerably more than the 1.5 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. [8] As opposed to the 5 year pre-production period average assumed in multiple studies, this dataset suggests the period is over 10 years for those that have reached production and nearly a quarter 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.

Figure 1 provides a breakdown of assets by region and presents the range of the time from discovery to production observed (or until 2020 for assets not yet producing). It shows that there is large variation between regions, with assets in the Americas on average being developed twice as quickly (6.4 years) as assets in Sub-Saharan Africa (16.8 years).

As shown in Figure 2 the data 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 8.1 years for fields discovered in democracies (polity score above 5 on -10 to 10 scale), it close to double or 15.5 years in autocracies (polity score below -5 on -10 to 10 scale). There is a similar gap for giant discoveries (9.1 year versus 16.9 years).

As shown in Figure 3 larger fields are slower to be developed. The size of the field matters especially for gas fields, where large fields may take triple as long as smaller ones. This could be explained by the need for more complex transport infrastructure to market larger gas fields if the amount of volume found greatly exceeds local demand.

Although the literature estimating the impact of petroleum discoveries tends to focus on giant discoveries, the remainder of my analysis focuses on all discoveries in order to maximize sample size.

[8] 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 1950 - 2010 period, there are 1171 giant discoveries in Horn (2011), while there are 1002 in Rystad's Ucube dataset.

Figure 1: Box plot of asset timelines by region

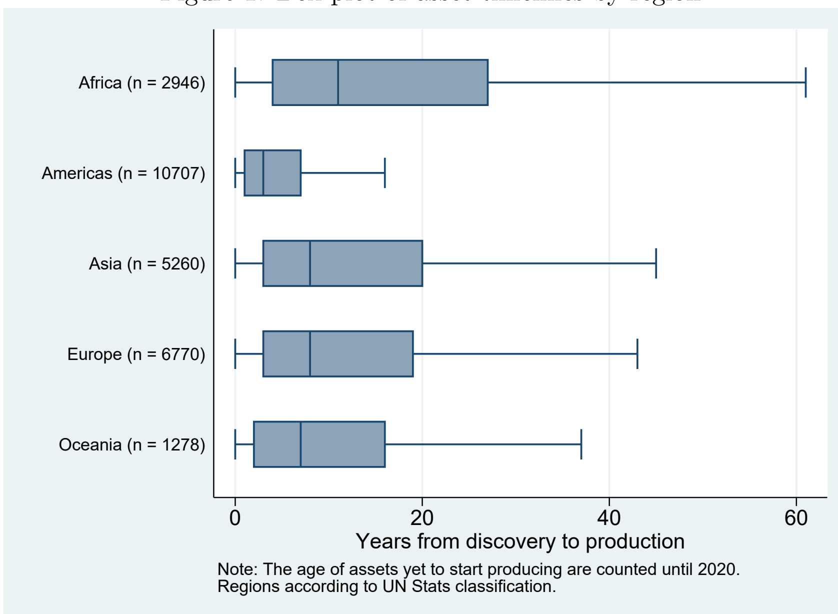


Figure 2: Box plot of asset timelines by regime type and asset size

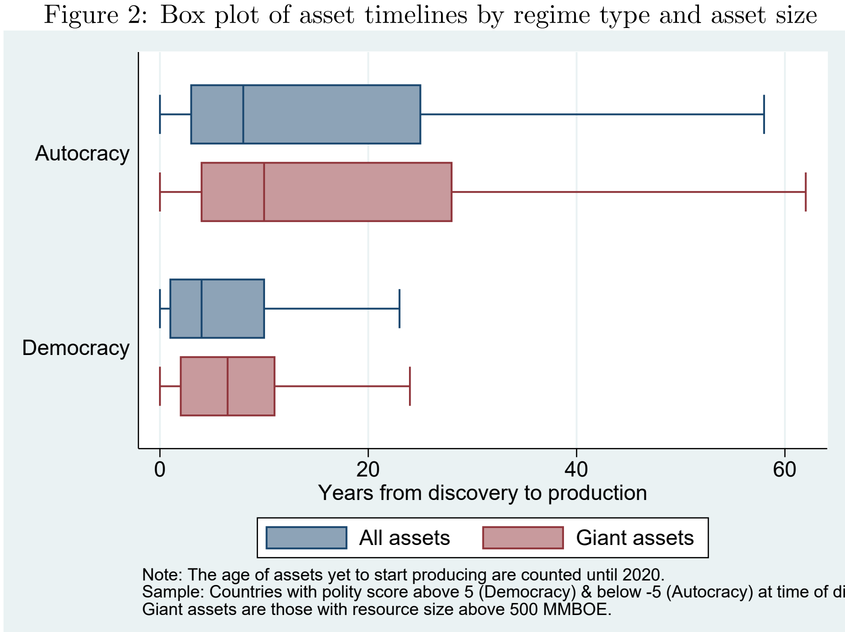
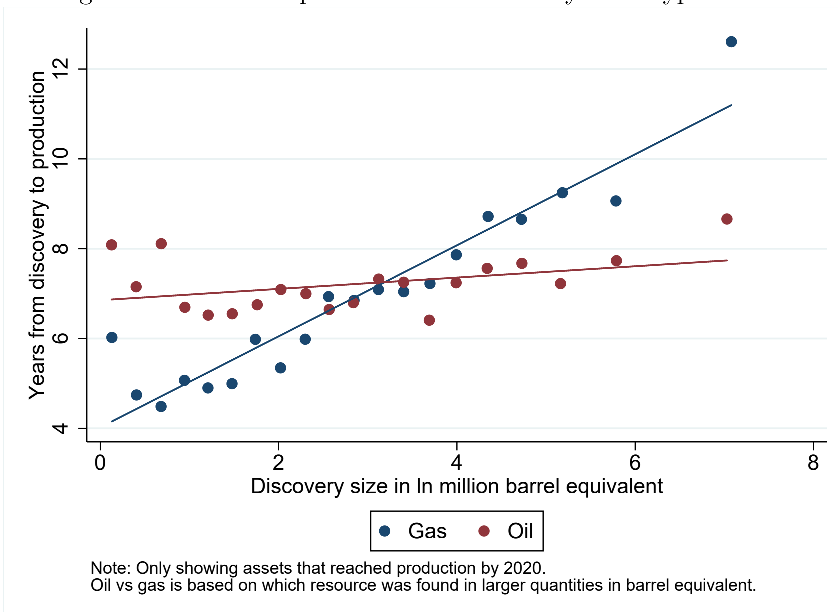


Figure 3: Binscatter plot of asset timeline by asset type and size



For each field, I obtained 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 *(lnAssetSize)*, the log of the water-depth of the field *(lnWaterDepth)*, the share of gas (vs oil) within the find *(GasShare)*, whether the asset is shale or not *(Shaledummy)*.

