CENTRE FOR APPLIED MACROECONOMICS AND COMMODITY PRICES (CAMP)

CAMP Working Paper Series
No 3/2022

Local economic development and oil discoveries

Jonas Hveding Hamang

Local economic development and oil discoveries

Jonas Hveding Hamang*

September 21, 2022

Abstract

In this paper I use data on the location of all historic petroleum discoveries onshore to establish a new stylized fact: Economically developed areas are significantly more likely (about five percentage points) to contain an oil discovery, compared to undeveloped areas. This result is robust to accounting for reverse causality, confounding geology and observed or unobserved country characteristics. By implication, there are large underexplored oil and gas deposits in currently undeveloped areas. I calculate these deposits to be approximately 600 billion barrels of oil — amounting to about 50% of the globe's currently known onshore endowment — and to be mainly located outside of Europe and North America. Exploring alternative mechanisms, I find that infrastructure access may explain the documented discovery differential.

1 Introduction

It is well established that natural resource endowments are important determinants of economic development (e.g., Sachs and Warner, 1995; Nordhaus et al., 2006; Henderson et al., 2018). However, discovered endowments are not necessarily determined by natural geography alone (Cust and Harding, 2020; Arezki et al., 2019), even at the local level. For example, once a village or city is established, its combination of access to labor, capital and infrastructure may increase the value of exploring for natural resources, potentially resulting in new discoveries. Thus, while locations' natural geography *per se* is given (i.e. by earth's physical characteristics), *knowledge* of their natural resource endowments might change as a function of local economic activity. The idea that second nature characteristics—e.g., the location of settlements, capital and infrastructure—might feed back to our knowledge of the existence and location of natural resources has so far not been articulated and empirically analyzed by others.

Empirically identifying second nature characteristics' impact on the discovery of natural resource endowment is intrinsically hard due to reverse causality and potential confounders. Ideally one would like to observe a natural resource that: is relatively newly discovered (i.e. discovered after the establishment of modern settlement patterns, to avoid reverse causality); requires at least some exploration effort to be discovered (i.e.

*Corresponding author: jonas.h.hamang@bi.no. Affiliation: BI Norwegian Bussiness School. I would like to thank Jørgen Juel Andersen, Torfinn Harding, Ragnar Torvik, David Weil, Thomas McKay and Morten Grindaker, your feedback has been vital. I would also thank participants at NOITS (2022), BI workshops 2018 and 2021, and to all of my colleagues at BI for comments and support.

is difficult to observe); is subject to high demand (i.e., has sufficient economic value); is unrelated to other natural characteristics (to avoid confounders); is spatially dispersed across many locations globally (to avoid spurious correlations); and for which there exist complete detailed records for all discovered locations. While this ideal case may not exist, I argue that oil and gas (henceforth just oil) discoveries comes close. Oil was first commercially discovered in 1835, is located in the subsurface and requires extensive exploration to be observed and verified, is clearly of high economic value, is arguably unrelated to surface characteristics, is globally dispersed, and complete data of all historic discoveries has recently become available.

In this paper, I use a unique high resolution geo-referenced oil data (0.25 degree) set from Rystad Energy (UCube, 2020) to document that economically developed areas, defined by detectable night light, have been 5.2 percentage point (pp) more likely relative to undeveloped areas to ever contain an oil discovery. This is a robust finding that holds also when I account for country-level differences (like institutional quality) and potentially confounding geology. However, as the location of economically developed areas can be influenced by the location of oil discoveries it is unclear how to interpret this novel empirical fact. I address this potential reverse causality issue by instrumenting economically developed areas on the distance to their closest historic settlement that predates the first historic oil discovery by at least three decades, thereby obtaining variation in locations of economically developed areas that are unaffected by the location of future oil discoveries. In an IV-specification that includes country, biome and geology fixed effects (FEs) and a large set of surface controls, I estimate the higher likelihood to be 5.8 pp, just above my baseline estimate of 5.2 pp. With a resulting conditional discovery rate of 2.3% in undeveloped areas, this means that economically developed is estimated to be more than 3.5 times likely to contain a discovery. Given the historic accounts of the average size of a discovery, my IV-estimate suggests that 613 billion barrels of oil (Gbbl) of *additional* undiscovered oil exist in currently undeveloped areas compared to already developed areas. This amounts to approximately 50% of the globe's current endowment of known onshore oil. The potential impact of this insight are highly economically and geo-politically relevant—not only due to the location of future energy supply, but because it suggests vastly different costs of phasing out oil across regions and countries.[1] By subsampling regions, I show that Africa, South America, Asia and Oceania are all regions with areas that are underexplored relative to their subsurface potential in economically undeveloped areas. In contrast, I find no evidence of excess oil potential in Europe and North America. The latter two regions are those with the longest history of oil discoveries, and they also have a relatively small fraction of undeveloped areas, which may explain this finding.

In line with my main result, I show that pre-production expenditure on exploration and capital investment is higher for discoveries in developed areas. Higher expenditures indicate that making a discovery in a developed area is more valuable for the petroleum companies, as profit maximizing behaviour dictates the marginal dollar to flow wherever it provides the highest expected return. In contrast, I find no statistically significant differences in discovery size and well construction expenditure, both of which are largely determined by geology. Hence, this is consistent with the relatively higher level of expenditures in the economically developed areas *not* being

[1] Examples of papers that investigate the impact of natural resource endowments: Arezki et al., 2017 (oil discoveries on the current account and savings); Asheim et al., 2019 (climate agreements); Nordvik, 2019 (oil on coups); Berman et al., 2017 (mineral prices on conflict) ; Andersen et al., 2022a (supply-side climate policy); Andersen et al., 2022b (post oil transition on democracy)

driven by differences in geological conditions.

While geological differences cannot explain why companies spend more on exploration in developed areas, another potential explanation is more favorable geographic surface characteristics. Clearly the location of economic development in general is partly determined by surface characteristics and, hence, the same might be true for oil activity. Including 18 different geographic surface characteristics (e.g. ruggedness, elevation, coastal, etc.) in my IV-regressions (all taken from Henderson et al., 2018), I find that they have high explanatory power for the location of current economic development in the first-stage, however, they do not affect my second-stage estimate of interest, neither economically nor statistically. In addition, the oil industry itself has identified sets of locations that have more or less favorable climate conditions for oil activity (from benign to different types of hazards and wear on facilities and operations). Adding these to the specification, my estimate of interest remains almost unchanged. The latter potentially reflects that most onshore areas are categorized as benign and, hence, that my main result is likely to be driven by variation *within* the set of benign locations. In sum, differences in natural surface geography do not seem to explain why the discovery rate is higher in economically developed areas, and I therefore proceed by exploring three alternative explanations: Economic intensity (i.e., the degree of economic development); geographic centrality (i.e., distance to economic centres); and structural connectedness (i.e., access to infrastructure, in particular, railroad).[2]

More economically developed areas may be both more productive and more energy intensive (i.e., have higher local demand for oil and gas), two plausible explanations for why a discovery may be associated with a higher expected return for the oil companies and, hence, higher expenditure and rate of discoveries. However, testing this hypothesis by comparing areas with different predicted night light *intensity*, I find no indication that areas with a higher intensity of night light have experienced a higher discovery rate. This indicates that it is the location of *any* level of economic development (extensive margin), rather than the intensity, or degree, of economic development (intensive margin) that matters for the discovery differential across developed and undeveloped areas.

Next, I investigate whether the distance to economic centres (geographic centrality) can explain the higher rate of discoveries in economically developed areas. If developed areas tend to be located closer to economic centres, this can reduce transportation costs of both output and input factors, thereby potentially explaining why exploration companies spend more per discovery. I investigate this by comparing the difference in discovery rates between economically developed and undeveloped areas that are located near a city (within 100 km). The idea is that if I restrict my sample to areas that are close to a city, we would not expect to see a noticeable difference in the discovery rates, if geographic centrality indeed is a key explanation. However, I still find a significantly higher discovery rate in economically developed areas in this restricted sample. This indicates that centrality is not a main explanation for my results.

The final potential explanation that I explore is access to infrastructure. As with centrality, access to infrastructure can reduce marginal transportation costs of output and input factors, as well as reduce the need

[2] All three explanations are emphasized as features of economic development (Feldman, 1999; Krugman, 1991b; Redding and Turner, 2015)

for pre-production investments in infrastructure, thereby making areas close to infrastructure more attractive to explore. Proxying access to infrastructure by closeness to a railroad (within a radius of 100 km) and limiting the sample to these areas, I find only a very small and statistically insignificant discovery rate differential between economically developed and undeveloped areas. Hence proximity to infrastructure appears to be a potential explanation for the location of oil discoveries across both developed and undeveloped areas, and is therefore also a potential explanation why developed areas in general—characterized by more infrastructure—have a higher likelihood of oil discoveries.

This paper contributes to the New Economic Geography literature by its focus on the two-way relation between economic variables and predetermined geography. Seminal papers like Krugman (1991b) and Krugman (1993) argue that economics of scale, transportation cost and factor mobility influence the relative value of predetermined geography.[3] However, no papers that I am aware of consider the knowledge of geographic characteristics (or measured characteristics) as something that could be influenced by the location of economic activity. The closest relatable empirical paper is Henderson et al. (2018), who show that changes in technology influence the comparative advantage of a location's specific geographic characteristics. My contribution to this literature is to provide empirical evidence that the location of economic development influences *where* natural resources are discovered, in the context of oil.

Another main strand of literature can be sourced back to David and Wright (1997), that questions whether a country's resource abundance is solely determined by geological conditions.[4] They argue that USA's lead as a mineral producer from 1870 to 1919 was a function of 'socially constructed' conditions, rather than 'first nature' factors, like geology. Bohn and Deacon (2000) takes this a step further by accompanying empirical work using a model highlighting ownership risk to explain observed differences in investment and production related to petroleum and forest resources. Both papers argue that the institutional setting matters for investment and resource accessibility. More recently, Arezki et al. (2019) and Cust and Harding (2020) provide additional evidence that institutional differences have an effect on oil wealth through increased inflow of exploratory investments. Arezki et al. (2019) focus on market orientation as a source for attracting investment capital, whereas Cust and Harding (2020) exploit the similar geological conditions close to country borders to identify the effect on exploratory drilling. My paper is related to the aforementioned papers, but is novel in its focus on geographic variation within countries, as measured by the exact location of economically developed areas.

My paper and findings also contribute to the resource curse literature (see Ross, 2001; Mehlum et al., 2006; Van der Ploeg, 2011; Andersen and Ross, 2014). Though it is well recognised that the location of oil wealth may be endogenous (Cassidy, 2019), my paper suggests that the challenge of obtaining sources of exogenous variation is not restricted to the cross-country level.

The remainder of the paper is structured as follows. The data is presented in section 2. In section 3, I describe the empirical strategy, followed by a presentation of the empirical specifications in subsection 3.1. Next, in section 4, subsection 4.1, I present the main results. In subsection 4.2 I use the results presented

[3]See also Henderson (1996), Helpman (1998) and Ellison and Glaeser (1999)

[4]See also: Southgate et al. (1991) and Deacon (1994)

in subsection 4.1 to quantify how much more oil have been discovered due to economic development and how much more underexplored subsurface potential there exist in currently underexplored areas. In section 5, I investigating regional heterogeneity, splitting the sample on Africa, Oceania, Asia, South America, Europe and North America. This section is followed by two subsections, where I discuss potential reasons for the observed difference between regions. In section 6, I investigate discovery specific outcomes. This is followed by a discussion and empirical analysis of potential mechanisms in section 7. Lastly, in section 8 I conclude with some final remarks. Supplementary material is given in the Appendix.

2 Data

The spatial data set used in this paper is created by merging four separate data sets using GIS software. All original data is geo-referenced with latitudes and longitude, and operates with a resolution of 0.25 degree or smaller. The base layer comes from Henderson et al. (2018), which consists of 242,184 grid cells of 0.25^2 degrees. The layer covers the entire globe, with the exception of offshore areas and places that are permanently covered with ice.

