Determinants of Oil and Gas Investments on the Norwegian Continental Shelf

forthcoming in Energy

Martin Berntsen[1], Kristine Skjong Bøe[1], Therese Jordal[1] and Peter Molnár[2]

February 8, 2018

Abstract

This paper studies the investment decisions by oil and gas companies operating on the Norwegian Continental Shelf. We account for the heterogeneity across the fields by including field-specific variables, including geological and geographical variables. We find that the most important factors influencing the investment decisions are the size of the oil and gas reserves, geological variables, and the price of oil. The effect of oil price volatility is insignificant.

Keywords: oil exploration, investment timing, oil price

[1] Norwegian University of Science and Technology, Department of Industrial Economics and Technology Management, Trondheim, Norway

1 Introduction

Oil and gas exploration and field development have been subject to economic research for decades. In the existing literature, the main focus has been on the US and UK oil and gas data sets, because international data sets provide much less detailed information (Bøe et al., 2018). In this paper, we investigate the relation between investment activity and economic and field-specific variables on the Norwegian Continental Shelf from 1967 to 2015.

The decision to develop a natural resource asset such as oil and gas fields can be considered to be an irreversible investment. The process involves three separate but closely interrelated activities: exploration, development, and extraction. From a real options point of view, having a license for exploration drilling may be seen as owning an option. When a drilling decision is made, the exploration option is exercised and a development option is acquired. In most cases, the development decision requires a large investment, which makes this decision particularly interesting to investigate. Hurn and Wright (1994) argue that once the exploration stage is completed, economic factors are the main determinants of the development decision. Since new information should continuously affect an investment decision, we allow economic variables to vary over the appraisal time. The aim of this paper is to examine how the investment decision is influenced by the oil price and the oil price uncertainty while also accounting for field-specific variables such as the size of the oil and gas reserves. We study the appraisal lag (the time elapsed from discovery to develop-

[2] University of Stavanger, UiS Business School, Stavanger, Norway and University of Economics, Department of Finance and Accounting, Prague, Czech Republic; corresponding author, email: peto.molnar@gmail.com

ment approval) for this purpose, using it as a proxy for the time spent waiting before investing.

Extensive empirical literature examining the uncertainty-investment relationship exists and a large proportion of these studies focus on natural resource industries. Moel and Tufano (2002) examine the impact of resource price volatility on gold mine openings and closings and find that real options models are useful for describing the decisions of mines by incorporating the effect of uncertainty, but claim such models fail to capture aspects of firm-level decision making. Dunne and Mu (2010) report the impact of price uncertainty on petroleum refinery investments and provide evidence of the wait-and-see response of investment to a rise in uncertainty, conforming to the standard predictions of the real options theory. Kellogg (2014) investigates well-level data on onshore drilling activity in Texas and find that firms reduce drilling activity under increased expected volatility. His findings agree with the predictions of the basic real options theory. However, Bar-Ilan and Strange (1996) provide a real options model in which the effect of price uncertainty on investment is weakened or even reversed in some cases. Similarly, Sarkar (2000) develops a model in order to demonstrate that an increase in uncertainty does not always have an inhibiting effect on investment. In some cases, particularly for low-risk and low-growth firms, he argues a higher level of uncertainty might actually have a positive effect on investment. Altogether, the real options theory provides mixed predictions regarding the impact of uncertainty on investments.

Duration analysis has become increasingly popular for testing the predictions of the real options theory and to generally analyze investment behavior empirically.

Dunne and Mu (2010) apply hazard models in their study on petroleum refineries. The framework has also been applied to analyze oil and gas field developments. Using duration analysis, Favero et al. (1994) examine empirically how oil price and oil price uncertainty affect the decision process on the United Kingdom Continental Shelf. They find that both oil price and volatility are significant determinants of the development decision. Hurn and Wright (1994) extend the work of Favero et al. (1994) by including monthly observations for each field. Hence, the oil price and volatility vary from discovery to approval in their study. Furthermore, their data set is extended to include assets discovered but not yet approved by the end of the study period. They conclude that the larger the known oil reserves and the higher the oil price – the shorter the period before oil companies decide to invest. However, they do not find a significant impact of the oil price volatility demonstrated by Favero et al. (1994).

A set of papers also examine the technical complexities of the offshore oil and gas industry. Aadnoy (2010) provides a qualitative evaluation of the relation between physical well characteristics and drilling speed. According to his study, considerable improvement has been made in equipment and technology in recent decades, but at the same time, operators are facing wells that are more difficult to drill. Osmundsen et al. (2012) complement Aadnoy (2010) by using an econometric approach on a large data set of individual exploration wells on the Norwegian Continental Shelf (NCS) to analyze the effect of different types of learning and experience. Data set utilized in this paper, based on a detailed database from the Norwegian Petroleum Directorate covering the 40-year history of the Norwegian Continental Shelf, was first employed in Mohn and Osmundsen (2008) to investigate exploration behavior.

We apply duration analysis on this data set in order to investigate the relation between the investment decision and economic and field-specific variables. We find that the size of oil and gas reserves are the most important explanatory factors of the investment activity on the Norwegian Continental Shelf. Oil companies are more likely to invest earlier when the estimated reserves of a discovery are larger and the oil price is high. We do not find any significant effect of oil price volatility on the investment decision.

The remainder of this paper is organized as follows. Section 2 describes the data, section 3 outlines the methodology used in our analysis, and section 4 presents our results and discusses the empirical findings. The last section summarizes and concludes.

2 Data

The data used in this paper is compiled from two sources. The primary source is a global upstream oil and gas database called UCube, owned and operated by the Norwegian oil and gas consultancy and research firm Rystad Energy. The data consists of 337 assets on the Norwegian Continental Shelf. From this total, 90 fields are currently producing, 18 have been producing but are now abandoned, 16 are under development, and the remaining 213 are discoveries not yet approved for development. The study period extends from the first recorded discovery on the Norwegian Continental Shelf in July 1967, when oil was found in the Balder field, to October

2015. The data is proprietary and have been released only for the purposes of this paper.

Our second source of data is the petroleum databases of the Norwegian Petroleum Directorate. This database consists of data on all wells drilled on the Norwegian Continental Shelf since the first wells were drilled in the late 1960s. This well level data has been aggregated to the field level in the Rystad Energy database by matching the well identifiers to the internal asset identifiers used by Rystad Energy. This has enabled us to calculate the number of appraisal wells within a field, as well as to obtain the geological age of the different fields.[1]

The final data set consists of 280 assets, 107 approved fields, and 172 discoveries. We have eliminated 57 assets, all discoveries, due to missing data on variables. This is particularly the case for discoveries having a highly uncertain resource size, hence Rystad does not provide a resource estimate. The field data together with the crude oil time series and oil price volatility constitutes our data set.

2.1 Crude oil price and volatility

The price of crude oil depends on its quality, which varies across the producing regions. Norwegian oils are mainly light, sweet blends as measured by the gravity and the sulphur level, respectively[2]. The most important blend is Ekofisk, which has

[1]The wells in the Norwegian Petroleum Directorate database are recorded with the oldest penetrated geological age of the rock. These have been grouped into four different categories according to closeness in age. When aggregated to field level, the final geological category assigned to the field has been the most frequently recorded category among the included wells.

[2]Gravity is a measure of how heavy or light a petroleum liquid is compared to water: if the gravity is greater than 10, it is lighter and floats on water; if less than 10, it is heavier and sinks. Furthermore, when the total sulfur level in the oil is more than 0.5% the oil is called "sour", and is generally sold with a discount.

a gravity of 37.8 and 0.3% sulfur content and is very similar to the UK's Brent. Brent Crude is a major trading classification of sweet light crude oil that serves as a major benchmark price for purchases of oil worldwide. Brent is the leading global oil price benchmark for Atlantic basin crude oils, which comprise all Norwegian reserves.