I supplement the dataset with some country level characteristics. These are the polity scores by Polity IV Project on the level of democracy *(polity2)*, the log of the per-capita level of GDP *(lnGDPpc)* from the Penn World Tables and the log of the number of assets that have already reached production prior to the asset in question (*(lnCountryProdHist)*. This latter variable captures the experience of a country in developing petroleum assets.

I also add the log of the nominal Brent oil price series from the World Bank commodity data tables *(lnOilPrice)*. Adding a *(Year)* numerical variable to regressions enables to capture the impact of technological progress.

For each asset, the time varying variables can be measured at time of discovery, production start or any year in between. I present the descriptive statistics with time varying variables measured at discovery year in Table 3, which is the preferred measure I use in the survival analysis.

Table 3: Description and summary statistics of additional variables

VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max
DiscoveryYear	26,961	1,989	17.62	1,950	2,020
Gas_Share	26,961	0.470	0.396	0	1
ln_Field_Size	26,961	2.890	1.835	0.000394	10.97
Shale_dummy	26,961	0.0707	0.256	0	1
ln_WaterDepth	26,961	1.566	2.209	0	8.423
ln_country_prod_hist_disco	26,961	5.276	1.986	0	8.649
ForeignNOCshare	26,961	0.0396	0.154	0	1
HomeNOCshare	26,961	0.277	0.407	0	1
ln_GDP_pc_disco	23,359	13.85	1.876	6.804	16.84
ln_OilPrice_disco	26,961	2.898	1.219	0.761	4.654
polity2_disco	23,312	0.726	0.358	0	1
NatYear	18,146	1,970	19.28	1,926	2,013
OpNatYear	16,548	1,973	19.54	1,938	2,013

A key explanatory variable in my event study empirical estimations is a country's choice of nationalizing the sector. For this I rely on the National Oil Companies (NOC) Dataset by Mahdavi (2020), which covers nationalization events across 175 sovereign countries over the 1905-2015 period. The key variable of interest from this dataset *(Nat)* is a dummy which denotes the setting up of an upstream nationalized oil company with over 50 percent state ownership. [9] I also add a variable *(OpNat)* from same dataset which denotes when NOC has reached de facto upstream production capacity. This means that is has the ability to physically operate and produce from petroleum fields, rather than just being a participant in projects operated by other companies. As discussed in Mahdavi (2014) these major nationalization events often happen in waves and triggered by a sentiment of resource nationalism.

About 2/3 of the assets within the dataset are located in countries where there was an oil sector nationalization event at some point. Of the 26,961 assets, 18,146 are in a country where the nationalization happened between 1926 to 2013, and 16,548 in a country where the NOC took on an operational role

[9] Includes partially privatized NOCs (e.g.Petrobras) but does not include NOCs only involved in the downstream sector, e.g. refining NOCs.

between 1938-2013.

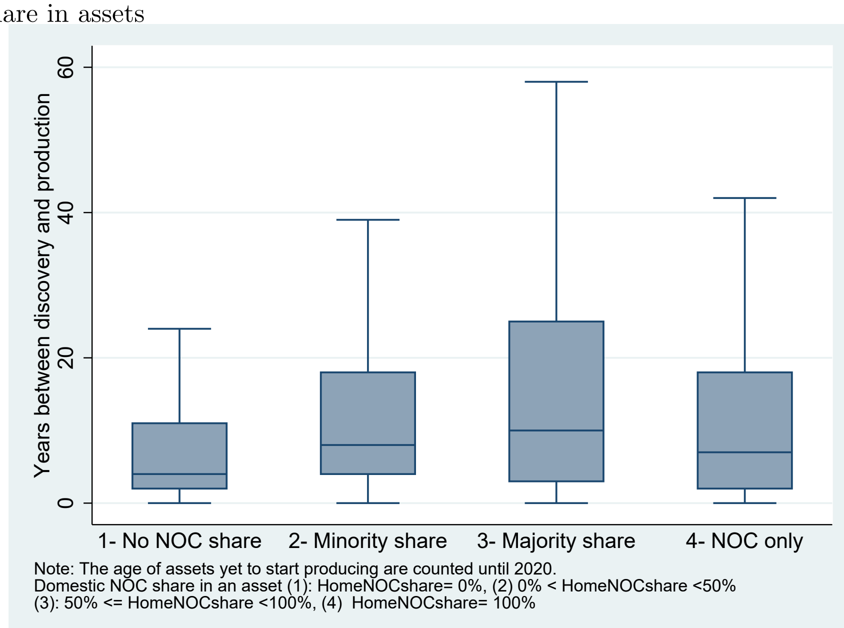
Finally, I analyze a variable which captures the share of state participation through a national oil company (NOC) in each asset. One variable *(Home-NOCshare)* codes for the share of domestic NOCs and one variable *(Foreign-NOCshare)* for the sum of stakes of foreign NOCs in each licence (data from UCube). I measure this variable at the end of nearest 5 year period after discovery. [10]

A few countries have no domestic NOCs (e.g. Australia, US) and some have fully state run sectors (e.g. Iran and Saudi Arabia) but most countries have mixed regimes, where partnerships between domestic NOCs and other companies are the norm. The role of NOCs vary within these partnerships, they may act as the operator or not, they often participate to monitor the project, collect additional revenues or to obtain know-how from the operator Heller and Mihalyi (2019). Figure 4 groups the *(HomeNOCshare)* variable into four categories and depicts project timelines accordingly. Having larger state ownership is correlated with slower project timelines although the association is weak (correlation is 0.13 across two variables).

I have explicitly decided not to include any data on extractions costs in the analysis, despite it being likely an important factor driving extraction decisions and project timelines. The reason for that is twofold. First, there is no public data on asset level costs, only expert estimates of costs, which may reflect various biases. For example, experts may assume that projects that are quick to move ahead are likely cheaper or have various beliefs about companies, especially NOCs, who decide to disclose less financial data. Secondly, the cost estimates will conflate cost drivers that are geological, therefore immutable (e.g. water depth) with those that reflect country factors subject to change (e.g. country risk or qualified staff). Separating out variables into geological, time varying global variables and other country variables can thus better identify the impact of institutions and policy change.

[10] This is a result of a data download limitation from the UCube database. For each asset the data codes the latest owners by default. To bulk retrieve historical ownership I had to narrow the years to take snapshots from. Ownership changes are rare events, hence this simplification is unlikely to alter results substantially.

Figure 4: Binscatter plot of asset timeline by degree of domestic NOC ownership share in assets



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 two estimation techniques: survival analysis and discrete-time event-history analysis (or event study) approach.

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. The approach extends on Khan et al. (2016) who analyzes similar issue in the mining sector.

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.

Figure 5: Kaplan-Meier estimate on the likelihood of an asset not moving to next stage after given number of years.