2.1 Oil discoveries

The oil-discovery data comes from Rystad Energy (UCube, 2020).[5] The data set consist of cross-sectional data on all onshore oil field discoveries throughout history, per 2019. The total number of recorded onshore discoveries is 19,238, including discoveries that have never produced anything and/or are abandoned, hence the only criterion to be included in the data set is that oil has been discovered. It is important to note that a field is not the same as a well, and each field will likely contain more than one well. To facilitate merging with the base-layer provided by Henderson et al. (2018), I collapse the data on a quarter of a degree, ending up with 7,170 grid cells, about 3% of the total number of grid cells, that have had at least one discovery throughout history.[6] The median discovery year of a field is 1985, and the average discovery size per grid cell with an discovery is 185 MMbbl (million barrels). The Rystad Energy data also includes discovery-specific records on *Discovery size*, *Discovery year*, exploration expenditure (*Exploration capex*), well expenditure (*Well capex*), operational expenditure (*Opex*) and infrastructure and surface installations expenditure (*Facility capex*), where the words in italics refer to variable names.[7] All the discovery specific variables are then merged on a grid cell-level. The expenditure data is created by Rystad Energy using a combination of modelling and field-company specific data.[8] In the case where it is difficult to obtain reliable data, Rystad Energy relies solely on accounts from similar fields to model the economic data.

[5]There are number of papers that use Rystad's data, for example: Bornstein et al. (2017) Arezki et al. (2019) Nordvik (2019), Asker et al. (2019) Andersen et al. (2022a), and Ahlvik et al. (2022). However, non, to my knowledge have exploited the spatial component to the extent I do in this paper.

[6]Each oil field is geo-reference within a maximum of one quarter degree, depending on the geographical extent of a particular field.

[7]See appendix A.2 for verbatim definitions of the expenditure variables. See appendix C.1 for summary statistics

[8]In total I have 7,105 grid cells that include a complete expenditure record on for all discoveries located the particular grid cell (the exception is Exploration Capex, where I have 7,087 grid cells).

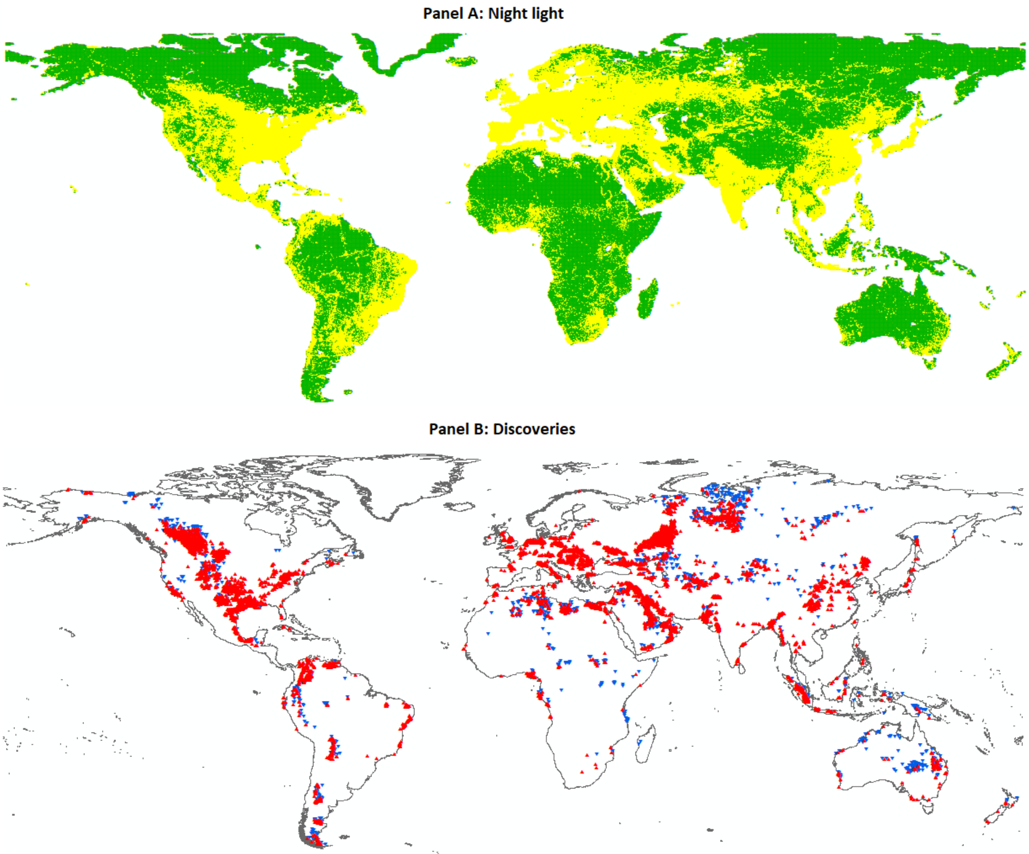

Figure 1: Panel A shows the location of all 242,184 grid cells. The green grid cells are those without any detectable night light, the yellow grid cells are those with detectable night light. Panel B shows the location of all oil discoveries. The red triangles are those located in a grid cells with detectable night light. The blue triangles are those located in a grid cells without detectable night light

2.2 Economically developed areas

To proxy economic development, I use night light data. The night light data is from Henderson et al. (2018), and is a 2010 radiance calibrated version.[9] I define a grid cell to be economically developed if the grid cell (EDC) has detectable night light. Original resolution of the night light data is of 0.5 arcminutes (approximately 1 km), but is aggregated up 0.25 degree to facilitate the merging of the other explanatory variables used in Henderson et al. (2018). An advantage of using night light data to proxy economic development is that it provides high-resolution data, without the risk of census-bias. Landscan (Bright et al., 2018), Worldpop (Tatem, 2017), The Gridded Population of the World (Doxsey-Whitfield et al., 2015) and G-econ (Nordhaus et al., 2006) are high-resolution data sets that could be considered as alternatives for proxying economic development. A problem with these alternatives, is that they rely on interpolating census data by using geographic data, which will limit the interpretation of my results, making it mechanically impossible to isolate the impact of favourable geography.[10]

A potential problem with using night light data is that light emission from petroleum activity might be recorded in the night light data through flares or regular activity at active fields, creating econometrically a case of simultaneity in my OLS-estimates. To mitigate this, have Henderson et al. (2018) removed all flares

[9] Following the procedure explained in Elvidge et al. (1999)

[10] Or, they assume uniform population density between census areas, as in the case of the Gridded Population of the World.

following the method of Elvidge et al. (2009). It is important to note, that even though flares (in addition to other types of reverse causality) could potentially influence my OLS-estimates, this is not a problem in my IV specifications, as the use of my instrument lets me source out spatial variation in economically developed grid cells that predates the first discovery.

2.3 Historic settlements

To form my instrument I use geo-referenced historic settlements, taken from Reba et al. (2016)'s global data set on all urban settlements from 3700 BC to 2000 AD.[11] To create the data set, Reba et al. use a combination of archaeological data and historical records. They include only the largest cities in a given era and geographic region, that is, only cities with population number above 20,000 from 800-1850 AD (above 40,000 in Asia for the same period), while in the period 3500-1000 BC they include cities that have more than 10,000 inhabitants. I use the distance to the nearest historic urban settlements as defined by Reba et al. (2016) to form my instrument. The idea is that the closer a grid cell is to a historic settlement, the more likely it is to contain modern economic development. In my analysis I only use historic settlements that are recorded for the years 1500-1800 AD, meaning that the most recent records are from year 1800, 35 years before the first discovery in 1835. In total, there are 654 settlements that satisfy this criterion see Fig. 2.

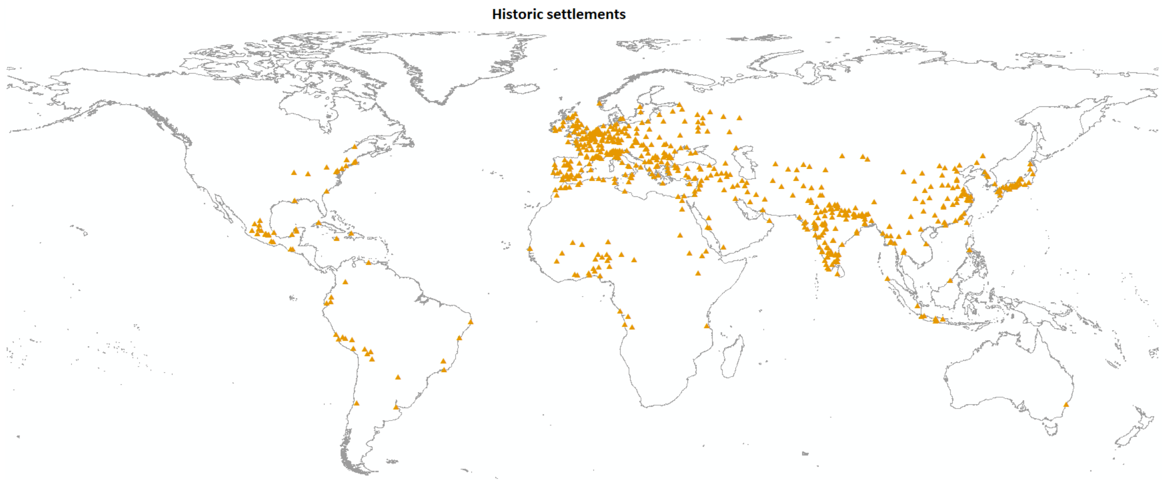

Figure 2: This figure shows the location of all 654 historic settlements (orange triangles) used to form my instrument.

2.4 Geology

The geology data comes from United States Geology Survey's (USGS) data bank, originally presented as 13 separate geological maps that, when combined, covers the entire globe: Africa (USGS, 2022l), Arabian Peninsula (USGS, 2022b), Arctic (USGS, 2022a), Australia and New Zealand (USGS, 2022d), Caribbean (USGS, 2022k), Europe (USGS, 2022e), Far East (USGS, 2022g), Former Soviet Union (USGS, 2022h), Iran (USGS, 2022m), North America (USGS, 2022c), South Pacific Asia (USGS, 2022f), South America (USGS, 2022j) and South Asia (USGS, 2022i). The maps are built of polygons, which encloses the area of any given bedrock type. The

[11] Reba et al. (2016) build on the work of Chandler (1987) and Modelski (2003)

Geology: Arabian Peninsula

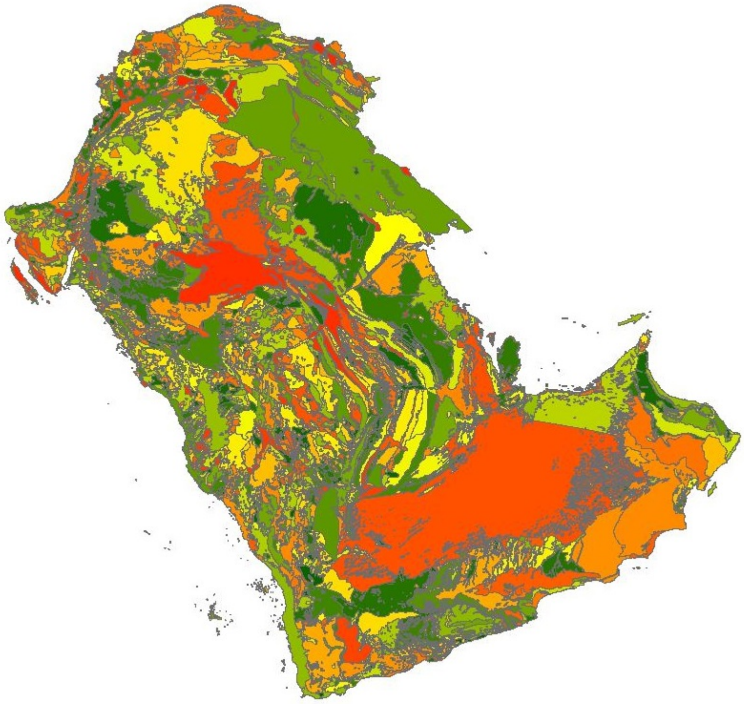

Figure 3: This figure shows the bedrock geology on the Arabian peninsula. Each color indicate a separate bedrock-type.

earth's bedrock is the first rock layer that are exposed if all deposits and loosely packed sediments were removed from the surface.[12] Geologists use bedrock geology to determine the subsurface potential of oil and gas.[13] The data used consists of 813 different types of bedrock, where 158 types have had at least one discovery. An alternative to bedrock geology data is to use data on sedimentary basins, which captures the deeper geological conditions. However, since the aim of using the geology data is to control for the potential correlation between the location of economically developed areas, a surface characteristic, and oil-forming geology, I argue that it is better to use bedrock geology, because it is the geological layer closest to the surface, and therefore the geology layer that is most likely to correlate with economic development. Another advantage of using bedrock geology instead of sedimentary basins, is that sedimentary basins have a more spatially lumpy distribution. This could result in spurious correlations with surface characteristics that are fundamentally independent of the geology. An example is the case of Cassidy (2019), who finds that the location of Muslim population correlates with certain types of sedimentary basins. Figure 3 contains an example of the type of geology data used. The figure shows the bedrock-geology distribution of the Arabian peninsula.