The Brent Crude price data used in our analysis are retrieved from Reuter's EcoWin Pro database. The data series covers the period from 1970 to 2015. For the years prior to 1970, we assume the 1970 Brent Crude price. Oil was not actively traded before the end of the 1970s. Prior to this, prices were based on tariffs set by the Organization of Petroleum Exporting Countries (OPEC). Hence, there are periods spanning several months prior to 1978 over which the price remains constant. Between 1978 and 1986, prices are available only on a monthly basis but since 1987 onward, prices have been recorded daily. The Brent Crude price over the studied period is plotted in Figure 1. Note that several price shocks occurred, with the oversupply in 2014 and the financial crisis in 2008 being the most recent.

The relevant oil price in the field development decision is the future expected oil price. Oil futures prices could be used as a predictor for the future spot price of oil. However, Alquist and Kilian (2010) show that oil futures prices tend to be a less accurate predictor than the current spot price. Additionally, they consider the use of long-term futures prices and conclude that the low liquidity limits the practical use of these contracts as a predictor for the long-term spot price. As the spot price is an adequate predictor for the expected future oil price and is easily obtained, it is often used in practice. Moreover, oil futures prices are not available for a large part of the period we study. Therefore, we use the oil spot price in our analysis.

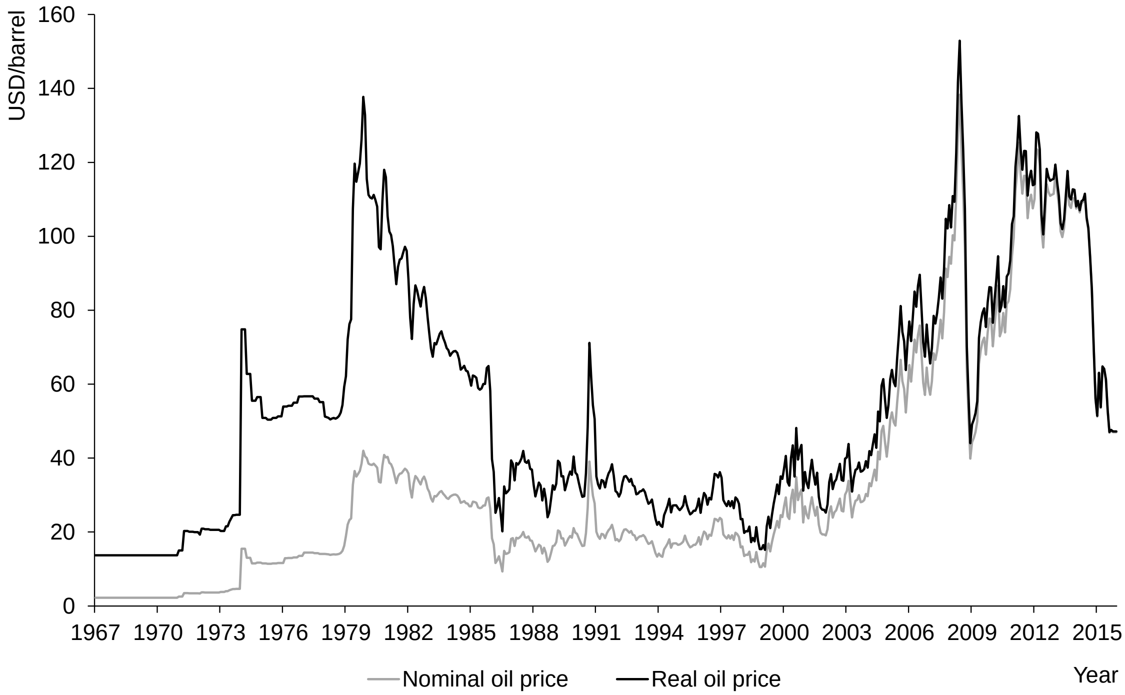


Figure 1: The Brent Crude price

A moving average of the squared returns from the last twelve months is used as a proxy for the historical volatility. The resulting monthly time series is plotted in Figure 2. Ideally, a measure of the expected future volatility, such as implied volatility calculated from the prices of options on oil futures, would have been used. However, such data are not available for the complete time period under consideration.

2.2 Field variables

The Norwegian Continental Shelf is a relatively mature hydrocarbon province with the first discovery made in 1967. A number of discoveries were made in subsequent years, which laid the foundations for a new industry in Norway. In 2014, total

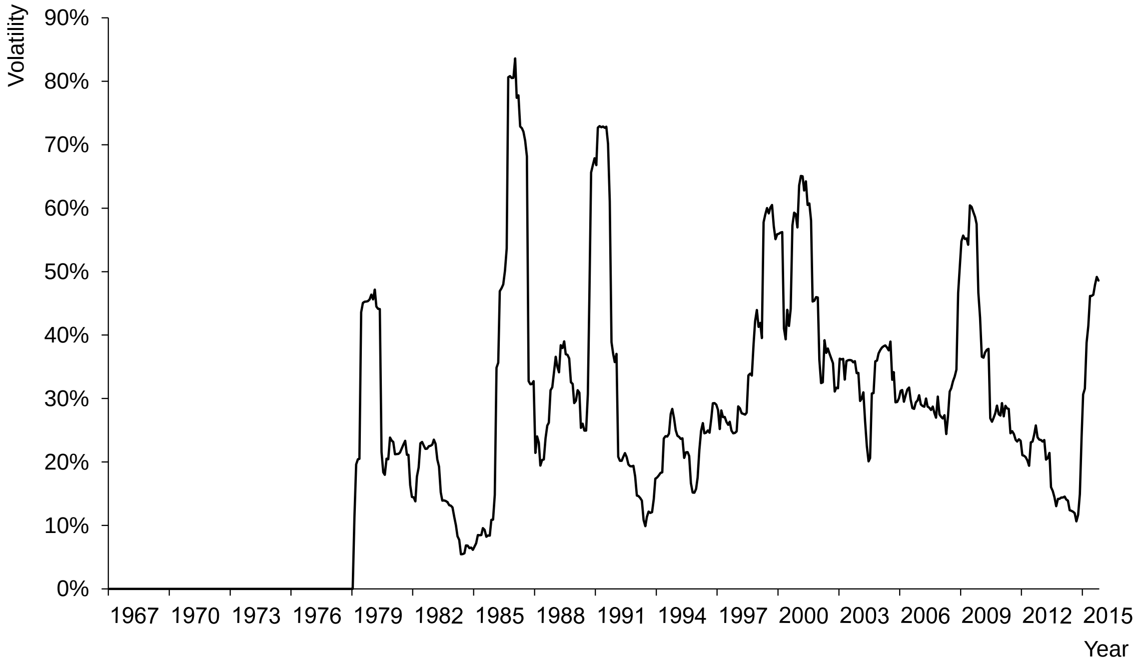


Figure 2: Oil price volatility

petroleum production was approximately four million barrels of oil equivalents per day (mmboepd), making Norway one of the largest oil exporting countries in the international oil market. All of Norway's reserves are located offshore on the Norwegian Continental Shelf, which can be divided into three main areas; the North Sea, the Norwegian Sea, and the Barents Sea. Most of the current Norwegian production is located in the central and northern areas of the North Sea. The data set used in our study includes the following field specific variables:

- *Discovery date* is the date that the field was discovered.