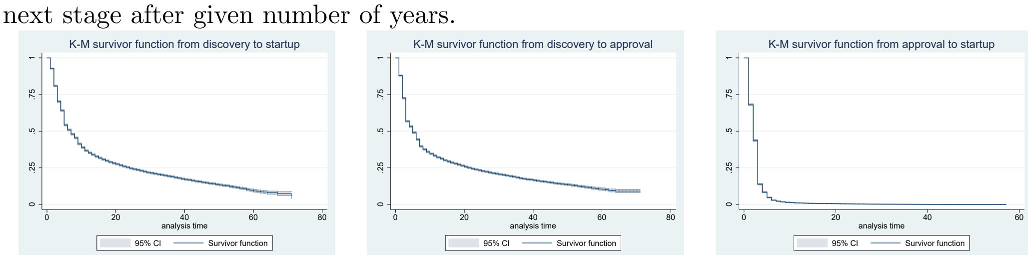
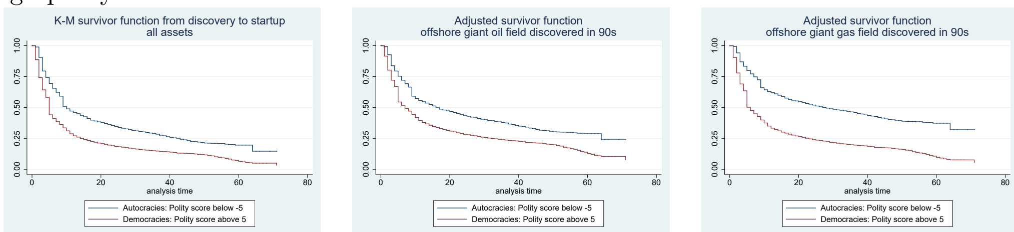


Figure 6: Timeline from discovery to startup for assets in countries with low vs high polity scores - with various controls



I present the K-M estimates for the three different periods in Figure 5. 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 in Figure 6. On the one hand one may speculate that weaker institutional settings have less ability to execute complex petroleum projects. Conversely, it is possible that autocracies are better able to fast track important infrastructure projects by discarding local resistance to it.

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 6, there is a large difference in timeline across institutional scores when comparing only offshore giant oil fields discovered in the 90s. While the odds of such an oil asset remaining underground within 20 year window is 31 percent when located in a country with high institutional score, there is a 45 percent chance when located in a country with low score. That gap increases even further when comparing fields that are mostly gas rather than oil. The odds of offshore giant gas fields remaining undeveloped within 20 years is 55 percent when located in countries with low institutional score at time of discovery or roughly double the odds (27 percent) of those of an asset with similar characteristics but located in one with strong institutional score, see plot 3 of Figure 6. The more marked difference for gas timelines 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), hence may be more dependent on additional country factors. Altogether, the above evidence finds that countries with weaker institutions are slower to execute petroleum projects.

Survival model- regressions

In order to evaluate the significance of individual variables on project timelines, there are a number of specifications to consider. The dependent variable can be three different periods (the full period from discovery to startup, from discovery to approval and from approval to startup), there are a range of explanatory variables to consider. Additionally within survival analysis set-up, there are multiple models: a semi-parametric model, the Cox regression or a parametric model, such as the Gompertz, Weibull, exponential, etc. I present results from the cox model, additional model results from multiple parametric models are shown in appendix.

Key results from analysis

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

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

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 4 for the effects on the full project timeline from discovery to project start. Column (1) shows time varying and country level characteristics, while the preferred specification, Table 4 column (4), uses both country and year fixed effects instead. Time varying control variables (GDP, polity, oil price, production history) are measured at the year of discovery for each asset. The year fixed effects capture the discovery year for each asset. I also show results when including either only year (column 2) or only country (column 3) fixed effects to better understand the drivers of any difference between the other two specifications.

In table 5 I break down to effects across time periods, separating effects on discovery to approval stage and on approval to startup periods. For both I show the results from pooled regression with longer list of controls and the preferred specification with both country and year fixed effects but shorter list of controls.

The results shown highlight the importance of various asset level geological characteristics. Field size matter, where larger fields are quicker to get approved but slower to get from approval to startup (overall sign positive but not all significant). Assets at deeper water depth and which contain more gas (rather than oil) are slower to complete. Shale gas is much quicker to get from approval to start, but slower reaching approval (overall signs are mixed). These results are broadly intuitive and aligned with reporting on the topic in industry press.

The time variant variables have mixed significance. The oil price is not significant (which may be because asset development decisions are based on future oil price expectations and not the ones at discovery). Discoveries found in earlier years were quicker to get approval than newer ones, but are slower to be executed. This latter result would be consistent with increasing petroleum abundance and more scrutiny in deciding which field to develop, but also technological improvement ensuring that fields selected are then developed more quickly.

Country level variables show that richer countries and those with stronger institutions at the time of discovery are quicker to develop their assets. This is in line with intuition that such countries are better able to attract investment and deliver on complex projects. Worth noting that the effect of the polity variable disappears in the project execution phase. Countries with more experience in developing petroleum assets in the past are quicker to develop subsequent finds. But results lose much of their statistical significance once including country fixed effects, suggesting that any new learning over time within country is slow. (Similar patterns of variables losing their significance can be observed when including a country's GDP and polity score together with country fixed effects, results not shown).

The domestic NOC's participation share in assets shows mixed results. Larger domestic NOC share is associated with slower project timelines in the specifications without country and year fixed effect. Results are very similar when adding only time fixed effects, therefore this association is not driven by time periods where both NOCs are more dominant and projects are slower. More interestingly, the domestic NOC variable switches signs after adding country fixed effects, therefore showing quicker timelines on assets with higher state share within the same country. One interpretation is that heavy state-ownership in a country may be correlated with various factors which slow down projects. On the other hand, within that country, it is the projects where the NOC plays a larger role which are more likely to go ahead.

The participation share of foreign NOCs in assets is associated with slower project timelines across all specifications. While the list of NOCs with foreign activities is a subset of those operating at home, it suggests that NOCs have an inherent disadvantage in developing assets globally, which they more than make up for when developing domestic assets.

Table 4: Results from Cox regressions on discovery to startup time

	(1)	(2)	(3)	(4)
VARIABLES	Disc-Start	Disc - Start	Disc - Start	Disc - Start
ln_Field_Size	1.007	0.986***	1.061***	1.059***
	(0.00457)	(0.00394)	(0.00455)	(0.00459)
ln_WaterDepth	0.899***	0.932***	0.873***	0.872***
	(0.00357)	(0.00338)	(0.00398)	(0.00400)
Gas_Share	0.940***	1.047**	0.838***	0.840***
	(0.0188)	(0.0186)	(0.0168)	(0.0169)
Shale_dummy	1.061*	1.290***	0.776***	0.744***
	(0.0336)	(0.0408)	(0.0254)	(0.0262)
ln_OilPrice_disco	0.981			
	(0.0143)			
DiscoveryYear	0.991***			
	(0.00105)			
polity2_disco	1.182***			
	(0.0335)			
ln_GDP_pc_disco	1.077***			
	(0.00832)			
ln_country_prod_hist_disco	1.188***	1.210***	1.077***	1.036*
	(0.00929)	(0.00596)	(0.00805)	(0.0192)
HomeNOCshare	0.893***	0.862***	1.062**	1.073***
	(0.0216)	(0.0169)	(0.0285)	(0.0290)
ForeignNOCshare	0.847***	0.842***	0.826***	0.818***
	(0.0454)	(0.0408)	(0.0451)	(0.0448)
Observations	22,558	26,959	26,959	26,959
Country FE	NO	NO	YES	YES
Year FE	NO	YES	NO	YES

The table shows the impact of various variables on the hazard ratio of an asset discovered getting from discovery to start of production. *, **, *** indicate significance at the 10%, 5%, 1% level, respectively.