[12] A bedrock type is defined on the basis of lithology (general characteristics of sediments and rocks) and the era of formation (e.g. Mesozoic, Lower Jurassic, Pleistocene). Geologists analyse an area's petroleum potential by mainly using three techniques: surface mapping; core analysis; and seismic profiling, a technique that is used to obtain measures of the surface-to-subsurface geological correlation (Southard, 2007)

[13] Bedrock geology data is an essential part of the surface mapping and seismic profiling on which the subsurface conditions are partly determined. In addition, the bedrock itself might also contain oil and gas (Feng et al., 2015)

2.5 Biomes and geographic surface characteristics

The biome and geographical surface characteristics (GSC) data comes from Henderson et al. (2018), and consist of 14 biomes and 18 GSC-variables, all geo-referenced to the grid cell layer. A biome is an area categorized by biologists to have similar fauna and flora. Examples are *Tundra*, *Temperate Conifer Forest Jungle* and *Desert*. The biome-data is mutually exclusive, meaning that each grid cell is located within one of the fourteen biomes.[14] The GSC-variables include: *Ruggedness* (Nunn and Puga, 2012), *Elevation* (meter over sea level), *Malaria Mosquito survivor index* (Kiszewski et al., 2004), *Land suitability index* (ranks the grid cells suitability for agricultural crops), *Coastal* (dummy that indicates whether the grid cell contains a coastal area), *Distance to nearest river* (in km), *On river* (dummy for whether a river passes through the grid cell), *Distance to coast* (in km), *Temperature* (average monthly temperature from 1960- 1990), *Precipitation* (average monthly precipitation from 1960 1990), *Distance to nearest natural harbour* (in km), *Distance to nearest big lake* (>5000 km2), *Growing days* (length of growing period, in days), *Land* (area of grid cell on land), *Natural harbour within 25 km* (dummy), *Navigable river within 25 km* (dummy), *Big lake within 25 km* (dummy) and *Absolute latitude*.[15]

2.6 Climate assessment and other data

In order to evaluate differences in climatic conditions as a potential explanation for my result, I use data from an oil-industry climate assessment (ICA) conducted by Rystad Energy AS. The purpose of the assessment is to evaluate whether a location's climate increases the cost or risk for oil extraction and/or exploration. Locations are categorized into one of five categories: *benign*, *arctic hazard*, *arctic wear*, *harsh hazard* or *harsh wear*.[16] A location without any additional costs or risks associated with the climatic conditions is categorized as *benign*. Rystad uses detailed local information about a location's climatic conditions in their categorization. Examples are: "Hot arid steppe" and "Snow area, full year humid, cool summer".[17] The assessment is global, but does only include certain locations, and not complete areas, which makes the location of assessments potentially endogenous (see Figure 4). I will therefore only exploit local variation around a location of assessment (within 400 km), following the logic that grid cells close to a location of assessment have the same or very similar climate. I also note, that though the climate assessment only contains certain locations, they suggest some underlying patterns, for instance, the majority of assessments in Africa are *benign* (see figure 4.)

I also use data on country borders, which enables me to include country fixed effects in my specifications (Earth, 2017). And I will use GIS data on railroads (WTP, 2017b) and cities with a population more than 100k

[14] List of biomes: *Tundra*; *Tropical Subtropical Moist Broadleaf Forests*; *Tropical Subtropical Grasslands, Savannas Shrublands*; *Tropical Subtropical Grasslands, Savannas Shrublands*; *Tropical Subtropical Dry Broadleaf Forests*; *Temperate Grasslands, Savannas Shrublands*; *Temperate Conifer Forests*; *Temperate Broadleaf Mixed Forests*; *Montane Grasslands Shrublands*; *Mediterranean Forests, Woodlands Scrub*; *Mangroves*; *Flooded Grasslands Savannas*; *Deserts Xeric Shrublands*; and *Boreal Forests*.

[15] The absolute value of latitude makes it possible to control for the mechanical increase in grid cell size, when moving away from equator.

[16] The verbatim definitions of each category are reported in the appendix A.1

[17] The ICA is from 2019. This means that if the climatic conditions that creates an adverse impact on oil exploration and production have changed over time this would influence the value of this assessment, for instance due to technological changes. However, as I do not find any statistical significant relation between discovery year (see Table 3) and whether a discovery is located in an economically developed are or not, which means that the ICA should be able to capture systematic differences between areas.

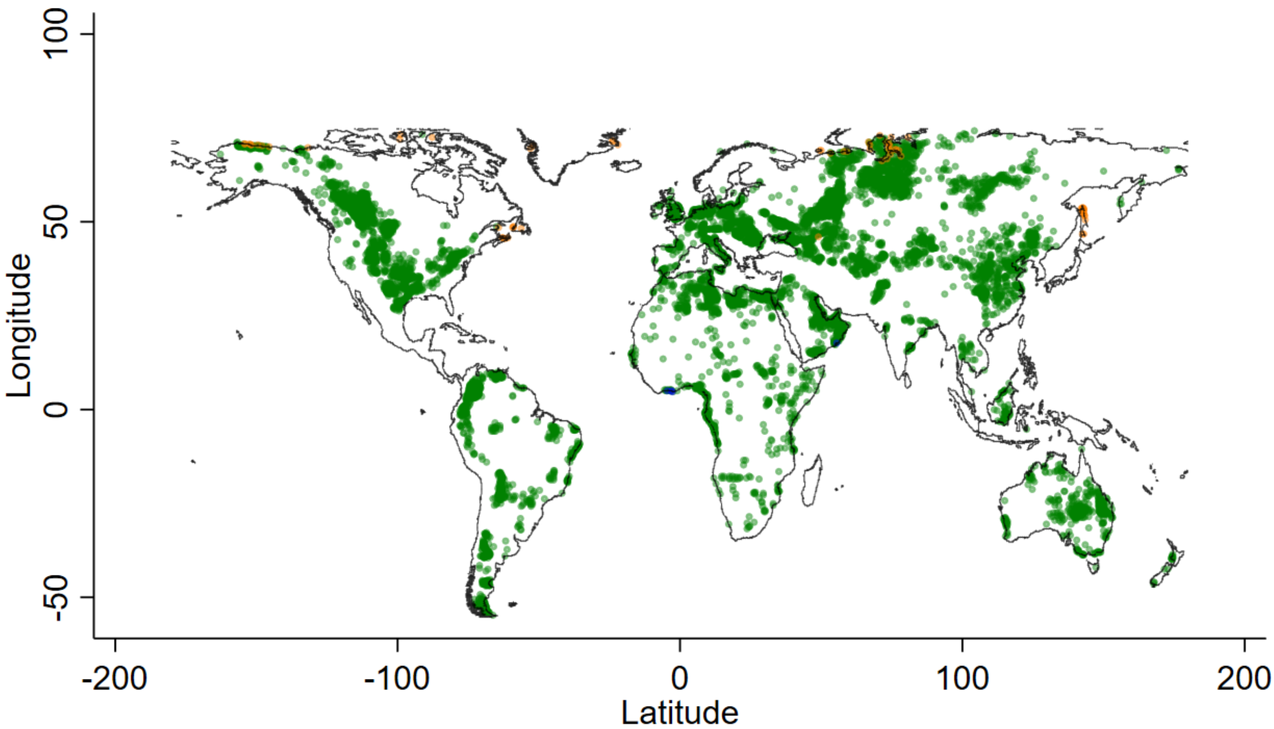

Figure 4: This figure shows the locations of the industry climate assessments conducted by Rystad Energy. The green dots are the locations categorized as *benign*. The light blue dots are the locations categorized as *harsh hazard*. The orange dots are the locations categorized as *artic wear*. The grey dots are the locations categorized as *arctic hazard*. The black dots are the locations categorized as *harsh wear*.

(WTP, 2017a) to proxy infrastructure access and closeness economic centers, respectively. Figure A1 reports the extent of the railroad and city data.

3 Empirical strategy

The aim of my research design is to obtain spatial variation in economic development that is independent of the exogenous subsurface conditions resulting in oil formation. This can be violated either by reverse causality, namely that the containment of oil discoveries within a grid cell increases the likelihood for economic development within the same grid cell, or if subsurface oil-forming geology influences surface characteristics which in turn affects the location of economically developed grid cells.

I address the problem of reverse causality by using spatial variation in economically developed grid cells that predates the first historical discovery. I do this by instrumenting economically developed grid cells on the distance to historic settlements. The settlements used dates back to before 1800 AD, 35 years before the first historical oil discovery, and is therefore unaffected by the location of future discoveries.

I address the potential worry that economically developed areas have different subsurface geology than undeveloped areas by including bedrock geology fixed effects. This means that I will exploit variation across grid cells within the same bedrock geology type. To further reduce the chance of subsurface confounders influencing my result, I include biome fixed effects and the set of GSC as controls.[18] The use of GSC controls,

[18] All of today's oil has been formed by being subjected to intense pressure and heat over hundred of millions of years. Accordingly

lets me also evaluate whether differences in geographic surface characteristic across developed and undeveloped grid cells are influencing the location of discoveries.

I also include country fixed effects to clear my estimate of country-level differences, like institutional quality, which are likely to influence both the location of development and the location of discoveries.

3.1 Empirical specification

My results are drawn from a two-stage-least-square estimation procedure, using the following equations:

$$\mathbf{EDC}_{icgb} = \gamma_0 + \gamma_1 \mathbf{DHS}_{icgb} + \Gamma' \mathbf{W}_{icgb} + \eta_c + \eta_g + \eta_b + \mathbf{u}_{icgb} \tag{1}$$

$$\mathbf{DISC}_{icgb} = \beta_0 + \beta_1 \widehat{\mathbf{EDC}}_{icgb} + B' \mathbf{W}_{icgb} + \delta_c + \delta_g + \delta_b + \mathbf{e}_{icgb} \tag{2}$$

The first stage is seen in equation (1), the second stage in equation (2). The instrument, $\mathbf{DHS}_{icgb}$, is *log*-distance to historic settlement, for grid cell i, in country c, within geology-type g and biome b. The endogenous variable, $\mathbf{EDC}_{icgb}$, is a dummy-variable equal to one if the grid cell is economically developed. Both equations includ three types of fixed effects (FEs): country FEs (δ_c and η_c), geology FEs (δ_g and η_g) and biome FEs (δ_b and η_b). Intercepts are captured by β_0 and γ_0. Controls are included in $\mathbf{W}$. The main outcome variable, $\mathbf{DISC}_{icgb}$, is a dummy-variable, with value one if a grid cell has experienced a discovery. I will also present estimation results where I have conditioned on grid cells that contain a discovery. The outcome variable is then changed to one of the field-specific economic variables (e.g. *Exploration capex*). In all specifications, the parameter of interest is β_1, which captures the linear relation between the instrumented endogenous variable, $\widehat{\mathbf{EDC}}_{icgb}$, and $\mathbf{DISC}_{icgb}$, the outcome variable, within an area of a country with the same geology- and biome-type. As $\mathbf{EDC}_{icgb}$ is a dummy-variable, β_1 is interpreted as the increased likelihood of discovery in an economically developed grid cell relative to an economically undeveloped grid cell in pp. $\mathbf{e}_{icgb}$ and $\mathbf{u}_{icgb}$ are error terms. Standard errors are adjusted for additional estimation uncertainty using a two stage procedure. All reported standard error are also clustered on the nearest 3x3 grid cells.[19]

4 Results

4.1 Main result

Table 1 reports my main results from estimating equations (1) and (2). The second stage IV results are reported in the row named IV-EDC. The table also includes OLS (EDC), reduced form (DHS) and first stage estimates.[20] Each column reports the results for specifications with different inclusion restrictions on FEs and controls.

to Cey et al. (2019) have the majority of known oil formed during the age of Pangea and the Mezoic age. Over the hundred millions of years from formation to first discovery there have been large continental shifts, multiple ice ages and major climatic changes, all of which further reduces the chance that the subsurface conditions needed in oil formation are correlated with surface characteristics.

[19] My results are robust to higher clustering level. In the appendix, table A2, I show that my main results in table 1 are robust to a clustering level of 9x9 grid cells.

[20] The OLS-estimates reported are produced by estimating only equation (2), using the endogenous variable $\mathbf{x}_{icgb}$. The reduced form estimates are produced by using my instrument directly in equation (2).