- *Approval date* is the date that the field received its approval.[3]
- *Location* is the area in which the field is located; North Sea, Norwegian Sea, or Barents Sea.
- *Oil reserves* is the recoverable crude oil reserves measured in mmboe.
- *Gas reserves* is recoverable gas reserves. In addition, natural gas liquids and condensates are included in this variable, all measured in mmboe.
- *Reservoir pressure* is the average recorded pressure in the reservoir measured in bar.
- *Reservoir depth* is the depth from the sea bottom to the reservoir measured in meters.
- *Appraisal wells* is the number of wells drilled in order to evaluate a discovery, between the time of discovery to the time of approval.
- *Block extension* indicates whether a field is located within one block or extends into several blocks.[4] This variable gives the number of blocks a field extends into.

[3]Before oil companies can develop a discovered field, the Norwegian authorities must approve a Plan for Development and Operation of the petroleum deposit. Ideally, the date on which the decision to develop a field would have been used. However, this date is not available. The approach in this study, as in previous studies, is to use the date that an oil company receives approval from the government to develop a field. The time lag from the discovery date to the approval date is an approximation of the time oil companies spend considering whether to invest. This includes the time spent by the government reviewing the development application although we assume this additional time lag is approximately the same for all applications in such a way that it cancels out across different appraisal lags.

[4]An exploration block is a large area of land awarded to drilling and exploration companies by a country's government.

- *Geological age* is the main geological age of the rock penetrated in the development of the field.

The approval date and discovery date are used to calculate the appraisal lag, which is simply the time difference between them, denoted in months. The appraisal lags in the data set range from 6 months to 498 months (42 years) for approved assets, while for discovered but unapproved assets, the appraisal lags vary from 4 months to 518 months (43 years). A large part of the assets is approved less than 200 months (17 years) after discovery. Figure 3 shows the distribution of the time lag from discovery to development for the assets that have been approved. For the assets not yet approved, the figure shows the distribution of the time lag from discovery until October 2015.

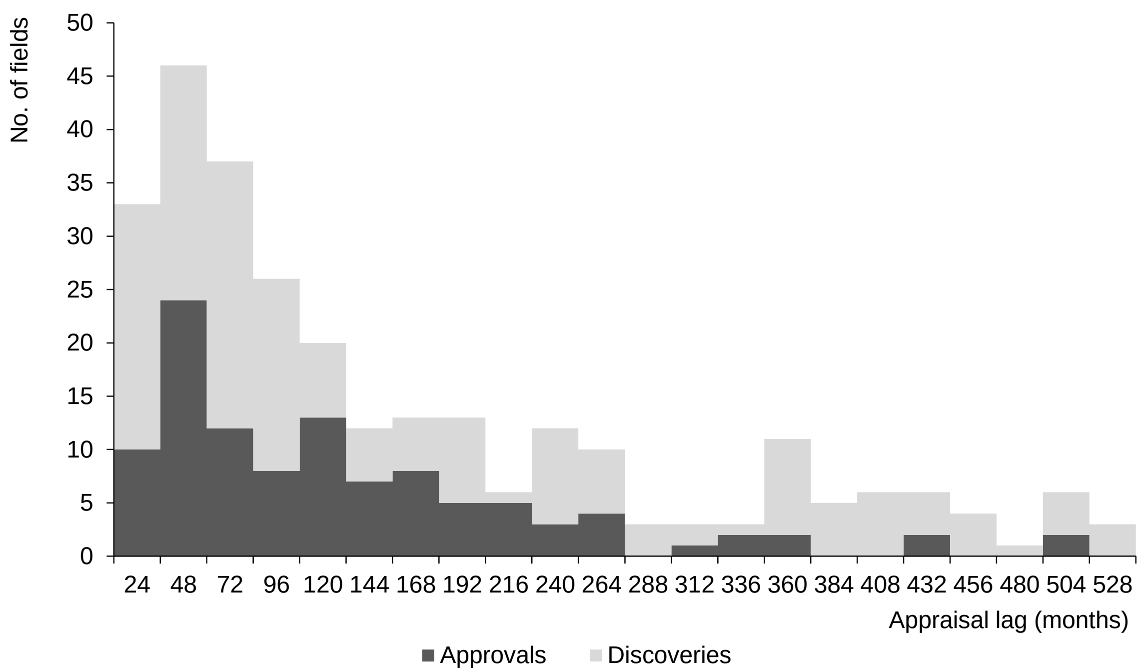


Figure 3: Appraisal lag for approved (dark) and censored (light) assets

The data set has a larger number of discovered fields than approved fields; for some of the discovered assets, development has not yet commenced, while others will probably never be developed. There has been a substantial increase in the number of discoveries on the Norwegian Continental Shelf during the last decade as well as in the number of approvals. This can be seen in Figure 4. Researchers and industry experts attribute this evolvement to an increasing oil price and a favorable regulatory environment[5] (Mohn and Osmundsen, 2008).

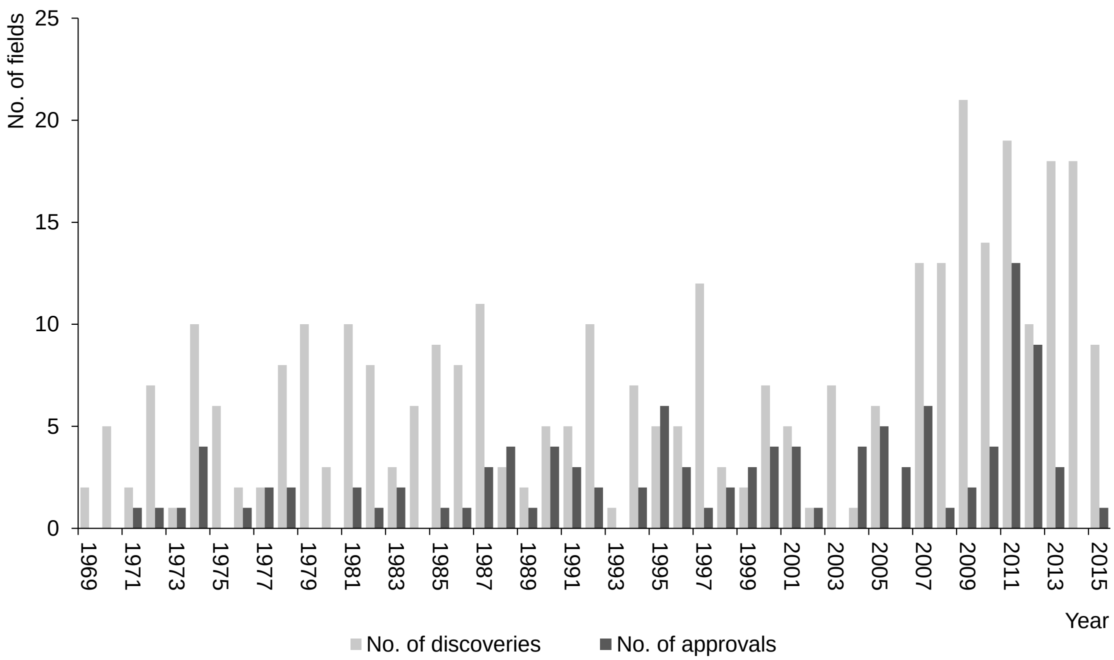


Figure 4: Number of assets discovered and approved annually

The distribution of total reserves including both crude oil, gas, and other reserves,

[5]All exploration costs are deductible and may be offset against profits from production, and any net losses may be carried forward. From the fiscal year 2005 onward, companies have been able to claim an annual cash refund of the tax value of direct and indirect exploration costs before ordinary petroleum tax and special tax, amounting to 78% of such costs.

is highly right-skewed, meaning that most of the assets in the data set are relatively small. This can be seen in Figure 5, also noting that a large part of the reserves with resources less than 100 mmboe has not been approved for development while larger fields have mostly been approved for development.