Table 5: Results from Cox regressions on partial project timelines

VARIABLES	(1) Disc-Appr	(2) Disc-Appr	(3) Appr-Start	(4) Appr-Start
ln_Field_Size	1.010**	1.057***	0.972***	0.990**
	(0.00456)	(0.00456)	(0.00455)	(0.00449)
ln_WaterDepth	0.910***	0.887***	0.953***	0.924***
	(0.00355)	(0.00398)	(0.00403)	(0.00476)
Gas_Share	0.961**	0.864***	0.897***	0.870***
	(0.0191)	(0.0173)	(0.0183)	(0.0182)
Shale_dummy	0.923**	0.642***	1.883***	1.765***
	(0.0292)	(0.0225)	(0.0630)	(0.0665)
ln_OilPrice_disco	0.977		1.010	
	(0.0141)		(0.0148)	
DiscoveryYear	0.990***		1.004***	
	(0.00104)		(0.00105)	
polity2_disco	1.213***		0.958	
	(0.0340)		(0.0273)	
ln_GDP_pc_disco	1.072***		1.042***	
	(0.00817)		(0.00833)	
ln_country_prod_hist_disco	1.191***	1.022	1.026***	1.034*
	(0.00919)	(0.0188)	(0.00818)	(0.0189)
HomeNOCshare	0.898***	1.063**	0.960*	1.086***
	(0.0214)	(0.0283)	(0.0227)	(0.0304)
ForeignNOCshare	0.854***	0.842***	0.853***	0.813***
	(0.0454)	(0.0454)	(0.0466)	(0.0459)
Observations	22,558	26,959	17,343	20,751
Country FE	NO	YES	NO	YES
Year FE	NO	YES	NO	YES

The table shows the impact of various variables on the hazard ratio of an asset getting to approval stage or from approval to startup.
*, **, *** indicate significance at the 10%, 5%, 1% level, respectively.

Model selection and limitations

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 A.1 and A.2 of the Appendix. Results are very similar to those of the Cox model presented above across the various specifications after taking to account that specification in Table A.1 are results in terms of proportional hazard (meaning a value above 1 is a quicker timeline), while models in Table A.2 are accelerated failure time models (where a value below 1 is a quicker timeline).

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 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 present in the appendix are more flexible in this regard, they do not require such assumption to hold. Without clear guidance from theory on the appropriate function form, one may want to select among parametric functions based on the best fit. This can be done using the Akaike Information Criterion (AIC). As reported in the last row of Table A.1 and A.2 in the Appendix, the AIC test suggests that the best fitting model is the one relying on a logn distribution (Table A.2 - column 2 and column 4 which have the lowest AIC number).

Discussion

The survival analysis 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 and higher GDP at time of discovery are quicker to be developed. I obtain quantitatively similar results using a number of specification of survival models.

A key insight from the analysis is that while larger domestic NOC shares are correlated with slower project timelines, within a country, it is the assets with larger NOC shares that are quicker to get developed. This suggests that countries with higher degree of state of ownership also exhibit additional factors which may slow down project timelines (e.g. lack of human capital, access to technology or regulatory barrier), but that state ownership actually helps in having an asset developed quicker.

One limitation is that the fixed effect model might capture time-invariant omitted variables, such as longstanding "cultures" of corruption that may both increase odds of nationalization but slow down the time between discovery and startup. Such hypothesis would suggest that the domestic NOC effects are actually underestimated.

Another limitation of this methodology is that it only allows to compare assets using time invariant characteristics within the life of the asset. Research by Arezki et al. (2019), Cust and Harding (2019) and Brunnschweiler and Poelhekke (2019) has established that the process of resource discovery is itself dependent on institutional factors. For example Brunnschweiler and Poelhekke (2019) finds that switching to foreign asset ownership results in more exploration and more finds. Therefore the results I present may be biased as certain countries may find more fields or fields with characteristics for which we do not control which in turn could affect project timelines as well. I use an event study approach to mitigate this risk.

5.2 Event study approach

The event study approach allows to estimate changes in likelihood of an asset reaching production in the time periods surrounding a particular event. In this case, I present results from analyzing likelihoods of production start in the years before and after the country nationalizes the sector through setting up a national oil company with a role in domestic production.

In order to implement that I transform the data into 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 linear panel model with a range of explanatory variables. 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 presented above).

I follow a linear panel event study approach using the regression presented in Equation 2.

Although the explanatory variable is binary, I use a linear panel model with many levels of fixed effects (Correia, 2016). [11] I use country-level fixed effects and year fixed effects and robust standard errors clustered at the country-level. [12]

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

where $Start_{i,c,t}$ represents a dummy, which takes the value of 1 if asset i in country c starts production in year t. The main variable of interest is $Post_N OC_{t,c}$, taking a value of 1 if the country c has established an NOC in any given year prior to t. I also include an asset age variable age and age squared age^2 variable to capture the fact that the petroleum field has a decreasing likelihood of opening as years progress. A series of control variables are denoted Z. The list of asset level controls are the same as in section above: $Shale\_dummy$, ln_Field_Size, Gas_Share, $ln\_WaterDepth$ and $ln\_country\_prod\_hist$). I do not add a control variable on NOC participation share given the main sock variable of interest is closely related. I add country and year fixed effects (α_c and δ_t) to all specifications, which capture country characteristics (such as resource endowments or human capital) and time trends (including the changes in oil price and effects of technological progress).

I use this approach to test for the significance of countries nationalizing the industry as an explanatory variable. I analyze observations around this event in a way which includes some assets that spent all the time prior, only post the event but also some that have spent some years both prior and after the shock.

[11] This follows (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.

[12] I use robust standard errors clustered at the country-level for experimental design reasons: the level of treatment (nationalization) is at the country-year level, while observations are at asset-year level (Abadie et al., 2017). In appendix I show asset level clustering also, where results are stronger.

A set of dummy variables capture all possible lags and leads to the event. [13]

Key results from analysis

The Table 6 shows the results of the main regression. It measures the impact of various variables on the likelihood of an oil asset reaching start up stage in any given year. The *age* variable and *age-squared* variable capture the fact that assets have a decreasing likelihood of opening as years progress albeit at diminishing rates. (While the likelihood of opening drops sharply in the initial years it later decelerates.) Additional asset level controls used in earlier regressions are also included and show similar results although not always significant. Larger fields and shale assets are quicker, deeper fields and those with higher proportion of gas are slower.