Outcome: DISC	(1)	(2)	(3)	(4)	(5)
IV-EDC	0.053***	0.069***	0.067***	0.057***	0.058***
	(0.003)	(0.004)	(0.004)	(0.006)	(0.010)
1st stage	-0.199***	-0.233***	-0.233***	-0.175***	-0.121***
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
1st stage F-test	24,031	13,574	13,169	5,800	2,548
EDC	0.052***	0.054***	0.044***	0.039***	0.036***
	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)
DHS	-0.011***	-0.016***	-0.016***	-0.010***	-0.007***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Country FEs		X	X	X	X
Geology FEs			X	X	X
Biome FEs				X	X
GSC					X
N	242,184	242,082	242,082	242,082	242,082

Table 1: This table reports estimates IV-estimates corresponding to equation 1 and 2. The OLS and reduced form results are reported in the rows noted as EDC and DHS, respectively. Each separate column reports the result with different fixed effects and controls included. The reported standard errors are calculated by clustering on 3x3 grid cells.

The IV-estimates show that there is a strong statisticaly positive relationship between the location of economic development and oil discoveries. A result that holds across all different specifications. The first stage estimates indicate no signs of a weak instrument, showing that the likelihood of a grid cell to contain economic development decreases with the *log*-distance to nearest historic settlement. I also need to assume monotonicity, that is that the likelihood of economic development is monotonic decrease with the distance to historic settlement. In the appendix, Table A1, I report the results with the additions of squared and cubed DHS in the first stage. The result is insignificant different from the results reported in Table 1.[21] My preferred IV-specification, reported in column 5, estimates that economically developed grid cells have 5.8 pp higher likelihood of having experienced a discovery than economically undeveloped grid cells.

Next, note that both the IV-EDC and 1st stage estimates are unchanged with the inclusion of geology fixed effects. This is consistent with the identifying assumption, namely that the predicted locations of economically developed grid cells are uncorrelated with subsurface oil-forming geology. In column 4, we can see that adding the biome FEs slightly reduces the estimates. However, conditional on the biome FEs, I find no significant

[21] Note that the results in Table A1 relies on stronger monotonicity assumption, as I need all varieties of the instrument to invoke monotonic selection into treatment.

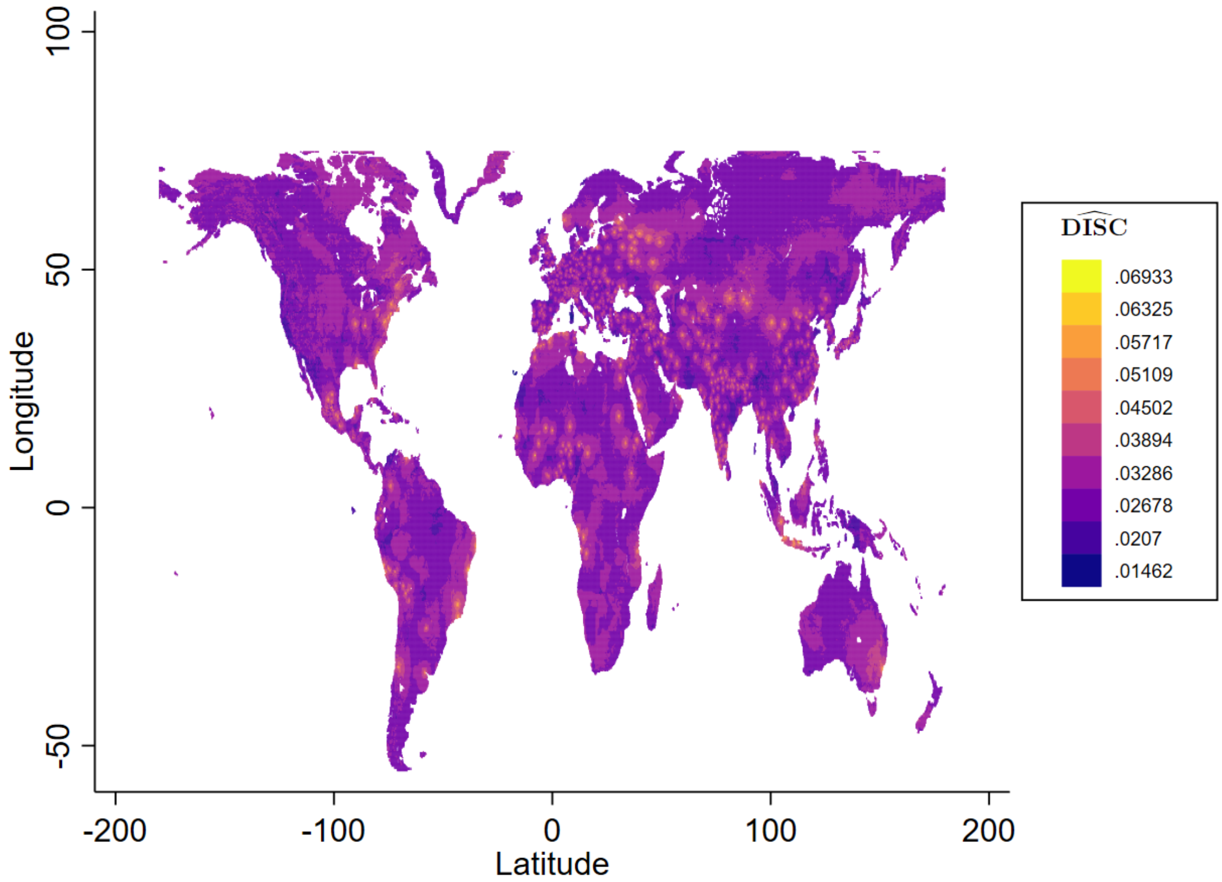

Figure 5: This figure shows the estimated probability for a grid cell to contain a discovery given the estimated probability for a grid cells to be economically developed. The estimated probabilities are drawn from a IV specification corresponding to the specification used to produce the estimates in Table 1, column 5. Redder grid cells have higher predicted values, bluer grid cells have lower predicted values.

change in my IV-estimate when I control for geographic surface characteristics (column 5). Note also that the geographic surface characteristics have a rather large impact on the first stage estimate. Given that the second stage is unaffected, this can be interpreted as a sign of homogeneous treatment effect, which will (if assumed to be true) increase the figures in my quantification exercises (see section 4.2).

I find that the OLS-estimates, across all specifications, are smaller than the IV-estimates. As the OLS-estimates might be influenced by reversed causality, a possible explanation for the lower estimates is that oil-discoveries negatively impact the likelihood of a grid cell to be economically developed. One can, for instance, imagine that oil activity puts pressure on the local economy by driving up housing and input factor prices, thereby discouraging other non-oil related economic activity to be located in a grid cell with a discovery. As the instrument invokes only a part of the variation in the location of economically developed grid cells, an alternative interpretation is that the IV-estimates deviate from the OLS-estimates because of heterogeneous treatment effects.

I also find that the reduced form estimates (DHS) are all significant and negative, meaning that grid cells located farther away from historic settlements are less likely to have experienced a discovery.

Figure 5 shows differences in estimated probabilities of a grid cell to experienced a discovery due to economic development. Redder grid cells have higher predicted value, the bluer grid cells have lower. Conversely, the bluer grid cells are those predicted to be underexplored in reference to the subsurface potential due to low probability of economic development.[22] When interpreting the figure it is important to note that the predicted values are cleared of country-, geology- and biome-FEs and geographic surface characteristics. One should therefore interpret the figure having in mind that the only variation between grid cells are the probability of economic development, i.e. geology, country, and natural geographic surface characteristics are fixed.

4.2 Quantification

My preferred estimate tells us that the economically developed grid cells have 5.8 pp higher likelihood to have experienced a discovery. Under the assumption that my estimate are not influenced by systematic differences in geology, this implies that there exist relative more undiscovered oil in undeveloped grid cells. The aim of this section is to translate my estimate into *barrels*, thereby quantify how much more oil have being discovered due to economic development, and how much underexplored subsurface potential there exists in currently economically undeveloped grid cells. For simplicity I start by assuming homogenous treatment effect. This assumption is later relaxed, by taking into account the shares of alwaystaker and nevertakers.

The law of total probabilities enables me to find the estimated conditional probabilities. In other words, the probabilities of a grid cell to have experienced an oil discovery conditional on it being developed or undeveloped. This gives me $\mathbb{P}(DISC|EDC = 1) = 0.081$ and $\mathbb{P}(DISC|EDC = 0) = 0.023$, which means that the discovery rate is estimated to be about 3.5 times higher in economically developed areas. As I now have the conditional probabilities, I can use these in combination with the average discovery size (185 MMbbl) to calculate the expected number of barrels in economically developed and undeveloped grid cells. The result is about 15 MMbbl and 4.25 MMbbl, respectively. This means that if I were to draw many grid cells at random I would expect on average that each economically developed grid cell to contain about 15 MMbbl (of discovered oil) and economically undeveloped grid cell to contain 4.25 MMbbl. To find the total discovery differential in barrels, I take each expected value and multiply it with the number of grid cells that are economically developed (97,182) and undeveloped (145,002) before taking the difference. This suggests that 839 Gbbl of oil have been discovered due to local economic development. Next, I calculate how much underexplored subsurface potential there exist in economically undeveloped grid cells relative to economically developed grid cells in amount of barrels. I do this by calculating the increase in expected amount of oil discovered in economically undeveloped grid cells given an discovery rate increase of 5.8 pp. Under the assumption of homogenous treatment effect, I find that if currently economically undeveloped grid cells have had the same discovery rate as economically developed grid cells this would amount to an increase in total oil supply of about 1,556 Gbbl, 1.17 times more than the current *total* number discovered onshore oil of 1,327 Gbbl.

The homogenous treatment effect assumption entails that the predicted variation in economic development sourced from my instrument has the same impact on the discovery rate as other sources of economic development.

[22] See Appendix section C for a detailed description of the procedure to create the map. Note also that a map of the first-stage will look identical if I normalize the scale as the predicted first-stage values are proportional to the predicted second-stage values.

More formally, I have assumed that the local average treatment effect (LATE) I have identified is equal to the average treatment effect (ATE). I can relax this assumption by only using the share of compliers in my quantification, i.e use only the variation in economic development sourced from my instrument. To do so I need to estimate the share of alwaystakers and nevertakers, i.e. the share of grid cells that are unresponsive to my instrument. I do this my summing over the predicted probabilities of economic development conditional on the grid cells that are economically developed and undeveloped. The total shares of alwaystakers is estimate to be 23.61% and the share of nevertakers is estimated to be 36.26%. Under the monotonicity assumption this implies that the total share of compliers is 40.13%. Taking the shares of compliers into account, I find that 350 Gbbl of the discovered oil is due to differences in economic development, and that the underexplored potential in currently undeveloped grid cells is 613 Gbbl. See section B in the Appendix for more details.[23]

5 Regional heterogeneity

My main result tells us that in general there have been discovered more oil in economically developed grid cells, and that there exist large undiscovered reserves distributed across the economically undeveloped areas of the globe. However, the global estimates reported in section 4.1 and 4.2 may overlook important heterogeneity. In this section I increase the flexibility of my empirical design by dividing the global sample into different geographic regions.[24] This enables me to identify potential cross regional heterogeneity. In particular, it allows me to identify the underexplored potential per region, which are of special interest due to the importance of future energy supply locations.

In Table 2, I report IV-EDC estimates over six different subsamples, corresponding to the following geographic regions: Africa, Oceania, South America, Asia, Europe and North America. For all regions, except for Europe and North America, I find a significantly higher discovery rate in economically developed grid cells.

To obtain the underexplored subsurface potential I multiply the IV-EDC estimates with the average discovery size and number of economically undeveloped grid cells. I use the within-region average discovery size to account for potential cross-region differences in geology. I find that the total underexplored potential is 1345 Gbbl, somewhat lower than the global estimate: 1,556 Gbbl. See Table 2, row "Underexplored (Gbbl)|ATE". When I only use the variation in economic development sourced from my instrument I get that the that the total additional oil discovered in economically developed grid cells is 523 Gbbl, again somewhat lower than the global estimate: 613 Gbbl. See Table 2, row "Underexplored (Gbbl)|LATE". I find that Asia is by far the region with most underexplored potential without the homogeneous treatment effect assumption. Note that the large reduction in the estimated underexplored potential in Africa when the homogenous treatment effect assumption is relaxed is do to the low share of compliers.

Table 2 shows that there exists substantial heterogeneity in the IV-EDC estimates cross regions. In the following two subsections I discuss potential explanations for the observed heterogeneity across regions.

[23] The quantification procedure is similar to what used in Dahl et al. (2014).