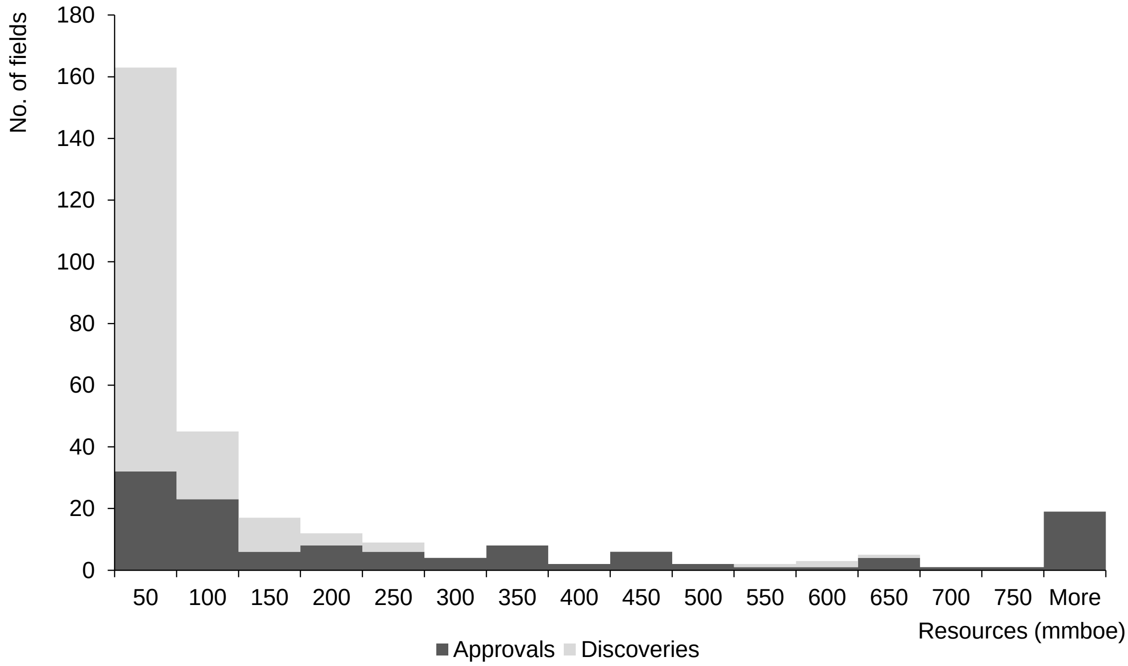


Figure 5: Number of assets discovered and approved annually from 1969 to October 2015

2.3 Model variables and descriptive statistics

We separate the variables into two main categories; economic variables and field-specific variables. Economic variables include the real price of Brent Crude oil and the oil price volatility as described in the previous section. The *Oil price* variable is the nominal price of oil adjusted for inflation.[6] The historical volatility measure, *Volatility*, is also included. Both *Oil price* and *Volatility* are recorded at a monthly frequency.

Field-specific variables are included in an effort to eliminate heterogeneity across the assets. The approach is similar to that of Hurn and Wright (1994), who emphasize that the different characteristics of the assets should be modeled as unobserved heterogeneity inherent in the investment problem. Both numerical and categorical variables are included. The set of numerical variables contains *Oil reserves*, *Gas reserves*, *Reservoir pressure*, *Nearby approvals* and *Block extension.* To obtain results comparable to those of Hurn and Wright (1994), we also include *Appraisal wells.* The set of categorical variables include *Location* and *Geological age.*

In measuring the size of the reserves, we use the two variables *Oil reserves* and *Gas reserves.* These variables are the natural logarithm of the recoverable resources in mmboe. Readers are referred to Appendix A for the mathematical derivation of the best fitting functional form of the reserve variables.

By using *Reservoir pressure* we seek to capture the differences in technological complexity. Extracting oil and gas under higher pressure often requires more advanced and robust equipment.

The variable *Nearby approvals* is time-varying and includes the number of approvals in a nearby area within a window of three years. Developed fields in proximity to discoveries are known to improve field economics as existing infrastructure may be exploited, and thus contribute to a shorter appraisal lag.

[6] This adjustment has been made using the approach employed by the US Energy Information Agency in which the price is adjusted with respect to the American Consumer Price Index.

The *Block extension* variable gives the number of blocks a field extends into. This is included in order to capture extra costs and complexities that may be induced if a field spans into several blocks.[7]

There are three categories represented in the *Location* variable; the North Sea, the Norwegian Sea, and the Barents Sea. Petroleum activities on the Norwegian Continental Shelf began in the North Sea and have gradually expanded northwards. Since 1980, there has also been activity in the Norwegian Sea and the Barents Sea. The North Sea is still the main area for Norwegian petroleum extraction, with 64 fields producing in 2015. Additionally, there are 16 producing fields in the Norwegian Sea, and 2 (Snohvit and Goliat) in the Barents Sea. The Norwegian Sea has large gas reserves, and is less mature and less thoroughly explored than the North Sea. The Barents Sea has a rougher climate than the North Sea and Norwegian Sea, has large unexplored areas, and limited existing infrastructure. We include *Location* to account for the differences between the three areas.

The variable *Geological age* includes the following categories: Paleogene and Neogene (Category 1), Cretaceous (Category 2), Triassic and Jurassic (Category 3), and Silurian, Devonian, Carboniferous, and Permian (Category 4). This is included in order to capture differences in complexity for oil and gas extraction due to differences in the rock type from the different geological ages. Approximately 60% of the assets

[7]Blocks may be owned by different companies, which has been shown to introduce additional frictions in the development stage of oil and gas projects on the Norwegian Continental Shelf. The firms owning the rights to each block must negotiate the share of the production that will be attributed to each company when the field comes on stream. If the companies are not able to agree, additional regulatory authorities must be involved, therefore contributing to further costs and duration. The development of the Johan Sverdrup field is a recent example. The field consists of multiple discoveries owned by Statoil, Maersk, Det Norske Oljeselskap and Petoro.

fall into the categories of Triassic and Jurassic (Category 3), while Paleogene and Neogene (Category 1) have the largest ratio of approvals to discoveries.

Table 1 provides a summary of all model variables included in our analysis. Descriptive statistics for the numerical variables are reported in Table 2 from which we can observe a mean reserve estimate for oil and gas of approximately 135 mmboe and 118 mmboe, respectively. These numbers are considerably smaller than the size of the largest reservoir, which is the result of having a data set containing many small fields. The high standard deviation, compared to the mean, certifies this.

Table 1: Model variables summary

Variable name	**Variable type**	**Description**
Oil price	Numerical	The real time-varying oil price
Volatility	Numerical	The time-varying oil price volatility
Oil reserves	Numerical	The logarithm of the estimated oil reserves
Gas reserves	Numerical	The logarithm of the estimated gas reserves, including natural gas liquids and condensates
Reservoir pressure	Numerical	The pressure of the reservoir
Nearby approvals	Numerical	The number of approvals in a nearby area within a window of three years
Block extension	Numerical	The number of blocks a field extends into
Location	Categorical	The area in which the field is located: North Sea, Norwegian Sea or Barents Sea
Geological age	Categorical	The main geological age of the rock penetrated in the development of the field: Paleogene and Neogene (Category 1), Cretaceous (Category 2), Triassic and Jurassic (Category 3), Silurian, Devonian, Carboniferous and Permian (Category 4)

Table 2: Summary statistics

Variable	Mean	SD	Min.	Max.
Oil reserves	2.67	2.10	-0.24	8.20
Gas reserves	2.37	2.28	-4.99	9.14
Appraisal wells	1.47	3.50	0.00	29.00
Block extension	1.82	1.10	1.00	6.00
Reservoir pressure	350.79	161.45	30.00	911.00
Nominal oil price	33.17	31.03	2.23	138.29
Real oil price	52.99	32.12	13.68	152.90

3 Methodology

In this section, the procedures and models applied to analyze and explain the investment decision for oil companies on the NCS are presented. Duration analysis is the main framework applied. Duration analysis was developed for investigating data from a well-defined time origin until an end point at which a particular event occurs.