The new insight comes from the inclusion of a dummy variable on whether the country has nationalized its industry through setting up a national oil company at any point in time. Four fifth of all observations are located in countries which eventually set up an NOC. [14] In Table 6 I show that assets are 1.4 percentage point more likely to open up after a national oil company was set up *(Post-Nat)* or after a national oil company with an operational role*(Post-OPNat)* was set up.

Having included year fixed effects capture spurious correlations in case years with more oil sector nationalization events globally coincided with periods when more project were about to start up. The country fixed effect capture spurious correlations where geography may be correlated with both nationalized oil sectors and petroleum assets which may be easier to develop. The *ln_country_prod_hist* variable captures spurious correlation where country production trends may drive both increased country-level knowledge on how to develop assets and desire to nationalize the industry. The robust standard errors clustered at the country level ensure that the results are not overly driven by very few countries with many assets.

The results presented here indicate that there is an increase in likelihood of assets turning to production in the years following an NOC being set up. While the typical asset has 6.4 percent likelihood of starting up in any given year,

[13] I follow (Clarke and Schythe, 2020) in implementing the event study.

[14] The US and Australia are the two petroleum producers with no NOCs with the largest number of assets alongside some other countries with fewer assets.

the odds increase by about 1.4 percentage point after NOC is set up (Table 6, column 1). This is equivalent to a 20 percent increase in likelihood of project start up in any given year. Results are similar when measuring what happens after an NOC takes on an operational role (Table 6, column 2). [15].

Next, I analyze the effects measured above over time. The Figure 7 depicts how the chances of an asset starting up changes in the 15 years prior to and up to 30 years after an NOC is being set up.[16] The reference year used, where the coefficient is set to zero, is the year prior to opening up: the results shown for all other years are in comparison to this one.

While there are no strong trends in the years prior to establishing the NOC, within 15 year of setting up the NOC there is a positive and significant increase in asset start up likelihood (bars show 95 percent confidence intervals). These effects are similar but somewhat less pronounced when looking at national oil companies taking on an operational role (Figure 7, right panel).

In the appendix, I also show a regression extending Table A.4 with dummies coding for 5 year time periods prior and after nationalization events similar to those shown Figure 7. Using the 5 year prior to nationalization as the base period, it also shows a significant jump in asset startup likelihood in the years 5+ after nationalization events.

[15] The two variables are not jointly significant when included in same regression. This is likely a result of strong overlap between two variables, with two events either coinciding or following each other with small timelag

[16] 40 percent of all observations (including those where no NAT event happened) fall within this time window. I show a histogram in Appendix Figure **??**.

Figure 7: Asset starting up around nationalization / operational nationalization events

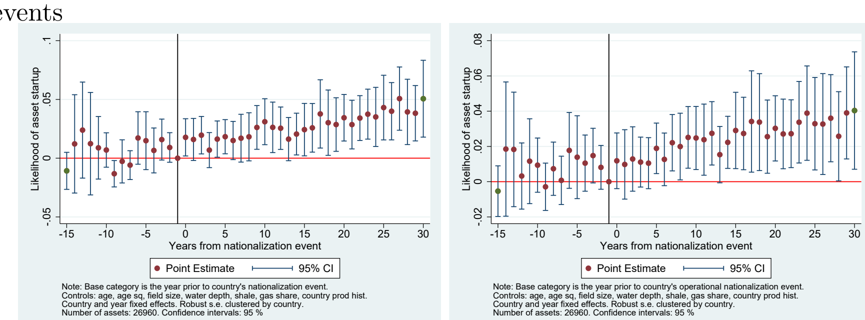


Table 6: Regression with discrete-time event-history

	(1)	(2)
VARIABLES	Start	Start
age	-0.00413***	-0.00413***
	(0.000854)	(0.000854)
age_sq	5.76e-05***	5.75e-05***
	(1.08e-05)	(1.08e-05)
ln_Field_Size	0.00562**	0.00560**
	(0.00232)	(0.00232)
ln_WaterDepth	-0.00931***	-0.00928***
	(0.00109)	(0.00110)
Shale_dummy	0.0118	0.0120
	(0.0249)	(0.0248)
Gas_Share	-0.0124	-0.0125
	(0.0116)	(0.0116)
ln_country_prod_hist	0.0390***	0.0392***
	(0.00546)	(0.00545)
post_nat	0.0145***	
	(0.00516)	
post_opnat		0.0143***
		(0.00537)
Constant	-0.104***	-0.104***
	(0.0273)	(0.0272)
Observations	317,194	317,194
R-squared	0.057	0.057
Country FE	YES	YES
Year FE	YES	YES

The table shows the impact of various variables on an production start dummy where each observation represents a year of the asset's life from discovery to production start. Standard errors are robust and clustered at the country level. *, **, *** indicate significance at the 10%, 5%, 1% level, respectively.

Robustness

I carry out a number of alternative specifications to ensure the robustness of the key results, with results shown in the Appendix.

First, I replicate the main specification by dropping any countries with over 100 petroleum assets one at a time. I check how the central estimate of the *(Post-Nat)* dummy changes in regressions where one country is left out. As shown in Figure A.3, the results barely change irrespective of which country is dropped.

Second, I remove the restriction on the time window observed prior and after the setting up of national oil companies and one with operational role Figure A.4 . Given the sharp drop in observations when measuring larger time lags, this leads to less robust estimates but the overall pattern is still visible.

Third, I measure the effects of setting up a national oil company on asset approval rather than on asset startup. As shown in Figure A.5, I find similar impacts as earlier shown .

Fourth, I calculate results using robust standard errors clustered at the asset-level. As shown in Figure A.6 results remain unchanged in magnitude and confidence intervals become smaller. (This suggests that the variance is correlated at the country-level and not at the asset-level).

Fifth, I have repeated the analysis on three samples that differ somewhat to the original one, as shown in Figure A.7. This includes dropping all shale assets, which tend to have longer approval timelines but quicker execution timelines (left plot). Alternatively, I add assets which represent subsequent phases of existing fields back to the sample (center plot), which had been dropped as considered less pertinent for this analysis. Finally I exclude all assets that have spent at least 40 years without production (right plot). These may be considered as outliers in terms of the slowness of their development. The results remain largely unaffected by either of these sample changes.

One important caveat to the main results is that the coefficients on project start prior to nationalization are not all null. They are often negative across specifications, especially when looking back multiple decades prior to nationalization. This would suggest that there may be an increased likelihood of nationalization after longer periods of under-performance in asset start-up. This pattern would broadly fit the observation by Mahdavi (2014) that nationalization events are aimed at maximizing sector revenues. **Discussion**

I have shown that setting up a national oil company within a country is followed by a 1.4 percentage point (or about 20 percent) increase in likelihood of assets turning to production in any given year. The geological variables, the country, year fixed effects and production history variable control for the effects of potential confounders such as trends in oil price, technology, differences in endowments and country experience. The results are robust to alternative specifications and to dropping groups of observations. This approach still has some limitations, as it can't discern any hidden third factor that both contributes to countries setting up national oil companies and quicker project timelines. Further investigation will be required to firmly prove causality.