[24] An alternative specification is to add treatment-interactions, see Table A3. However, a downside of doing this is that the controls will not be region-specific without additional interaction. And adding interactions for all controls is not feasible, due to

Outcome: DISC	Africa (1)	Oceania (2)	S. America (3)	Asia (4)	Europe (5)	N. America (6)
IV-EDC	0.174***	0.374***	0.097***	0.067***	0.012	0.024
	(0.041)	(0.101)	(0.024)	(0.022)	(0.038)	(0.057)
EDC = 1	9,079	2,417	11,139	29,639	25,402	19,370
EDC = 0	32,266	10,803	14,212	21,609	34,632	31,480
Share of compliers if EDC =1	22.29%	18.66%	45.37%	58.54%	42.77%	38.34%
Share of compliers if EDC =0	21.87%	18.20%	42.89%	56.89%	42.00%	38.12%
Average discovery size	77	28	84	348	114	176
Underexplored (Gbbl)|ATE	432	113	116	504	47	133
Underexplored (Gbbl)|LATE	95	21	50	287	20	50
1st stage	-0.060***	-0.109***	-0.161***	-0.104***	-0.106***	-0.062***
	(0.007)	(0.019)	(0.008)	(0.004)	(0.005)	(0.005)
1st stage F-test	79	33	430	546	519	134
N	41,345	13,220	25,351	51,248	60,034	50,850

Table 2: This table reports the IV-estimates corresponding to equation 1 and 2 (IV-EDC), and underexplored subsurface potential over six different regions. Each regression-estimate a produced by regressions on mutually exclusive subsamples, corresponding to grid cells located within the regions reported in the column headings. All regressions include the complete set of GSC variables, geology-, country-, and biome-FE. The reported standard errors are calculated by clustering on 3x3 grid cells. I have in rows with suffix "|ATE" assumed homogenous treatment effect. The rows that include in the name "Discovered oil" contain the amount of additional oil discovered in economically developed grid cells, compared to undedeveloped grid cells.

5.1 Exploration duration

One potential reason for the differences in estimates across regions might come from their different histories of oil-discovery duration, that is, for how long in each region there has been oil exploration. Assuming that the oil price is unpredictable and that there is a limit on exploration resources available to companies at any given point in time within each region, then, if being economically developed increases the value of a potential discovery, it is presumably favorable to begin exploration for oil in the economically developed grid cells. For example, if we assume that the oil price is a random walk, it follows that the oil-price variance increases over time, thereby making it less risky to start exploring straight away. And then, if there is a limit on available exploration resources each time period, there is a need to prioritize, which renders exploration in the areas with highest expected net present value most favorable. Over time, though, the relative depletion of new discoveries in economically developed areas is likely to increase, making exploration in undeveloped areas relatively more attractive. Hence, we would expect, depending on the general profit level, that the relative discovery rate across areas to somewhat converge over time.[25] Therefore, if regions differ in the oil-discovery duration, this could explain why the relative discovery rate across developed and undeveloped areas differs across regions.

the curse of dimensionality. I find it therefore better to split the sample, thereby reducing the parameter space.

[25] Since economically developed areas are independent to the subsurface conditions needed in oil formation we would have that if the oil price is high enough, so that all areas have a positive expected profit from exploration, we will see a total convergence given enough time.

I empirically investigate this by plotting the IV-EDC estimates (corresponding to the output in Table 2) as a function of the year of the first discovery in each region.[26] As can be seen in Figure 5, the IV-EDC estimates increases (approximately linearly) with the year of the first discovery, in line with my conjecture.

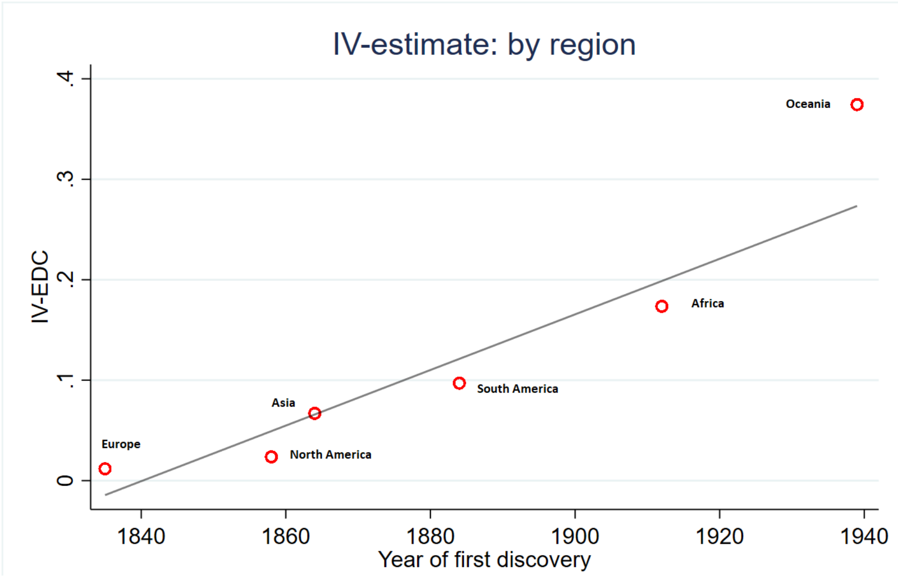

Figure 6: This figure reports the IV-estimates for all 6 regions across their respective first discover year. The red circles corresponds to the point estimates in table 4, row 1. The black line depicts the linear relationship between IV-estimates and the first discovery year in a region.

5.2 Fraction of developed areas

Another potential explanation for the heterogeneous IV-EDC estimates is the fraction of developed areas within a region.[27] To see why, consider an example: imagine a region of 1000 grid cells, where only 10 grid cells are economically developed. Then, assume that the subsurface geology required in oil formation, is randomly distributed, and that the within-grid cell probability of a (profitable) discovery, if one decide to explore, is 0.1. Also assume that the available exploration capital is constrained, limiting exploration to 20 grid cells. Given that economically developed grid cell increases the potential value of a discovery, we would expect that all 10 economically developed grid cells will be explored. The remaining exploration capital is used in undeveloped grid cells. The expected number of discoveries is then 1 in both economically developed and undeveloped grid cells, with discovery rates of 10% and 0.101%, respectively. The resulting discovery rate differential is approximately 9 pp. Conversely, if 990 of 1000 grid cells were economically developed, we would have a discovery rate difference of approximately 0.2 pp. Note that without capital constraints we would see no difference in the discovery rate.

This example illustrates why regions with higher fraction of developed grid cells are expected to have a lower discovery rate differential. The pattern in Figure 7 supports this interpretation, showing a linear relation between the IV-EDC estimates and fraction of economically developed grid cells across regions. Regions with a small fraction of economically developed grid cells, like Africa and Oceania, have a higher IV-EDC, in line with the intuition from the example above.

[26]Explaining why different regions have discovered oil at different points in time is beyond the scope of this paper

[27]Note that the inclusion of country-FEs will only equate the means within each regions and not across the samples of regions.

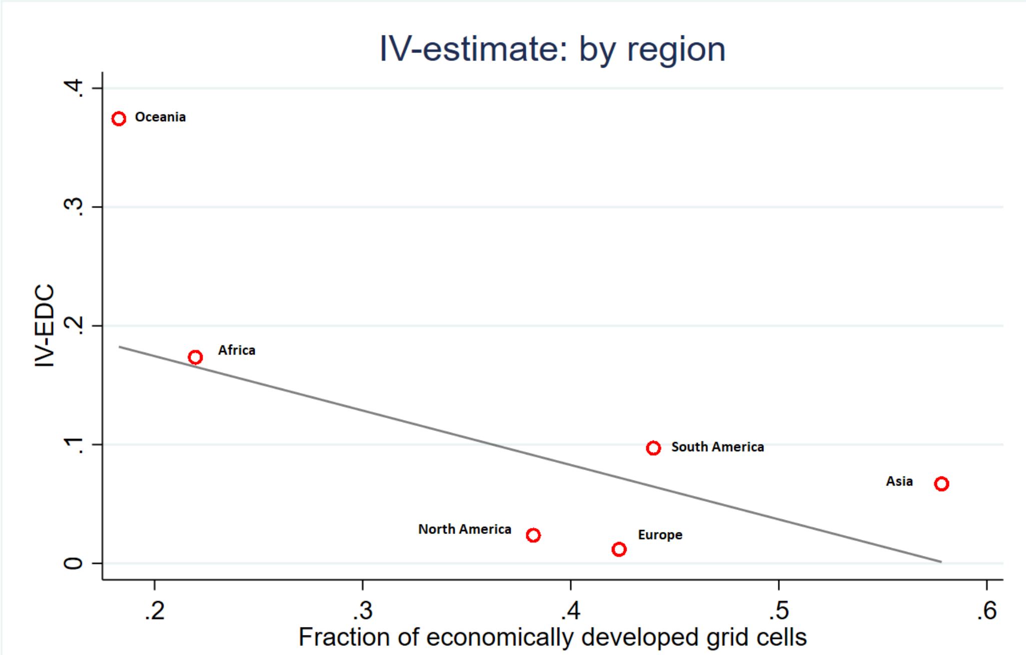

Figure 7: This figure reports the IV-estimates for all 6 regions over the fraction of economically developed grid cells. The red circles corresponds to the point estimates in table 4, row 1. The black line depicts the linear relationship between IV-estimates and the fraction of developed grid cells in a region.

6 Discovery specific outcomes

Table 3 reports the result on various discovery specific outcomes. The sample is restricted to only consist of grid cells that have experienced at least one discovery. The IV-EDC estimates are obtained by regressing my IV-specification, with the full set of fixed effects and controls. The sample is consistent across all columns, with the exception of column 2, where I have excluded 18 grid cells due to missing data.[28] All outcomes are in *logs*.

Column 1 shows that the *Exploration Capex* for predicted economically developed grid cells is statistically higher, with a point estimate of 2.62 log-difference. This means that oil extractors on average direct much more exploratory efforts toward discoveries located in (predicted) economically developed grid cells than (predicted) economically undeveloped grid cells. Under the assumption of profit-maximizing oil extractors, this behavior is consistent with a higher valuation of oil discoveries in economically developed grid cells, as an extractor that searches for oil will always direct the marginal dollar of investment toward the location with highest expected profits from exploration.

Column 2 and 3 report the results for *Discovery Size* and *Well Capex*. Here I find that both estimates variables are lower in (predicted) economically developed grid cells, but the difference is statistically insignificant. As *Discovery size* and *Well capex* is likely to be driven predominantly by geological factors, the insignificant results are consistent with the assumption of independence between subsurface oil-forming geology and the predicted location of economically developed grid cells.

Column 4 reports a statistically higher *Facility capex* in predicted economically developed grid cells. This suggest that oil extractors on average allocate more resources to discoveries located in (predicted) economically

[28] I am missing expenditure data on 65 grid-specific discoveries. This makes the sample size lower than 7,170, which is the total number of grid cells that contain a discovery.

developed grid cells in the pre-production phase, for example investing in factors such as infrastructure and production facilities. Higher levels of expenditure in infrastructure and production facilities can potentially also explain why I find higher, though insignificant, *Opex* in predicted economically developed grid cells (column 5), as an increase in *Facility capex* is likely to increase the chance of production start, and, hence, explain the higher production expenditure.

Column 6 reports a negative insignificant estimate on *Discovery year*. Even though the estimate on *Discovery year* is insignificant, this tells us that the average year of discovery in (predicted) economically developed grid cells is lower. It is therefore possible that factors developing over time, like technology, are influencing the expenditure and discovery size estimates. I investigate this by I running the same regressions on a subsample of grid cells that contain discoveries with an average discovery year higher than 1980. Thereby, limiting the period the discoveries dates back to, resulting in a reduced *Discovery year* difference of five years, compared to the previous 43. When doing this, I find a clear reduction in the absolute value of the *Discovery Size* and *Well Capex* estimates, with resulting estimates of -0.007 (s.e. 1.363) and 0.060 (s.e. 2.08), respectively. The rest of the estimates are more or less unchanged.[29] This tells us that the given interpretations of the results in Table 3 also holds for discoveries in the later part of my sample.

Outcome:	(1) Exploration capex	(2) Discovery size	(3) Well capex	(4) Facility capex	(5) Opex	(6) Discovery year
IV- EDC	2.62**	-1.96	-1.58	3.10*	2.27	-43.64
	(1.142)	(1.336)	(1.819)	(1.690)	(1.839)	(70.107)
1st stage	-0.038***	-0.038***	-0.038***	-0.038***	-0.038***	-0.038***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
1st stage F-test	30	30	30	30	30	30
N	7,087	7,105	7,105	7,105	7,105	7,105

Table 3: This table reports estimates IV-estimates corresponding to equation 1 and 2, with the outcome variable reported in the column heading. The regressions include the complete set of GSC variables, geology-, country-, and biome-FE. The sample consist only of grid cells that have experienced a discovery and for which I have data on the outcome variable. The reported standard errors are calculated by clustering on 3x3 grid cells.