In duration regression, the dependent variable is a period of time: a duration. We will present the theoretical framework of duration analysis in a way adapted to our study in which the study time is the calendar time during which a firm decides to develop an asset or not. There are several reasons why duration data are not amenable to standard statistical procedures. Firstly, the data are generally not symmetrically distributed. A normal distribution is thus not a fair assumption. Secondly, survival times are frequently censored, meaning that the event of interest, the end-point, may not be observed. Right-censored data is the most common type of censored data. In our study, right censoring means that a firm has not invested in the development of a field by the end of the study time in that the event of interest has not occurred. It is the only type of censoring relevant to our study. When censoring is present, the basic analyses of finding standard mean and median time to event occurrence using traditional methods fail. In these cases, the failure time is known only to be greater than the end time, which makes the estimate of the mean biased downward. For the median, the standard method for calculation is to order the observations and report the middle data point. In the presence of censoring, this ordering is impossible to obtain and therefore, duration analysis procedures must be applied (Collett, 2015).

Duration analysis can be non-parametric, semi-parametric or parametric, depending on assumptions about the survival function. We apply semi-parametric and non-parametric procedures in our analysis.

3.1 Non-parametric analysis

In duration analysis, the initial step is to present numerical or graphical summaries of the durations of objects. The survival function and the hazard function are of particular interest when summarizing and analyzing as they do not require any specific assumptions about the underlying distribution of the durations. We will now define these, and also review methods for estimating them from the duration data.

The dependent variable in duration analysis, the duration, is assumed to have a continuous probability density function *f(t)*. We denote the associated cumulative distribution function as *F(t)*. The probability that the duration time will be *less than* t is given by

$$F(t) = Prob(T \leq t) = \int_0^t f(u)du. \tag{1}$$

The survival function is defined as the probability that the duration will be *at least*

t:

$$S(t) = 1 - F(t) = Prob(T \geq t). \tag{2}$$

The Kaplan-Meier estimate is the most important and widely used estimate of the survival function (Collett, 2015). One assumes that the event of interest occurs independently over the sample. Then, the estimated survival function at any time in the interval from t_k to t_{k+1}, will be the estimated probability of surviving beyond t_k. This is the same as the probability of surviving through the interval from t_k to t_{k+1}, and all preceding intervals.

The Kaplan-Meier estimate of the survival function is specified as

$$\hat{S}(t) = \prod_{j=1}^{k} (\frac{n_j - d_j}{n_j}). \tag{3}$$

The Kaplan-Meier estimate can be used to calculate the median of duration data. The median is defined as the time t at which 50% of subjects are expected to survive, that is having $S(\tilde{\mu}_T) = 0.5$.

The hazard rate is the probability that an object will experience the event of interest at time t conditional on not having experienced the event already. In our case, it is the probability that a firm will invest at time t given that it has not yet invested. The hazard rate is thus the instantaneous rate of experiencing an event for an object surviving to time t.

To express this more formally, consider the probability that the random variable associated with an object's survival time T, lies between t and $t + \delta t$, conditional on T being greater than or equal to t. The hazard function *h(t)* is then the limiting value of this probability divided by the time interval δt, as δt tends to zero. The hazard rate can be defined as

$$h(t) = \frac{f(t)}{S(t)} \tag{4}$$

. It may be estimated from ungrouped data by taking the ratio of the number of events at a given time to the number of objects at risk at that time. If there are d_j events occurring at the jth event time, the hazard function on the interval from t_j to t_{j+1} can be estimated by

$$\hat{h}(t) = \frac{d_j}{n_j \tau_j}. \tag{5}$$

The log-rank test is useful for comparing the survival experience of different groups in categorical variables. The null hypothesis is that there is no difference in the survival rate between them. Consider r groups, and assume that in all groups combined, there are k distinct failure times. At failure time t_j one assumes that there are n_j subjects at risk, of which d_j fail and $n_j - d_j$ survive. The log-rank test is then computed by constructing, at each of the k distinct failure times, an r x 2 contingency table and then combining the results from these k tables. The expected number of failures in group i at time t_j is $E_{ij} = n_{ij} d_j / n_j$. The chi-squared test statistic with $r - 1$ degrees of freedom is calculated as

$$u^T V^{-1} u \tag{6}$$

using the row vector

$$u^T = \sum_{j=1}^{k}(d_{1j} - E_{1j}, ..., d_{rj} - E_{rj}) \tag{7}$$

and the variance of the log-rank statistic V, where the individual elements are calculated as the following:

$$V_{il} = \sum_{j=1}^{k} \frac{n_{ij}d_j(n_j - d_j)}{n_j(n_j - 1)}. \tag{8}$$

In (8), $i = 1, \ldots, r$; $l = 1, \ldots, r$; and $\delta_i \cdot l = 1$ if $i = l$ and 0 otherwise.

3.2 Semi-parametric analysis: Cox hazard model

The fairly simple non-parametric procedures above are not able to analyze the effect of different explanatory variables on the duration. For this purpose, we will use the Cox (1992) regression model with several covariates. This model has the following form

$$h(t, x, \beta) = h_0(t)r(x, \beta), \tag{9}$$

in which the hazard is the product of two functions. Such models are often referred to as proportional hazard models because they are multiplicatively related, i.e. their ratio is constant over the survival time. In (9), the function $h_0(t)$ characterizes how the hazard function changes as a function of the duration time, and is often called the *baseline hazard*. The other function, $r(x, \beta)$ characterizes how the hazard function changes as a function of the subject covariates.

Duration models that have a fully parametric regression structure but leave their dependence on time (the baseline hazard) unspecified are called semi-parametric regression models. The Cox proportional hazard model is the most widely used semi-parametric model. By using this model, there is no need to assume a particular probability distribution for the survival times. As a result, the hazard function is not restricted to a specific functional form and the model has flexibility and widespread applicability. The model may be extended to include time-dependent variables as well as variables that are constant over the time period. Time-dependent variables are simply defined as variables whose value change over time. There are two types of time-dependent variables; internal variables and external variables. While internal variables relate to a particular subject in the study, and can therefore only be measured before the asset experiences an event, external variables may still be recorded after an event and may be totally independent of any particular asset. In our study, we only encounter external time-varying covariates. In the case where time-varying covariates are encountered, the hazard ratio also becomes a function of time. The log hazard ratio is no longer constant and we, therefore, no longer have a proportional hazard model (Collett, 2015).

Cox (1992) was the first to propose a parametrization in (9) of $r(x, \beta) = e^{x\beta}$ so that the *cox proportional hazard model* is defined as

$$h(t, x, \beta) = h_0(t)e^{x\beta}, \tag{10}$$

where x is a vector of covariates and β is a vector of parameters. This parametrization may be used because the relative hazard cannot be negative. Furthermore, it is easy to use for the cases where we are only interested in how the covariates shift the hazard function. Then the estimation of h_0 is no longer necessary and with cox, a partial likelihood estimator for β may be obtained that does not require estimating h_0.