There are a number of potential hypothetical channels which may explain the observed association. For example, the NOC's involvement in an asset may be help overcoming bureaucratic setbacks, more able to garner support for developing the project or they may be more willing to take larger financial risks (as suggested by Marcel (2019)).

Another hypothesis consistent with the results is that a government which wants a priori to increase depletion rates can only effectively do so with NOC control. The government cannot reasonably force foreign companies to produce quicker or startup assets faster if the companies do not believe it wise to do so. Governments with NOCs may deliberately want to speed up the extraction process even if it comes at a future cost – e.g., rapid depressurization of wells. This is consistent with results by Mahdavi (2020) who suggests leaders who have constrained time horizons are more likely to opt for operational control in the hands of NOCs.

Part of the association may be indirect, driven by a third factor, such as a greater desire by the government to achieve energy independence. This could drive both larger likelihood of nationalization and also accelerated production.

Nevertheless, the observed association between higher state control and quicker project timeline is telling irrespective of whether it is caused by the national oil company directly or an underlying third factor, such as resource nationalism. Both point to the ability of governments to influence the speed at which oil assets are developed.

6 Conclusion

I presented a detailed analysis on the factors that influence the speed at which petroleum assets are being developed globally. I have shown that on top of geological factors, which are beyond the country's own control, a country's institutional and developmental characteristics also matter matter for the speed at which petroleum assets are being developed. My findings imply that earlier research relying on lagged impacts of giant petroleum discoveries produced biased estimates of subsequent oil production. They underestimated the post-production impacts from countries that are slower to extract their resources. Alternatively, some of the impacts (e.g. increase in conflict or borrowing) these articles had captured may in fact have happened prior to the start of production.

While state ownership of resource sector is associated with slower project timelines overall, this correlation is misleading. Countries with high degree of state of ownership are likelier to also exhibit other factors (geological or institutional) which may slow down project timelines. Within a country, it is the assets with larger state share that are the quicker to be developed. In an event study I have also shown that the likelihood of assets getting developed increases after setting up a national oil company. These results suggest state ownership may in fact help rather than hinder asset development timelines. Further research is need to firmly establish causality.

In order to mitigate climate change, a large share of already discovered petroleum resources need to remain underground. But economists have also warned of the risk of a green paradox, where oil companies decide to accelerate fuel extraction in anticipation of a shift to renewable energy. Further work building on my research could help evaluate how the race to extract the last drop of oil may unfold.

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.

Anderson, S. T., R. Kellogg, and S. W. Salant (2018). Hotelling under Pressure. *Journal of Political Economy 126*(3), 984–1026.

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. and S. Poelhekke (2019). Pushing one's luck: petroleum ownership and discoveries.

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.

Clarke, D. and K. Schythe (2020). Implementing the panel event study.

Correia, S. (2016). Linear models with high-dimensional fixed effects: An efficient and feasible estimator. Technical report. Working Paper.

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.*

Dale, S. and B. Fattouh (2018). Peak oil demand and long-run oil prices. *Energy Insight 25.*

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.

Heller, P. and D. Mihalyi (2019). Massive and misunderstood: Data-driven insights into national oil companies.

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.

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.

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

Mahdavi, P. (2014). Why do leaders nationalize the oil industry? the politics of resource expropriation. *Energy Policy 75*, 228–243.

Mahdavi, P. (2020). *Power Grab: Political Survival Through Extractive Resource Nationalization*. Cambridge University Press.

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*.

Manley, D. and P. R. Heller (2021). Risky bet:national oil companies in the energy transition.

Marcel, V. (2019). National oil companies of the future. In *Annales des Mines-Responsabilite et environnement*, Number 3, pp. 133–136. FFE.

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.

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.

Merrill, R. K. and A. W. Orlando (2020). Oil at risk: Political violence and accelerated carbon extraction in the middle east and north africa. *Energy Economics 92*, 104935.

Mihalyi, D. and T. Scurfield (2021). How africa's prospective petroleum producers fell victim to the presource curse. *The Extractive Industries and Society 8*(1), 220–232.

Perez-Sebastian, F., O. Raveh, and F. van der Ploeg (2021). Oil discoveries and protectionism: Role of news effects. *Journal of Environmental Economics and Management*, 102425.

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.

Sinn, H.-W. (2008). Public policies against global warming: a supply side approach. *International tax and public finance 15*(4), 360–394.

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. Vézina (2018). Resource discoveries, fdi bonanzas, and local multipliers: Evidence from mozambique. *Review of Economics and Statistics*, 1–36.

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

Van der Ploeg, F. and C. Withagen (2012). Is there really a green paradox? *Journal of Environmental Economics and Management 64*(3), 342–363.

A Appendix

Table A.1: Results from proportional hazard parametric regressions w AIC test results

VARIABLES	(1) exp	(2) gom	(3) wei	(4) exp	(5) gom	(6) wei
ln_Field_Size	1.020***	1.010**	1.018***	1.076***	1.063***	1.077***
	(0.00454)	(0.00456)	(0.00455)	(0.00459)	(0.00459)	(0.00460)
ln_WaterDepth	0.886***	0.896***	0.889***	0.865***	0.872***	0.864***
	(0.00347)	(0.00355)	(0.00350)	(0.00387)	(0.00396)	(0.00387)
Gas_Share	0.967*	0.948***	0.963*	0.843***	0.846***	0.843***
	(0.0192)	(0.0189)	(0.0191)	(0.0169)	(0.0170)	(0.0169)
Shale_dummy	1.023	1.067**	1.018	0.676***	0.724***	0.675***
	(0.0320)	(0.0337)	(0.0319)	(0.0235)	(0.0253)	(0.0235)
ln_OilPrice_disco	0.969**	0.974*	0.969**			
	(0.0140)	(0.0141)	(0.0140)			
DiscoveryYear	0.997**	0.991***	0.996***			
	(0.00105)	(0.00105)	(0.00105)			
ln_country_prod_hist_disco	1.233***	1.201***	1.224***	1.045**	1.031*	1.045**
	(0.00961)	(0.00940)	(0.00957)	(0.0193)	(0.0191)	(0.0193)
HomeNOCshare	0.895***	0.897***	0.897***	1.088***	1.083***	1.088***
	(0.0218)	(0.0217)	(0.0219)	(0.0292)	(0.0292)	(0.0292)
ForeignNOCshare	0.884**	0.865***	0.883**	0.819***	0.826***	0.819***
	(0.0470)	(0.0463)	(0.0470)	(0.0449)	(0.0452)	(0.0449)
polity2_disco	1.280***	1.212***	1.266***			
	(0.0360)	(0.0342)	(0.0357)			
ln_GDP_pc_disco	1.105***	1.084***	1.098***			
	(0.00859)	(0.00841)	(0.00853)			
Observations	22,558	22,558	22,558	26,959	26,959	26,959
Country FE	NO	NO	NO	YES	YES	YES
Year FE	NO	NO	NO	YES	YES	YES
AIC	67326	65368	67186	75837	74779	75837

The table shows the impact of various variables on the hazard ratio of an asset discovered reaching the start of production.
*, **, *** indicate significance at the 10%, 5%, 1% level, respectively.