7 Mechanisms

Favorable natural geography. A potential explanation for the higher discovery rate in economically developed grid cells is that these grid cells are located in areas with geographical conditions that reduces the cost of doing extractions and exploration, and thereby making it more attractive to make a discovery. However, as seen in Table 1, row 5, I find that the inclusion of the set of GSC-variables as controls does not affect the IV-EDC estimate, which suggests that differences in the geographic surface characteristics are not driving my IV-EDC estimate. The same regression output also shows that the first stage estimate is significantly changed when I

[29]Table A5, in the appendix, includes all outcomes as in Table 3 over a sample of grid cells with average discovery year higher than 1980.

include the GSC-variables, indicating that the GSC-variables and EDC are correlated.[30] The unchanged second stage estimate, together with the correlation between EDC and GSC, suggests that though geographic surface characteristics impact the location of economically developed grid cells, they have small or no direct effect on the location of discoveries. Note that it is still possible that the standard geographic surface characteristics are less relevant for the oil sector, while other, unobserved characteristics are more relevant. To account for this possibility I investigate how my results are affected by controlling for Rystad Energy's ICA. I do this by adding five indicator variables in equation (1) and (2), one for each climate category: *benign*, *arctic hazard*, *arctic wear*, *harsh hazard* or *harsh wear*. I only find a very small and statistically insignificant change, with a resulting IV-EDC estimate of 6 pp (Table 4, column 1).

Economic intensity. I proxy the degree of economic development by using *Night light intensity*. *Night light intensity* captures a combination of agglomeration (i.e., how many people are living in an area) and the level of economic activity per capita, both of which are commonly used as proxies for economic intensity. The degree of economic development, or intensity, might influence both the supply side and the demand side of the location of discoveries. On the supply side, higher intensity might reduce cost for extractors by facilitating better access to more productive labor and capital. In addition, the local demand for the oil and gas resulting from increased economic intensity might also be higher. Table 4, column 1, reports the estimate of instrumented night light intensity in a grid cell (IV-NLI) on the DISC-variable, using a subsample which only contains economically developed grid cells, i.e. grid cells with detectable night light. I isolate the intensive margin by restricting the sample to economically developed grid cells. The resulting estimate is close to zero and statistically insignificant, hence, there is no indication that the degree of economic development influences the location of discoveries.

Geographical centrality. Another potential explanation for why economically developed grid cells have a higher discovery rate is their proximity to economic centers. The distance to economic centres could influence the discovery rate by providing shorter transportation of both output and input factors. I proxy closeness to economic centers by calculating the distance between each grid cell and the closest city.[31] I then subsample grid cells that are within 0.9 degree ($\approx$ 100 km) radius of a city, and regress my IV-specification with the full set of FEs and controls. If closeness to economic centers is a key explanation for the higher discovery rate in economically developed areas we would expect a reduced discovery rate difference between economically developed and undeveloped grid cells when all grid cells in the sample are located close to an economic centre. However, as reported in Table 4, row 2, I find a significantly higher discovery rate in (predicted) economically developed areas, with a IV-EDC-estimate of 15.5 pp. The higher estimate might suggest that there exist complementarities between closeness to an economic center and economically developed grid cells, making it

[30]Table A4 reports the whole set of GSC-estimates. Note that even though the GSC estimates show up as statistically significant, they should be interpreted as reduced form estimates because of the correlation with economic development, as seen from the first-stage.

[31]A potential worry of using distance to modern cities (per 2017) is that they are not relevant for historic discoveries. However, as the median discovery year is 1985, it is likely that cities present in 2017 also where also present at the time of discovery for most oil fields. This claim is supported by researchers that have found that historic development patterns have been very stable across time, making it a good predictor for where modern economic activity takes place. For instance, Bleakley and Lin (2015) find that early US cities have not experience any particular decline after the initial period of formation. The same goes with settlements in Africa, where Jedwab et al. (2017) shows that populations patterns in Kenya and Ghana have been relative stable since colonial times. See also Krugman (1991a)

even more attractive to discover oil in economically developed grid cell if the grid cell is located close to an economic centre.

Structural connectedness. The final potential explanation for the higher discovery rate in economically developed grid cells that I explore is access to infrastructure. I use closeness to a railroad to proxy infrastructure accessibility. If infrastructure access can explain why economically developed grid cells have experienced more discoveries we should expect to see a reduced, if not eliminated, discovery rate differential between economically developed and undeveloped areas when both are located close to infrastructure. I investigate this by, first, calculating the distance between all grid cells and rail roads. Then I subsample grid cells that are within 0.9 degree ($\approx$ 100 km) radius of a railroad, and regress my IV-specification with the full set of FEs and controls. The result is reported in Table 4, row 3. I find a statistically insignificant difference of 3.6 pp. Given that the analysis on closeness to economic centres indicated complementarities between the location of economically development nearby cities, I want to make sure that my analysis on closeness to rail roads are not influenced by the locality of cities, which potentially correlates with the location of rail roads. I do this by further restricting my sample to grid cells that are located within 0.9 degree of a railroad, but farther than 0.9 degree from a city. Here I find that my estimate is reduced to 0.9 pp (see Table 4, row 4), down from 3.6 pp on the less restricted sample. Both these results indicate that, among my alternative explanations, access to infrastructure stands out as the most important feature of economic development that can explain the higher discovery rate in economically developed areas.

Outcome: DISC	(1)	(2)	(3)	(4)	(5)
IV-EDC	0.060***		0.155**	0.036	0.009
	(0.011)		(0.073)	(0.026)	(0.026)
IV-NLI		-0.001			
		(0.003)			
1st stage	-0.118***	-0.528***	-0.039***	-0.065***	-0.074***
	(0.003)	(0.013)	(0.003)	(0.003)	(0.004)
1st stage F-test	2,141	1,552	159	578	391
ICA	*X*				
EDC		*X*			
City (0.9 degree)			$<$		$>$
Railroad (0.9 degree)				$<$	$<$
N	193,947	97,067	30,137	111,823	85,249

Table 4: This table reports the results from four different regressions, all of which includes the complete set of GSC variables, geology-, country-, and biome-FE. Column 1 reports the IV-EDC estimate from a regression with the addition of ICA-controls. The sample consist only of grid cells within 3.2 degree ($\approx$ 400 km) radius of a location where it has been conducted a climate assessment. Column 2 reports the result of a regression with *night light intensity* (NLI) as instrumented explanatory variable, conditional on grid cells that have detectable night light. Column 3 reports the IV-EDC estimate, conditional on the grid cells that are within 0.9 degrees ($\approx$ 100 km) of a large city. Column 4 reports the IV-EDC estimate, conditional on the grid cells that are within 0.9 degrees ($\approx$ 100 km) of a rail road. Column 4 reports the IV results from EDC, conditional on the grid cells that are within 0.9 degrees ($\approx$ 100 km) of a rail road, but farther away than 0.9 degrees ($\approx$ 100 km) of a large city. The reported standard errors are calculated by clustering on 3x3 grid cells.

8 Conclusion

This paper investigates how the location of economically developed areas have influenced the location of *known* oil resources. To do so, I make use of a high resolution geo-referenced oil data set that includes all historic discoveries. My main result shows that economically developed areas are 5.8 pp more likely to have experienced an oil discovery. This is a robust result, that holds when accounting for reverse causality, country-level, geology differences and natural geographic surface characteristics. I go on to consider different features of economic development as a source of the observed discovery differential, finding that infrastructure access might be a key explanation.

I believe that the paper's main contribution is the study of a new economic channel to explain the location of *known* natural resource endowment. This papers argues that the location of economically developed areas impact the relative discovery rate of oil through the rise of differences in economic incentives to make a discovery. Though this paper focuses on oil, it is imaginable that the location of economic development has a similar impact on other natural resources. For instance, one can imagine that minerals, timber and fish resources also would increase in value if economic development is located nearby. If the value of natural resources is endogenous to local economic development this impacts how we value spatial dispersion of economic development. In particular, the opportunity costs of increased agglomeration. However, studying how local economic development influences other natural resources may prove a challenge, mainly due to the threat of reverse causality. In contrast to oil, most natural resources have been valued for a very long time, serving as an obstacle to obtaining predetermined spatial variation.

A Appendix

Outcome: DISC	(1)	(2)	(3)	(4)	(5)
IV-EDC	0.053***	0.071***	0.068***	0.055***	0.053***
	(0.003)	(0.004)	(0.004)	(0.005)	(0.009)
1st stage:					
DHS	0.474***	0.225***	0.145 ***	0.362***	0.253***
	(0.038)	(0.035)	(0.035)	(0.034)	(0.034)
DHS^2	-0.124***	-0.047***	-0.027***	-0.074***	-0.048***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
DHS^3	0.007***	0.001**	-0.000	0.003***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
F-test	14398.76	5507.16	5769.14	2138.50	949.08
Country FEs		X	X	X	X
Geology FEs			X	X	X
Biome FEs				X	X
GSC					X
N	242,184	242,082	242,082	242,082	242,082

Table A1: This table reports estimates IV-estimates corresponding table 1, row 5 with additions of squared and cubed DHS in the first-stage. The standard errors are calculated by clustering on 3x3 grid cells

Outcome: DISC	(1)	(2)	(3)	(4)	(5)
IV-EDC	0.053***	0.069***	0.067***	0.057***	0.058***
	(0.005)	(0.007)	(0.007)	(0.008)	(0.013)
1st stage	-0.199***	-0.233***	-0.233***	-0.175***	-0.121***
	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)
1st stage F-test	4695.87	3861.05	4430.50	2216.74	1133.93
Country FEs		X	X	X	X
Geology FEs			X	X	X
Biome FEs				X	X
GSC					X
N	242,184	242,082	242,082	242,082	242,082

Table A2: This table reports estimates IV-estimates corresponding table 1, row 5. The standard errors are calculated by clustering on 9x9 grid cells (about 7000 km2). The number of clusters is 2,908.

Outcome: I{Discovery = 1}	(1)
IV-EDC (Africa)	0.088***
	(0.0155)
IV-EDC (Oceania)	0.120 ***
	(0.024)
IV-EDC (South America)	0.041***
	(0.013)
IV-EDC (Asia)	0.086***
	(0.016)
IV-EDC (Europe)	0.039**
	(0.017)
IV-EDC (North America)	0.022
	(0.016)
1st stage F-test	419.43
N	242,082

Table A3: This table reports all point estimates of one regression, corresponding to column 5 in table 1 with the addition of region-interactions. The reported standard errors are calculated by clustering on the nearest 3x3 grid cells.

Outcome: I{Discovery = 1}	(1)		
IV-EDC	0.0583***	Dist. natural harbor	-0.0000***
	(0.0098)		(0.0000)
Ruggedness	-0.0004***	Dist. big lake (km)	0.0000***
	(0.0001)		(0.0000)
Elevation (m)	-0.0023*	Growing days	-0.0001***
	(0.0014)		(0.0000)
Land suitability	-0.0078	Malaria index	0.0004***
	(0.0052)		(0.0001)
Coastal	-0.0016	Land	0.0000***
	(0.0022)		(0.0000)
Dist. nearest river (km)	0.0000**	Abs. latitude	0.0010***
	(0.0000)		(0.0002)
On river	-0.0056	Natural harbor (<25km)	-0.0129***
	(0.0055)		(0.0032)
Dist. coast (km)	0.0289***	River (<25km)	0.0017
	(0.0054)		(0.0046)
Temperature	0.0016***	Big lake (<25km)	-0.0102**
	(0.0003)		(0.0044)
Precipitation	0.0000		
	(0.0000)		
		1st stage F-test	2,548
		N	242,082

Table A4: This table reports all point estimates of one regression, corresponding to column 5 in table 1. The reported standard errors are calculated by clustering on the nearest 3x3 grid cells.