The model may be generalized to the situation in which some of the explanatory variables are time-varying. We let $x_{ji}(t)$ denote the value of the jth explanatory variable at time t, in the ith asset. The adjusted model is specified as follows:

$$h_i(t) = \exp\left\{ \sum_{j=1}^{p} \beta_j x_{ji}(t) \right\} h_o(t), \tag{11}$$

where the values of the variables $x_{ji}(t)$ depend on the time t, therefore the relative hazard $h_i(t)h_0(t)$ is also time-dependent. The interpretation of the β-parameters in this model, considering the ratio of the hazard functions at time t for two assets, say the rth and the sth asset, is given by

$$\frac{h_r(t)}{h_s(t)} = \exp\Big[\beta_1\big\{x_{r1}(t) - x_{s1}(t) + ... + \beta_p\big\{x_{rp}(t) - x_{sp}(t)\big\}\Big]. \tag{12}$$

The coefficient $\beta_j, j = 1, 2, ..., p$, can be interpreted as the log hazard ratio for two assets whose value of the jth explanatory variable at time t differs by one unit, with the assets having the same values as all the other variables at the time (Collett, 2015).

4 Results

Our results are presented and discussed in this section. Duration analysis is the main framework applied and is used to find determinants of the appraisal lag. This enables us to incorporate variations over time in the explanatory variables.

4.1 Nonparametric duration analysis

Figure 6 displays the estimated Kaplan-Meier curve. The graph shows the estimated survival rate at each time t, which can be interpreted as the proportion of discovered fields that is unapproved from discovery (time 0) until time t. Since none of the discovered fields have been approved immediately after the discovery, the curve begins at 1.00. The survival rate decreases as a stepwise function.

Looking at Figure 6 we can observe that at time $t = 240$ months, approximately 50% of the assets have been approved. Therefore, this point is the median of the data set and its exact value is 242 months. The estimated mean is 288 months, but since the largest observed duration is censored, the mean is underestimated.

To compare the survival time of the different categories in *Location* and *Geological age* we calculate the Kaplan-Meier estimate for each variable grouping by these categories. Looking at Figure 7a, we observe that North Sea assets generally have been approved faster than both the Norwegian Sea and the Barents Sea assets. The Barents Sea has clearly had the slowest development. The likely reasons for this observation are the high costs of developing the required infrastructure, as well as the rough climate and the political aspects involved. From Figure 7b we can

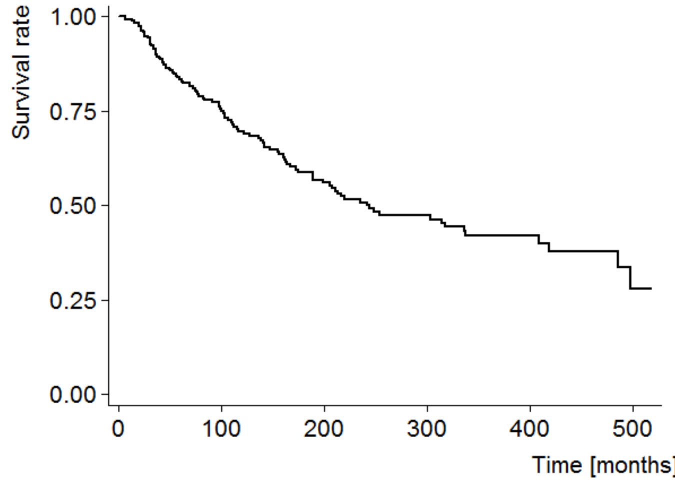
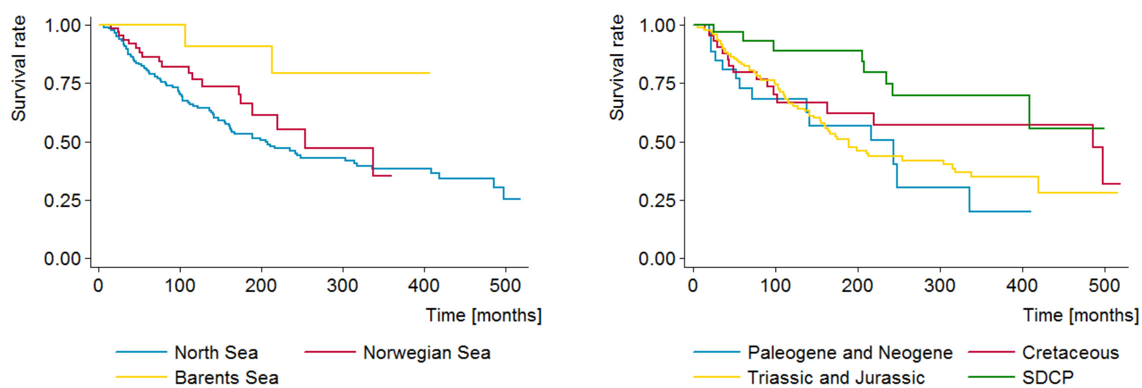


Figure 6: Kaplan-Meier estimate

observe that Silurian, Devonian, Carboniferous, and Permian (SDCP), the oldest of the four geological age categories, have a higher survival rate than the other categories meaning that oil companies wait longer before deciding to develop the fields dominated by this particular rock.



(a) Assets split on geographic area

(b) Assets split on geologic time age

Figure 7: Kaplan-Meier estimates for categorical variables

The smoothed hazard estimate is displayed in Figure 8.[8] Hess et al. (1999) found that the kernel-based hazard function estimators usually perform poorly if there are fewer than 10 subjects at risk. We limit the graphing range to $t = 350$ months as this seems to be the most appropriate cut off for our data set. The hazard rate is approximately 0.26% at $t = 50$ months and increases to 0.29% at $t = 100$ months. After this point of time, the hazard rate decreases towards 0.125% at $t = 350$, which is the end point of time. From these results, we conclude that the decision to develop an asset or not is mostly made within the first 200 months (17 years) after discovery.

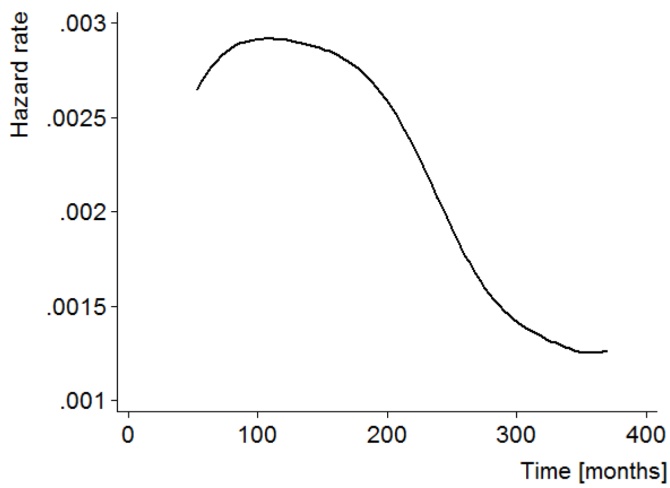


Figure 8: Hazard rate of the Norwegian continental shelf data

[8]Gaussian kernel smoothing with a width of 45 months is used to obtain these results, which requires averaging values over a moving data window. At the endpoints of the plotting range, these windows contain insufficient data for accurate estimation, and so these results are said to contain *boundary bias* and are therefore not plotted in the graph.

4.2 Cox proportional hazard regression

In order to build a complete Cox proportional hazard regression model, it is common practice to start by performing univariate regressions. This involves regressing the duration of each of the suggested continuous explanatory variables separately, in order to see whether they are significant. The results of the univariate analyses are displayed in the first column of Table 3.