Table A.2: Results from accelerated failure time model parametric regressions (estimates are reversed with values above 1 being slower) and AIC test results

	(1)	(2)	(3)	(4)
VARIABLES	logl	logn	logl	logn
ln_Field_Size	0.998	0.998	0.937***	0.937***
	(0.00514)	(0.00502)	(0.00425)	(0.00417)
ln_WaterDepth	1.158***	1.153***	1.185***	1.179***
	(0.00495)	(0.00483)	(0.00556)	(0.00547)
Gas_Share	1.175***	1.158***	1.323***	1.308***
	(0.0258)	(0.0248)	(0.0267)	(0.0260)
Shale_dummy	1.217***	1.176***	1.808***	1.689***
	(0.0430)	(0.0422)	(0.0657)	(0.0609)
ln_OilPrice_disco	1.043***	1.040***		
	(0.0162)	(0.0160)		
DiscoveryYear	1.006***	1.005***		
	(0.00113)	(0.00111)		
ln_country_prod_hist_disco	0.783***	0.795***	0.937***	0.938***
	(0.00659)	(0.00637)	(0.0173)	(0.0168)
HomeNOCshare	1.170***	1.186***	0.859***	0.877***
	(0.0293)	(0.0286)	(0.0229)	(0.0227)
ForeignNOCshare	1.339***	1.296***	1.298***	1.258***
	(0.0768)	(0.0716)	(0.0682)	(0.0641)
polity2_disco	0.790***	0.786***		
	(0.0243)	(0.0231)		
ln_GDP_pc_disco	0.946***	0.947***		
	(0.00775)	(0.00742)		
Observations	22,558	22,558	26,959	26,959
Country FE	NO	NO	YES	YES
Year FE	NO	NO	YES	YES
AIC	63521	63057	71483	71118

The table shows the impact of various variables on the accelerated failure time of an asset discovered reaching the start of production.
*, **, *** indicate significance at the 10%, 5%, 1% level, respectively.

Table A.3: Regression underlying event study on nationalization

VARIABLES	(1) Start	(2) Approv
age	-0.00411***	-0.00708***
	(0.000850)	(0.00151)
age_sq	5.65e-05***	0.000106***
	(1.05e-05)	(2.21e-05)
ln_Field_Size	0.00562**	0.00599**
	(0.00232)	(0.00272)
ln_WaterDepth	-0.00924***	-0.00935***
	(0.00106)	(0.00135)
Shale_dummy	0.0156	-0.00825
	(0.0221)	(0.0292)
Gas_Share	-0.0127	-0.0113
	(0.0112)	(0.0137)
ln_country_appr_hist		0.0233***
		(0.00404)
ln_country_prod_hist	0.0381***	
	(0.00508)	
pre_nat_15	-0.0207***	-0.0279***
	(0.00618)	(0.00999)
pre_nat_10_15	0.00281	-0.0179**
	(0.0110)	(0.00804)
pre_nat_5_10	-0.00701	-0.00464
	(0.00451)	(0.00582)
post_nat_0_5	0.00547	0.00814*
	(0.00387)	(0.00463)
post_nat_5_10	0.0116**	0.0139**
	(0.00493)	(0.00540)
post_nat_10_15	0.0124**	0.0186**
	(0.00535)	(0.00740)
post_nat_15_20	0.0215***	0.0301***
	(0.00764)	(0.0103)
post_nat_20_25	0.0258***	0.0400***
	(0.00754)	(0.0110)
post_nat_25_30	0.0317***	0.0462***
	(0.00966)	(0.0143)
post_nat_30	0.0412***	0.0647***
	(0.0135)	(0.0218)
Constant	-0.109***	-0.0117
	(0.0261)	(0.0187)
Observations	317,194	285,100
R-squared	0.058	0.080
Country FE	YES	YES
Year FE	YES	YES

The table shows the impact of various variables on an production start dummy and approval dummy where each observation represents a year of the asset's life from discovery to production start. Standard errors are robust and clustered at the country level.
*, **, *** indicate significance at the 10%, 5%, 1% level, respectively.

Table A.4: Regression underlying event study on operational nationalization

VARIABLES	(1) Start	(2) Approv
age	-0.00409***	-0.00704***
	(0.000839)	(0.00149)
age_sq	5.66e-05***	0.000107***
	(1.04e-05)	(2.19e-05)
ln_Field_Size	0.00559**	0.00592**
	(0.00232)	(0.00272)
ln_WaterDepth	-0.00916***	-0.00921***
	(0.00110)	(0.00141)
Shale_dummy	0.0144	-0.0104
	(0.0231)	(0.0314)
Gas_Share	-0.0127	-0.0113
	(0.0113)	(0.0139)
ln_country_prod_hist	0.0385***	
	(0.00545)	
ln_country_appr_hist		0.0235***
		(0.00413)
pre_opnat_15	-0.0141***	-0.0202**
	(0.00529)	(0.00985)
pre_opnat_10_15	0.00323	-0.00693
	(0.00693)	(0.00710)
pre_opnat_5_10	-0.00112	0.00193
	(0.00424)	(0.00559)
post_opnat_0_5	0.00389	0.00861
	(0.00427)	(0.00530)
post_opnat_5_10	0.0120***	0.0162***
	(0.00442)	(0.00562)
post_opnat_10_15	0.0145**	0.0203**
	(0.00563)	(0.00826)
post_opnat_15_20	0.0212***	0.0263**
	(0.00810)	(0.0114)
post_opnat_20_25	0.0230***	0.0339***
	(0.00842)	(0.0123)
post_opnat_25_30	0.0248**	0.0392**
	(0.0115)	(0.0156)
post_opnat_30	0.0313**	0.0468**
	(0.0153)	(0.0229)
Constant	-0.105***	-0.00200
	(0.0285)	(0.0212)
Observations	317,194	285,100
R-squared	0.058	0.079
Country FE	YES	YES
Year FE	YES	YES

The table shows the impact of various variables on an production start dummy and approval dummy where each observation represents a year of the asset's life from discovery to production start.
Standard errors are robust and clustered at the country level.
*, **, *** indicate significance at the 10%, 5%, 1% level, respectively.