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome:	Exploration capex	Discovery size	Well capex	Facility capex	Opex	Discovery year
IV-EDC	1.787	-0.007	0.060	3.401	2.927	-4.857
	(1.380)	(1.364)	(2.083)	(1.690)	(2.229)	(7.795)
1st stage	-0.0432***	-0.0432***	-0.0432***	-0.0432***	-0.0432***	-0.0432***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
1st stage F-test	19	19	19	19	19	19
N	4,023	4,039	4,039	4,039	4,039	4,039

Table A5: This table reports estimates IV-estimates corresponding to equation 1 and 2, with the outcome variable reported in the column heading. The sample is restricted to grid cells that contain discoveries with an average discovery year higher than 1980. The regressions include the complete set of GSC variables, geology-, country-, and biome-FE. The sample consist only of grid cells that have experienced a discovery and of which I have data on the outcome variable. The reported standard errors are calculated by clustering on 3x3 grid cells.

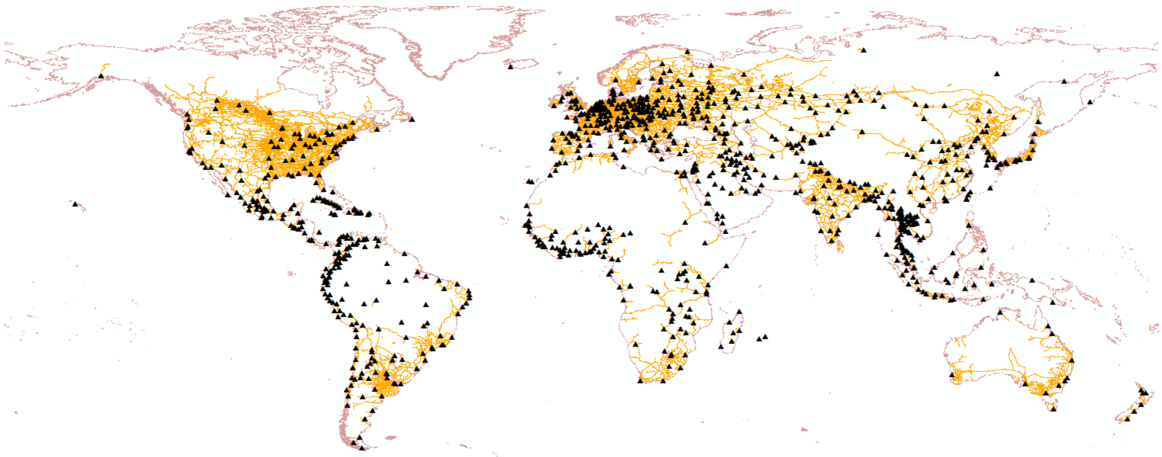

Figure A1: This figure shows the location of railroads and cities used in my analysis.

A.1 Rystad Energy: Industry climate assessment

All variable definitions are reported verbatim from Rystad Energy's Ucube.

Benign: Benign refers to a climate not putting additional wear or risk requirements on facilities and operations. It may still be hot, cold, or snow, but standard technology applies.

Arctic wear: Arctic wear refers to arctic conditions putting additional requirements on facilities and operations due to icing.

Arctic hazard: Arctic hazard refers to risks of ice bergs or ice extension threatening the integrity of facilities and production.

Harsh wear: Harsh wear refers to climatic conditions that put additional requirements on operation: particular vessels and facilities to operate in high waves, wind and currents.

Harsh hazard: Harsh hazard refers to risk of hurricanes or geohazards (such as earthquakes).

A.2 Rystad Energy: Expenditure variables

All variable definitions are reported verbatim from Rystad Energy's Ucube.

Exploration Capex: Cost incurred to find and prove hydrocarbons: seismic, wildcat and appraisal wells, general engineering costs, based on reports and budgets or modelled

Well Capex: Well capex is capitalized costs related to well construction, including drilling costs, rig lease, well completion, well stimulation, steel costs and materials.

Facility Capex: Development capex except well construction costs, includes costs to develop, install, maintain and modify surface installations and infrastructure. As reported by operators, field partners or officals, or modelled. Distribution of costs over time is largely modelled, one exception is Norway, where all historic costs are as reported in national budgets.

Opex = Transportation Opex + Production Opex + SG&A Opex

Transortation Opex: Represents the cost of brining the oil and gas from the production site/processing plant to the pricing point (only upstream transportation). The category includes transport fees and blending costs.

Production Opex: Represents operational expenses direcly related to the production activity. The category includes materials, tools, maintenance, equipment lease costs and operation related salaries. Depreciation and other non-cash items are not included

SG&A Opex: Represent operation expenses not directly associated with field operations. The category includes administrative staff costs, office leases, related benefits (stocks and stock option plans) and professional expensenes (legal, consulting, insurance). Only E&P (Exploration and Production) related SG&A are included.

B Quantification procedure

I start by the following relation: 5.8 pp is equal to the density of discoveries in economically developed areas minus the density of discoveries in economically undeveloped areas, that is:

$$0.058 = \mathbb{P}(DISC|EDC = 1) - \mathbb{P}(DISC|EDC = 0) \tag{3}$$

We can now use the law of total probabilities to back out the conditional probabilities:

$$\mathbb{P}(DISC) = \mathbb{P}(DISC|EDC = 1)\mathbb{P}(EDC = 1) + \mathbb{P}(DISC|EDC = 0)\mathbb{P}(EDC = 0) \tag{4}$$

$\mathbb{P}(DISC)$ is know, and is equal to 0.0296. The same goes for $\mathbb{P}(EDC = 1)$, which is equal to 0.401, and $\mathbb{P}(EDC = 0)$ which is equal to (1-0.401). We are now left with two unknowns, and two equations. Solving for $\mathbb{P}(DISC|EDC = 1)$ and $\mathbb{P}(DISC|EDC = 0)$ I get 0.081 and 0.023, respectively. As we now have the conditional densities, we can use them to calculate the number of *expected* barrels discovered in an economically developed grid cells versus an undeveloped grid cell.

If we randomly draw one economically developed grid cell we expect on average to discover $\mathbb{P}(DISC|EDC = 1)$*185 MMbbl = 14.985 MMbbl. If we do the same from the pool of grid cells that are economically undeveloped we expect on average to discover $\mathbb{P}(DISC|EDC = 0)$*185 MMbbl = 4.255 MMbbl. In total there exist 97.182 grid cells that are economically developed (EDC = 1) and 145,002 grid cells that are undeveloped.

We can now identify the number of expected barrels discovered due to economic development and the underexplored subsurface potential in economically undeveloped areas.

The number of expected barrels due to economic development is then $(97,182{*}0.081 - 145,002{*}0.023) * 185 =$ 839 Gbbl. The number of underexplored subsurface potential is $145,002 * (0.081 - 0.023) = 145,002 * 0.058 =$ $1,556$ Gbbl.

Up to now I have assumed that the nevertakers and alwaystakers have the same treatment effect as the compliers, i.e. asssumed homogenous treatment effect. This assumption can be relaxed by using only the shares of compliers in economically developed and undeveloped grid cells in the quantification.

I calculate the number of nevertakers by summing over the predicted values ($\widehat{EDC}$) in grid cells that are economically undeveloped ($EDC = 0$) and substract it from the number of grid cells that are economically undeveloped. Then I do the same for the grid cells that are economically developed ($(EDC = 1$):

$$145,002 - \sum_{x=1}^{145,002} (\widehat{EDC}|EDC = 0) = 145,002 - 57,179.77 = 87,822.3 \tag{5}$$

$$97,182 - \sum_{x=1}^{97,182} (\widehat{EDC}|EDC = 1) = 97,182 - 40,002.3 = 57,179.77 \tag{6}$$

The shares of nevertakers and alwaystakers in each group is then $\frac{87,822.3}{145,002} = 60.57\%$ and $\frac{57,179.77}{97.182} = 58.83\%$, respectively. This means that the shares of compliers in each groups is $1 - 0.6057 = 39.43\%$ and $1 - 0.5883 = 41.17\%$

I find the expected number of additional barrels discovered due to economic development by multiplying the share number I got when I assumed homogenous treatment effect with the share of compliers: $839*41.17\% = 350$ Gbbl. I do the same for to find the underexplored subsurface potential: $1,556 * 39.43\% = 613$ Gbbl.

C Heat map

To derive Figure 5 I start by regressing DHS_{icgb} and EDC_{icgb} on all covariates ($\mathbf{W}_{icgb}$, η_c, η_g and η_b) in two separate regressions. I then store the residuals. Next, I regress the residuals from the first regression (DHS_{icgb}) on the residuals from the second regression (EDC_{icgb}). I thereby obtain my first-stage estimate (same as in Table 1, column 5, which allows me to produce predicted values of EDC_{icgb}, noted as $\widehat{EDC}_{icgb}$. Next, I regress $DISC_{icgb}$ on the covariates and stor the predicted values. Finally, I regress $\widehat{EDC}_{icgb}$ on the residual variation of $DISC_{icgb}$ (the residuals from the previous regression), and obtain the second-stage coefficient (same as in Table 1, column 5). I use this coefficient together with $\widehat{EDC}_{icgb}$ to get the predicted values in Figure 5 ($\widehat{DISC}$). Each color represents a decile. The distribution of deciles is depicted in Figure A2.

C.1 Summary statistics

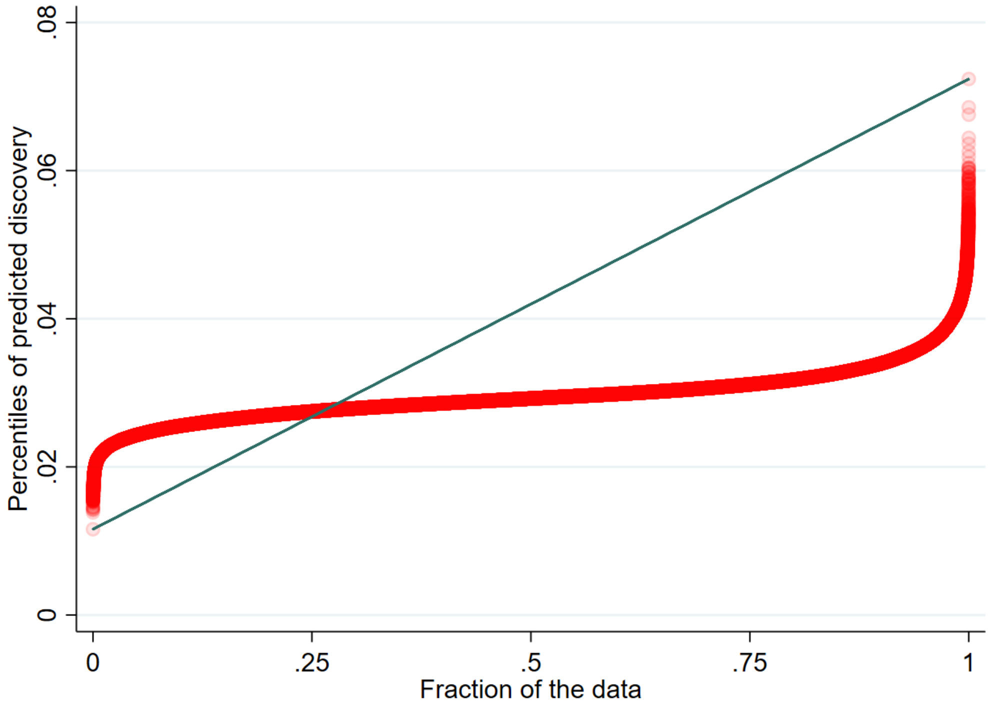
.08
.06
.04
.02
0
Percentiles of predicted discovery
0
.25
.5
.75
1
Fraction of the data

Figure A2

Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
EDC	242.184	0,401	0,49	0	1
DISC	242.184	0,03	0,169	0	1
Night light intensity	242.184	-3,357	3,119	-5,684	6,941
Ruggedness	242.184	2,781	4,852	0	95,814
Elevation	242.184	0,605	0,79	-0,187	6,169
Land suitability index	242.184	0,275	0,32	0	1
Coastal	242.184	0,097	0,296	0	1
Distance to nearest river	242.184	675,128	557,648	0	8.395,764
On river	242.184	0,014	0,119	0	1
Distance to coast	242.184	0,486	0,481	0	2,274
Temperature	242.184	10,018	13,768	-22,286	30,366
Precipitation	242.184	60,817	59,274	0,387	921,909
Distance to nearest natural harbor	242.184	567,858	477,108	0,894	2.355,688
Distance to nearest big lake	242.184	1652,471	1874,73	0	11.454,1
Growing days	242.184	139,63	99,043	0	366
Malaria Mosquito survivor index	242.184	1,921	5,289	0	38,081
Land	242.184	536,122	194,175	0,241	769,317
Absolute latitude	242.184	38,315	20,935	0,125	74,.875
Natural harbor within 25 km	242.184	0,027	0,163	0	1
Navigable river within 25 km	242.184	0,027	0,163	0	1
Big lake within 25 km	242.184	0,011	0,104	0	1

Figure A3

Descriptive Statistics: discovery specific outcomes

Variable	Obs	Mean	Std. Dev.	Min	Max
Discovery year	7.170	1.973,539	132,321	0	2.021
Discovery size (mmbbl)	7.170	2,877	2,074	0	10,653
Well capex	7.170	4,028	2,978	0	12,441
Exploration capex	7.152	2,605	1,871	0	9,385
Facility capex	7.170	3,918	2,651	0	11,872
Opex	7.170	4,47	3,096	0	13,695

Figure A4: This table shows summary statistics for all discovery specific outcomes. All variables are in *logs*, with the exception of *Discovery year*

References

Ahlvik, Lassi, Jørgen Juel Andersen, Jonas Hveding Hamang, and Torfinn Harding, "Quantifying supply-side climate policies," 2022.