Both oil and gas resource variables are significant, but it is conspicuous that neither the oil price nor the volatility appear to be so. Since economic theory suggests the oil price should play an essential role in these kinds of investment decisions, and also because several related empirical studies conclude this is the case, we do, however, keep this variable for further testing. Looking at the investment decision from the real options point of view, the price uncertainty could also play a role, as an increase in volatility would increase the value of the investment option. The initial test fails to confirm this.

Estimated Cox hazard rates models, both univariate and multivariate, are presented in Table 3. Table 3 displays the hazard ratios for the variables included. When larger than one, the hazard ratio indicates a positive effect of the covariate on the probability of a field being developed.

Overall, the model estimations indicate a significant positive effect of oil price on the hazard rate. As an example, an interpretation of the hazard rate in Model 5 yields that for a \$1 increase in the price of oil, there is a 0.67% increase in the probability of investing from one month to the next. Assuming a linear relationship between the price of oil and the hazard rate, a \$10 increase in the price of oil makes

*Table 3: Proportional Cox hazard estimates of appraisal duration, in which ** indicates 1% significance and * indicates 5% significance. The first column presents estimates from several univariate specifications.*

		Univariate	Multivariate specification				
Numerical variables			**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Oil price		1.002	1.025*	1.006	1.006*	1.006	1.007**
Volatility		0.727				1.302	1.305
Oil reserves		1.426**	1.423**	1.314**	1.370**	1.364**	1.370**
Gas reserves		1.232**	1.224**	1.177**	1.258**	1.261**	1.260**
Appraisal wells		1.073**			0.952	0.954	0.952
Block extension		0.876			0.814	0.821	0.816
Reservior pressure		0.999*			0.998**	0.998**	0.998**
Nearby approvals		1.096				1.050	
Categorical variables							
Location	1	-		-	-	-	-
	2	0.725		0.805	0.823	0.811	0.813
	3	0.197		0.341	0.326	0.336	0.323
Geological age	1	-		-	-	-	-
	2	0.612		0.454*	0.402*	0.396*	0.406
	3	0.847		0.621	0.498*	0.493*	0.498
	4	0.304**		0.299**	0.204**	0.205**	0.204**

it 6.7% more likely the decision to develop a field in a given month will be made. Model 1 suggests a particularly high impact from the oil price on the investment decision as it indicates a 24.8% probability increase from one month to the next based on a $10 oil price increase during the same month. However, Model 1 is a particularly simple model, in which it may be reasonable to believe that much of the heterogeneity between the assets is left unaccounted for. This is also reflected by a likelihood ratio test supporting the inclusion of the additional variables in Model 2. Considering only the models that account for differences in both location and geological age, the range of oil price impact varies considerably less; from 5.6% to

6.7%.[9]

None of the models are able to confirm the significance of oil price volatility. In addition, bearing the real options theory in mind, the hazard ratios corresponding to the volatility are not in accordance with the predictions from this theory. This is interesting in the light of the discussion by Bar-Ilan and Strange (1996), who explain that an increase in volatility may increase the probability of investing. It is also particularly interesting that industry experts (Rystad Energy, 2015) state that, to their knowledge, oil price volatility is typically not considered by oil companies when evaluating an investment. This contradicts the way volatility is emphasized in the traditional real options way of thinking. One possible explanation is the particularly long horizon for offshore oil and gas field development and production. Current price uncertainty may not be essential for a development decision, as long as the oil price is mean-reverting over a longer horizon.[10]

The sizes of crude oil and gas reserves have the most stable, significant positive effects on the hazard rates across all models. It is reasonable that larger assets are more likely to justify large start-up investment, which is required to develop a field. Both oil and gas reserves impact the hazard rate positively. The estimate of the oil reserves matters more in the decision to invest than the estimate of the gas reserves. Industry experts attribute this to the higher historical profitability of oil extraction compared to gas extraction on the Norwegian Continental Shelf (Rystad Energy, 2015). Interpretation of the hazard ratios corresponding to the reserves variables

[9] Also including Model 2 and Model 5 in which the oil price is insignificant at conventional levels.

[10] Bessembinder et al. (1995) and Smith and McCardle (1999) find the oil price to be mean-reverting over longer periods of time.

is slightly more complicated, due to the logarithm transformation. Increasing the *log* of crude oil resources by one increases the hazard rate by a proportion between 31.3% and 42.3% according to the different models. Thus, if the crude oil resources estimate increases by e (≈ 2.72) times, for instance from 100 million barrels to 272 million barrels, the effect on the investment decision would be as suggested by this percentage range. Therefore, the size of the oil and gas reserves is important when considering whether to develop a field. However, caution should be exercised to rely on the magnitude of the model estimates as a resource estimate at the time of discovery is not available and hence Rystad Energy's current resource estimates have been used. Time-varying resource estimates could provide more reliable results, if available, as also pointed out by Hurn and Wright (1994).

The covariates *Appraisal wells* and *Block extension* are not significant at conventional levels in any of the models, but contribute to the overall fit of the model.[11] *Appraisal wells* provides weak evidence that an increase in the number of appraisal wells drilled is associated with a longer approval lag. This may be caused by the time elapsed during drilling in addition to the fact that larger and more complex fields require more appraisal wells in order to determine their resource potential. If so, one may, however, debate whether this variable actually adds any new information or if it is a result of the time required to obtain sufficient field data. Tests of correlation between *Appraisal wells* and the size of resources indicate a positive relation. Hence, some of the variation in this variable is explained by the need to drill more appraisal wells when fields are large. In interpreting the *Block extension* variable, we see weak

[11] A likelihood ratio test does not reject the null hypothesis that the smaller specification fits the data better, hence we keep both variables in the model.

evidence that an increase in the number of blocks spanned by a field is associated with increased approval duration. This may be due to the reasons discussed in section 2.3; increasing bureaucracy and disputes when several operators must come to an agreement concerning several blocks.

The reservoir pressure is significant both in the univariate analysis and in Model 3-5 in which it is included. As argued in section 2.3, the pressure is important when deciding which equipment is needed for extracting resources. Advanced high-pressure equipment may be more expensive or even unavailable, and as a result, one may hesitate for a longer time before investing or decide not to invest at all. For instance, equipment sufficiently robust for extracting resources from high pressure, high-temperature reservoirs was not invented until recently (Rystad Energy, 2015).

As discussed in section 2.2, discoveries made in areas near existing fields may benefit from better field economics due to the possibility of utilizing existing infrastructure. We have included *Nearby approvals* to measure the effect this may have on the appraisal lag. However, this variable does not have a significant impact on investment decisions.

For the two categorical variables, *Location* and *Geological age*, the interpretation of the hazard rate differs somewhat from the continuous variables reviewed so far. In both cases, the first category is used as a reference. The hazard rate estimate for the other categories is the probability of event occurrence, relative to the reference category.[12] Hence, the results indicate a longer appraisal duration for assets located in the Norwegian Sea and the Barents Sea, compared to assets in the North Sea.

[12] As a result, it does not make sense to interpret the magnitude of this ratio on its own.

Although *Location* loses its significance at conventional levels when included with the other variables, this relation is consistent across all models. Hence, there is some evidence for what was argued in section 2.3. Also, despite the weak significance, it seems *Location* still contributes to the overall fit of considered models[13] and may, therefore, be helpful when accounting for heterogeneity across assets. Considering *Geological age*, there are consistent results across all models indicating that assets in the Silurian, Devonian, Carboniferous, and Permian (Category 4) category are more likely to have a longer duration compared to the other categories. The difference between Triassic and Jurassic (Category 3) and Cretaceous (Category 2) is small but they both indicate a higher probability of investment compared to Paleogene and Neogene (Category 1).