Figure A.1: Histogram of nationalization / operational nationalization events used in regressions (weights proportionate to number of assets)

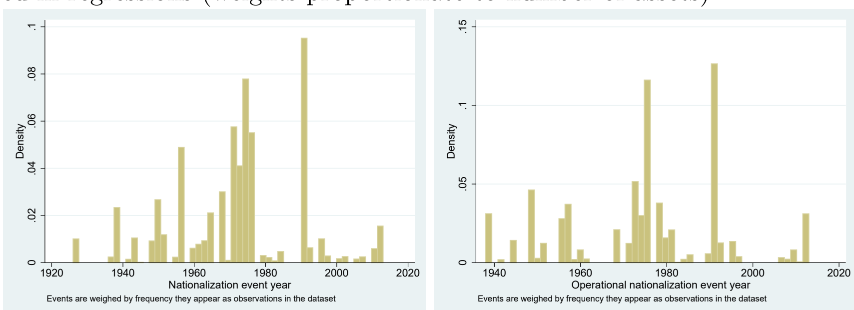


Figure A.2: Histogram on the number of observations around nationalization / operational nationalization events

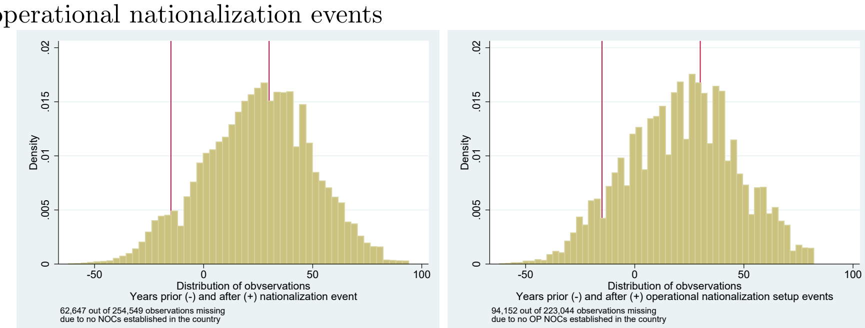


Figure A.3: Leave one country out regression results. The coefficient of the post-nationalization variable when dropping selected country.

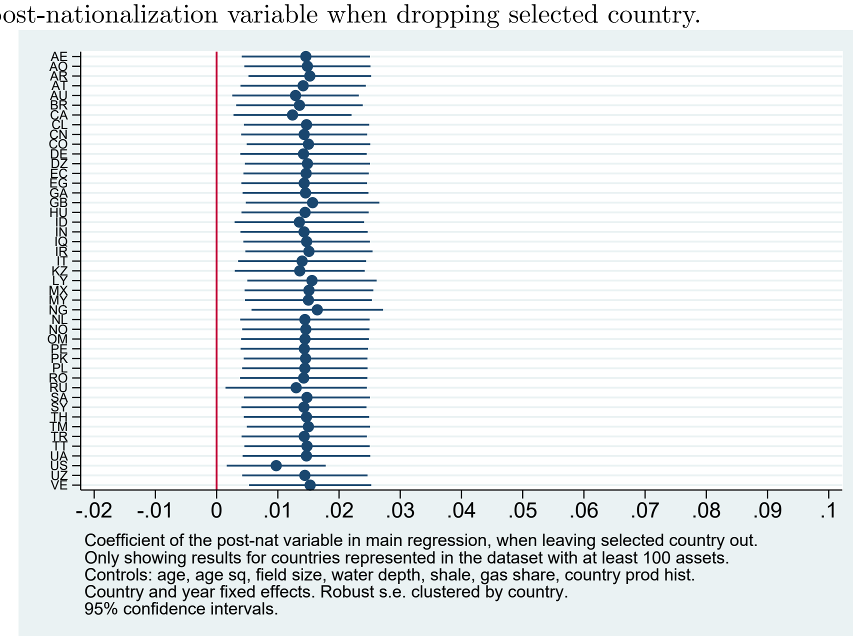


Figure A.4: Asset starting up around nationalization / operational nationalization events - all years

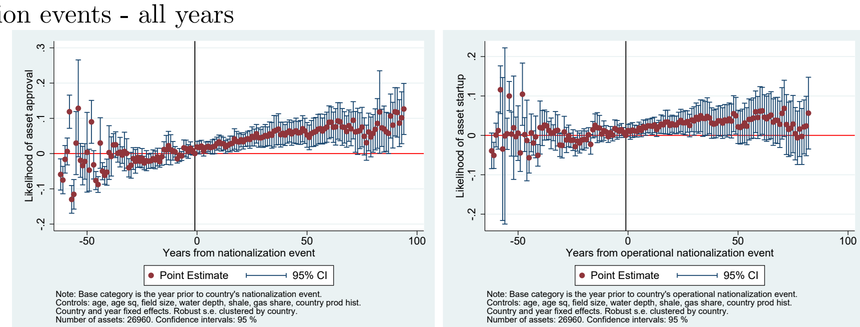


Figure A.5: Asset getting approved around nationalization / operational nationalization events

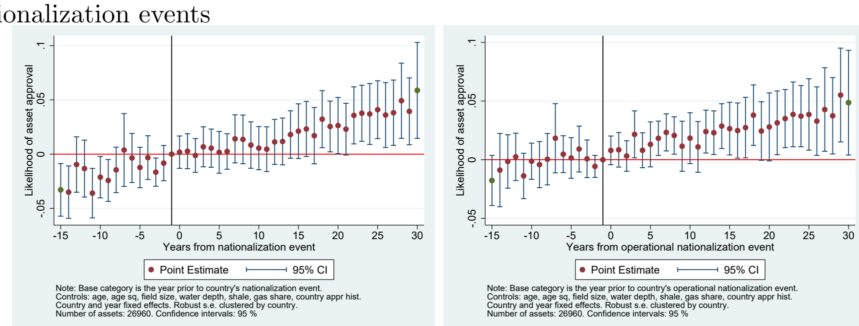


Figure A.6: Asset starting up around nationalization / operational nationalization events - errors clustered at asset level

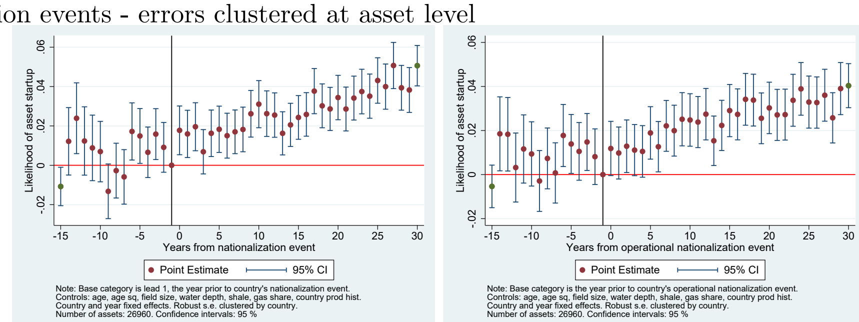


Figure A.7: Asset starting up around nationalization events - Modified sample