Andersen, Jørgen J and Michael L Ross, "The big oil change: A closer look at the Haber–Menaldo analysis," *Comparative Political Studies*, 2014, *47* (7), 993–1021.

Andersen, Jørgen Juel, Frode Martin Nordvik, and Andrea Tesei, "Oil price shocks and conflict escalation: onshore versus offshore," *Journal of Conflict Resolution*, 2022, *66* (2), 327–356.

Andersen, Jørgen Juel, Jonas Hamang, and Michael Ross, "Declining Oil Production Leads to More Democratic Governments," *Center for global development - working paper series*, 2022.

Arezki, Rabah, Frederick van der Ploeg, and Frederik Toscani, "The shifting natural wealth of nations: The role of market orientation," *Journal of Development Economics*, 2019, *138*, 228–245.

_ , Valerie A Ramey, and Liugang Sheng, "News shocks in open economies: Evidence from giant oil discoveries," *The quarterly journal of economics*, 2017, *132* (1), 103–155.

Asheim, Geir B, Taran Fæhn, Karine Nyborg, Mads Greaker, Cathrine Hagem, Bård Harstad, Michael O Hoel, Diderik Lund, and Knut Einar Rosendahl, "The case for a supply-side climate treaty," *Science*, 2019, *365* (6451), 325–327.

Asker, John, Allan Collard-Wexler, and Jan De Loecker, "(Mis) allocation, market power, and global oil extraction," *American Economic Review*, 2019, *109* (4), 1568–1615.

Berman, Nicolas, Mathieu Couttenier, Dominic Rohner, and Mathias Thoenig, "This mine is mine! How minerals fuel conflicts in Africa," *American Economic Review*, 2017, *107* (6), 1564–1610.

Bleakley, Hoyt and Jeffrey Lin, "History and the Sizes of Cities," *American Economic Review*, 2015, *105* (5), 558–63.

Bohn, Henning and Robert T Deacon, "Ownership risk, investment, and the use of natural resources," *American Economic Review*, 2000, *90* (3), 526–549.

Bornstein, Gideon, Per Krusell, and Sergio T Rebelo, "Lags, Costs and Shocks: An Equilibrium Model of the Oil Industry," 2017.

Bright, Edward A, Amy N Rose, Marie L Urban, and Jacob McKee, "LandScan 2017 high-resolution global population data set," Technical Report, Oak Ridge National Lab.(ORNL), Oak Ridge, TN (United States) 2018.

Cassidy, Traviss, "The long-run effects of oil wealth on development: Evidence from petroleum geology," *The Economic Journal*, 2019, *129* (623), 2745–2778.

Cey, Edwin, Jordan Hanania, Ashley Sheardown, Kailyn Stenhouse, and Jason Done, "Oil formation [Online]," https://energyeducation.ca/encyclopedia/Oil$_formationcite_note-2$.2019.

Chandler, Tertius, *Four thousand years of urban growth: An historical census*, Mellen, 1987.

Cust, James and Torfinn Harding, "Institutions and the location of oil exploration," *Journal of the European Economic Association*, 2020, *18* (3), 1321–1350.

Dahl, Gordon B, Andreas Ravndal Kostøl, and Magne Mogstad, "Family welfare cultures," *The Quarterly Journal of Economics*, 2014, *129* (4), 1711–1752.

David, Paul A. and Gavin Wright, "Increasing Returns and the Genesis ofAmerican Resource Abundance," *Industrial and Corporate Change*, 1997, pp. 203–245.

Deacon, Robert T, "Deforestation and the rule of law in a cross-section of countries," *Land economics*, 1994, pp. 414–430.

der Ploeg, Frederick Van, "Natural resources: curse or blessing?," *Journal of Economic literature*, 2011, *49* (2), 366–420.

Doxsey-Whitfield, Erin, Kytt MacManus, Susana B Adamo, Linda Pistolesi, John Squires, Olena Borkovska, and Sandra R Baptista, "Taking advantage of the improved availability of census data: a first look at the gridded population of the world, version 4," *Papers in Applied Geography*, 2015, *1* (3), 226–234.

Earth, Natural, "Admin 0 – Countries," https://www.naturalearthdata.com/downloads/10m-cultural-vectors/10m-admin-0-countries/ 2017.

Ellison, Glenn and Edward L Glaeser, "The geographic concentration of industry: does natural advantage explain agglomeration?," *American Economic Review*, 1999, *89* (2), 311–316.

Elvidge, Christopher D, Daniel Ziskin, Kimberly E Baugh, Benjamin T Tuttle, Tilottama Ghosh, Dee W Pack, Edward H Erwin, and Mikhail Zhizhin, "A fifteen year record of global natural gas flaring derived from satellite data," *Energies*, 2009, *2* (3), 595–622.

_ , Kimberly E Baugh, John B Dietz, Theodore Bland, Paul C Sutton, and Herbert W Kroehl, "Radiance calibration of DMSP-OLS low-light imaging data of human settlements," *Remote Sensing of Environment*, 1999, *68* (1), 77–88.

Feldman, Maryann P, "The new economics of innovation, spillovers and agglomeration: Areview of empirical studies," *Economics of innovation and new technology*, 1999, *8* (1-2), 5–25.

Feng, MA, YAN Cunfeng, MA Dade, LE Xingfu, Chenggang HUANG, SHI Yajun, Yongshu ZHANG, and XIE Mei, "Bedrock gas reservoirs in Dongping area of Qaidam Basin, NW China," *Petroleum Exploration and Development*, 2015, *42* (3), 293–300.

Helpman, Elhanan, "The size of regions," *Topics in public economics: Theoretical and applied analysis*, 1998, pp. 33–54.

Henderson, J Vernon, "Ways to think about urban concentration: neoclassical urban systems versus the new economic geography," *International Regional Science Review*, 1996, *19* (1-2), 31–36.

_ , Tim Squires, Adam Storeygard, and David Weil, "The global distribution of economic activity: nature, history, and the role of trade," *The Quarterly Journal of Economics*, 2018, *133* (1), 357–406.

Jedwab, Remi, Edward Kerby, and Alexander Moradi, "History, path dependence and development: Evidence from colonial railways, settlers and cities in Kenya," *The Economic Journal*, 2017, *127* (603), 1467–1494.

Kiszewski, Anthony, Andrew Mellinger, Andrew Spielman, Pia Malaney, Sonia Ehrlich Sachs, and Jeffrey Sachs, "A global index representing the stability of malaria transmission," *The American journal of tropical medicine and hygiene*, 2004, *70* (5), 486–498.

Krugman, Paul, "History and industry location: the case of the manufacturing belt," *The American Economic Review*, 1991, *81* (2), 80–83.

_ , "Increasing returns and economic geography," *Journal of political economy*, 1991, *99* (3), 483–499.

_ , "First nature, second nature, and metropolitan location," *Journal of regional science*, 1993, *33* (2), 129–144.

Mehlum, Halvor, Karl Moene, and Ragnar Torvik, "Institutions and the resource curse," *The economic journal*, 2006, *116* (508), 1–20.

Modelski, George, *World cities:-3000 to 2000*, Faros 2000, 2003.

Nordhaus, William, Qazi Azam, David Corderi, Kyle Hood, Nadejda Makarova Victor, Mukhtar Mohammed, Alexandra Miltner, and Jyldyz Weiss, "The G-Econ database on gridded output: methods and data," *Yale University, New Haven*, 2006, *6.*

Nordvik, Frode Martin, "Does oil promote or prevent coups? the answer is yes," *The Economic Journal*, 2019, *129* (619), 1425–1456.

Nunn, Nathan and Diego Puga, "Ruggedness: The blessing of bad geography in Africa," *Review of Economics and Statistics*, 2012, *94* (1), 20–36.

Reba, Meredith, Femke Reitsma, and Karen C Seto, "Spatializing 6,000 years of global urbanization from 3700 BC to AD 2000," *Scientific data*, 2016, *3* (1), 1–16.

Redding, Stephen J and Matthew A Turner, "Transportation costs and the spatial organization of economic activity," *Handbook of regional and urban economics*, 2015, *5*, 1339–1398.

Ross, Michael L, "Does oil hinder democracy?," *World politics*, 2001, *53* (3), 325–361.

Sachs, Jeffrey D and Andrew Warner, "Natural resource abundance and economic growth," 1995.

Southard, John, "12.110 Sedimentary Geology," *Massachusetts Institute of Technology: MIT OpenCourseWare, http://ocw. mit. edu*, 2007.

Southgate, Douglas, Rodrigo Sierra, and Lawrence Brown, "The causes of tropical deforestation in Ecuador: A statistical analysis," *World Development*, 1991, *19* (9), 1145–1151.

Tatem, Andrew J, "WorldPop, open data for spatial demography," *Scientific data*, 2017, *4* (1), 1–4.

UCube, "UCube, Rystad Energy data bank," https://www.rystadenergy.com/energy-themes/oil–gas/upstream/u-cube/ 2020.

USGS, "Arctic Geology (geoarcst)," https://certmapper.cr.usgs.gov/data/we/ofr97470j/spatial/shape/geoarclst.shp.xml 2022.

_ , "Bedrock geology of the Arabian Peninsula and selected adjacent areas (geo2bg)," https://data.doi.gov/dataset/bedrock-geology-of-the-arabian-peninsula-and-selected-adjacent-areas-geo2bg/ 2022.

_ , "Database of the Geologic Map of North America," https://pubs.er.usgs.gov/publication/ds424 2022.

_ , "Generalized Geology of Australia and New Zealand (geo3cl)," https://data.doi.gov/dataset/generalized-geology-of-australia-and-new-zealand-geo3cl 2022.

_ , "Generalized Geology of Europe including Turkey (geo4$_2l$)," $https://www.sciencebase.gov/catalog/item/60abc880d3$

_ , "Generalized Geology of Southeast Asia (geo3bl)," https://www.sciencebase.gov/catalog/item/60abc823d34ea221ce51e5f
2022.

_ , "Generalized Geology of the Far East (geo3al)," https://www.sciencebase.gov/catalog/item/60abc7f9d34ea221ce51e5ee 2022.

_ , "Generalized Geology of the Former Soviet Union (geo1ec)," https://www.sciencebase.gov/catalog/item/60ad3d53d34e40
2022.

_ , "Geologic map of South Asia (geo8ag)," https://www.sciencebase.gov/catalog/item/60c3b89fd34e86b93897ef19 2022.

_ , "South America Geologic Map (geo6ag)," https://data.doi.gov/dataset/south-america-geologic-map-geo6ag 2022.

_ , "Surface Geology of the Caribbean Region (geo6bg)," https://www.sciencebase.gov/catalog/item/60abf6aad34ea221ce51
2022.

_ , "Surficial geology of Africa (geo7$_2ag$)," $https://www.sciencebase.gov/catalog/catalogParty/show?partyId = U.S. + G$

— , “Surficial geology of Iran (geo2cg),” https://www.sciencebase.gov/catalog/item/60ad3db2d34e4043c850f2a0 2022.

WTP, “Global Main Populated Places,” https://geonode.wfp.org/layers/geonode:vw$_{w}ld_{p}op_{m}aincities_{g}nm_{w}fp/metadata_{d}e$

— , “Global Railways,” https://geonode.wfp.org/layers/esri$_{g}n : geonode : wld_{t}rs_{r}ailways_{w}fp$2017.

Centre for Applied Macroeconomics and Commodity Prices (CAMP)
will bring together economists working on applied macroeconomic issues, with special emphasis on petroleum economics.

BI Norwegian Business School
Centre for Applied Macro - Petroleum economics (CAMP)
N-0442 Oslo

www.bi.no/camp

CAMP Working Paper Series
ISSN: 1892-2198