5 Conclusion

This paper studies how economic and field-specific variables influence the decision to develop a field made by oil companies operating on the Norwegian Continental Shelf. To investigate how the explanatory variables affect the length of the appraisal lag; the time from discovery to approval, we apply a duration analysis. This method is favourable because it enables us to include unapproved fields in our analysis. Economic variables are incorporated as time-varying to capture the effect of variations in the oil price and volatility over the appraisal lag. We account for heterogeneity across the fields by including the field-specific variables, namely the size of the oil reserves, the size of the gas reserves, and the geographic and geological variables.

[13] A likelihood ratio test supports the inclusion of the variable.

We find that the size of the oil and gas resources are the most important explanatory factors of the investment activity on the Norwegian Continental Shelf. Oil companies are more likely to invest earlier when the estimated reserves size of a discovery is larger. Furthermore, the results indicate a positive impact of the oil price on the investment decision. However, none of our models confirm any significance of volatility on the appraisal lag.

This paper supports the results of previous studies suggesting that oil price and the size of reserves are crucial determinants in the investment decision. Our findings concerning volatility agree with those of Hurn and Wright (1994), thereby contradicting the findings of Favero et al. (1994). We believe these results are interesting in light of what Bar-Ilan and Strange (1996) argue; there are scenarios in which increasing volatility may, in fact, increase investment probability. After discussions with industry experts, we consider it probable that oil price uncertainty itself is not an important factor for oil companies when making offshore investment decisions.

Acknowledgements We would like to thank Rystad Energy represented by Andreas Eraker for providing data and advice, Bo Henry Lindquist for insight into the survival analysis, and Stein-Erik Fleten, Erik Haugom, Einar Cathrinus Kjenstad and Lukáš Lafférs for useful comments.

References

Aadnoy, B. S., 2010. Modern well design. CRC Press.

Alquist, R., Kilian, L., 2010. What do we learn from the price of crude oil futures? Journal of Applied Econometrics 25, 539–573.

Bar-Ilan, A., Strange, W. C., 1996. Investment lags. The American Economic Review 86, 610–622.

Bessembinder, H., Coughenour, J. F., Seguin, P. J., Smoller, M. M., 1995. Mean reversion in equilibrium asset prices: evidence from the futures term structure. The Journal of Finance 50, 361–375.

Bøe, K., Jordal, T., Štěpán Mikula, Molnár, P., 2018. Do political risks harm development of oil fields? Journal of Economic Behavior & Organization Forthcoming.

Collett, D., 2015. Modelling survival data in medical research. CRC press.

Cox, D. R., 1992. Regression models and life-tables. In: *Breakthroughs in statistics*, Springer, pp. 527–541.

Dunne, T., Mu, X., 2010. Investment spikes and uncertainty in the petroleum refining industry. The Journal of Industrial Economics 58, 190–213.

Favero, C. A., Hashem Pesaran, M., Sharma, S., 1994. A duration model of irreversible oil investment: Theory and empirical evidence. Journal of Applied Econometrics 9.

Hess, K. R., Serachitopol, D. M., Brown, B. W., 1999. Hazard function estimators: a simulation study. Statistics in medicine 18, 3075–3088.

Hurn, A. S., Wright, R. E., 1994. Geology or economics? Testing models of irreversible investment using North Sea oil data. The Economic Journal pp. 363–371.

Kellogg, R., 2014. The effect of uncertainty on investment: evidence from Texas oil drilling. The American Economic Review 104, 1698–1734.

Moel, A., Tufano, P., 2002. When are real options exercised? An empirical study of mine closings. The Review of Financial Studies 15, 35–64.

Mohn, K., Osmundsen, P., 2008. Exploration economics in a regulated petroleum province: the case of the Norwegian Continental Shelf. Energy Economics 30, 303–320.

Osmundsen, P., Roll, K. H., Tveteras, R., 2012. Drilling speed–the relevance of experience. Energy Economics 34, 786–794.

Rystad Energy, 2015. personal communication with Andreas Eraker, consultant at Rystad Energy.

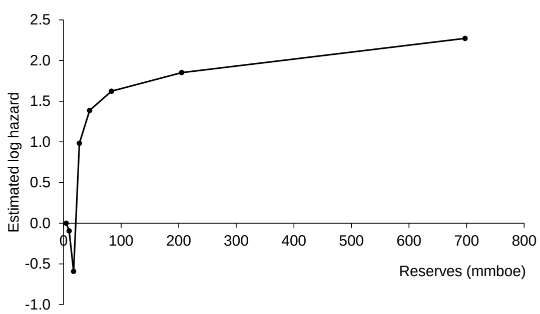
Sarkar, S., 2000. On the investment–uncertainty relationship in a real options model. Journal of Economic Dynamics and Control 24, 219–225.

Smith, J. E., McCardle, K. F., 1999. Options in the real world: Lessons learned in evaluating oil and gas investments. Operations Research 47, 1–15.

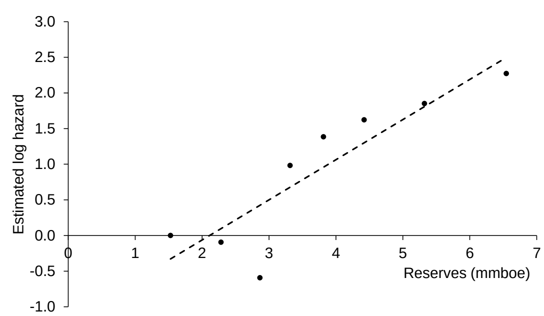
Appendix: functional form of the resources variables

In survival analysis, the relation between the predictor and the log hazard does not have to be linear. We particularly investigate the crude oil and gas reserves variables as their spans of values are particularly large and the economic interpretation suggests a nonlinear relationship. For instance, an increase in reserve estimates of 100 million barrels may be significant for a field believed to contain 50 million barrels, but have very little significance for a field with an estimated size larger than 1 billion barrels.

The simplest method for investigating the linear effect is to replace the covariate with design variables formed from its quartiles. The estimated coefficients for the design variables are plotted with the midpoints of the intervals defined by the cut points. At the midpoint of the first interval, a point is plotted at zero. If the correct scale is linear in the log hazard, the polygon connecting the points should be approximately a straight line. The results are displayed in Figure 9a and Figure 9b.



(a) Estimated coefficients versus 8 quantiles for resources



(b) Estimated coefficients versus 8 quantiles for ln(resources)

The effect of the resources on the log hazard appears to be approximately logarithmic in Figure 9a. This is confirmed by transforming the variable into its logarithmic equivalent, shown in Figure 9b, as the points form an approximately straight line.

Applying the method of fractional polynomials further confirms the logarithmic transformation proposed in the last section. Stata compares a range of different specifications for the variable examined, in this case *Oil reserves*. Since economic intuition suggests that the function should be monotonic, we consider only the fractional polynomials of the first degree. The results are displayed in table 4. The best fitting model is the approach where *Oil reserves* are included using a log transformation (power of 0), which is significant for all relevant significance levels. The same transformation results apply for the variable *Gas reserves*. The results are similar to what we find using the graphical model.

Table 4: Results from comparing different first degree transformation of the Oil reserves variable. The best fitting transformation is obtained by a log transformation. P-value is the significance of the deviance difference comparing the reported model with m = 1 model.

Oil reserves	**DF**	**Deviance**	**Dev. dif.**	**p-value**	**Powers**
Not in model	0	999.92	28.28	0.00	0
Linear	1	988.18	16.54	0.00	1
m = 1	2	971.65	-	-	0