FACULTY OF SCIENCE AND TECHNOLOGY

MASTER'S THESIS

Study program/specialization: Industrial Economics Project Management	Spring semester, 2017 Open
Author: Kasper Stensrud Kleivbo	………………………………………. (signature of author)
Internal supervisor: Roy Endré Dahl	
Title of master's thesis: Cost overruns in Norwegian projects – An econometric study	
Credits: 30	
Keywords: Cost overruns Econometrics Macroeconomics Statistical analysis	Number of pages: 66 + supplemental material/other: 0 Stavanger, June 15th, 2017

Title page for Master's Thesis

Faculty of Science and Technology

Acknowledgements

This thesis is written as my final part of the Master's program Industrial Economics at the University of Stavanger. The thesis is rated to 30 credits and was completed during the spring of 2017.

I would like to thank my supervisor, Dr. Roy Endré Dahl at the University of Stavanger, for his enthusiasm and constructive feedback along the way. His enthusiasm and expertise regarding the subject has greatly inspired me to perform my very best to meet his expectations. Our cooperation has worked well throughout the whole process, and has made my work with the thesis both challenging and enlightening. I would also like to thank other supporters who have contributed with important data for my thesis. This includes Dr. Morten Welde at the Concept Research Programme at Norwegian University of Science and Technology, Geir Saxebøl from Statens Vegvesen and Morten Haveraaen from Statsbygg. Their contribution has been crucial in order to conduct the analysis in this thesis.

Kasper Stensrud Kleivbo Stavanger, 15. June 2017

Abstract

Cost overruns are a global phenomenon. By assuming that companies are profit maximizing, we imply that inaccurate estimates of project costs are unwanted as the basis for investment analysis is weakened. The aim of this thesis is therefore to identify which factors affect the ability to set accurate budgets and meet the estimated costs. I highlight this topic by analyzing the differences in cost overruns between projects from the Norwegian oil industry and the public sector, by introducing macroeconomic variables for analysis and by looking into whether the cost overruns from one sector affects the other.

Firstly, descriptive statistics and univariate regressions were run in order to obtain a better overview of the topic. Both public and oil projects are statistically more prone to cost overruns than underruns, however, oil projects experience overruns of larger magnitude. Public projects show a trend where increasing project size reduces cost overruns, while cost overruns in oil projects tend to increase with the duration of the project.

To further analyze the dynamics of cost overruns, multivariate regressions were performed. This includes using forward selection by iterative processes in order to arrive at the final models. For oil projects, I find the variables Duration, Pension fund surprise, GDP growth and NCS investment surprise to significantly affect the magnitude of cost overruns, explaining about 25% of the variability in cost overruns for oil projects. As for public projects, the corresponding model includes the variables Duration, Employment level, GDP from marine activities and Export, explaining about 13% of the variability.

The models above indicate that cost overruns in both sectors depend on the macroeconomic environment at the time of project execution. The explanatory power for the oil model is rather acceptable, however, I fail to find a good general model for cost overruns in public projects. When comparing the two sectors by running the models on the opposite dataset, no causal relationship in cost overruns between the two was found, and I therefore fail to confirm whether the different sectors affect the cost overruns of each other.

Table of contents

List of tables

List of figures

This page is intentionally left blank

1 Introduction

1.1 Scope and purpose of this thesis

The purpose of this thesis is to study the nature of cost overruns across Norwegian industries. Cost performance, in this thesis referred to as cost overruns, is one of many Key Performance Indexes (KPIs) to measure a projects overall performance. In project management, an important task is to forecast the estimated costs of the execution of a project, ultimately playing a major part in whether or not a project is selected for execution. Hence, the accuracy of estimated costs emerges as an important area of focus in order to optimize the portfolio of projects to be executed.

This study is limited to analyzing cost overruns in the Norwegian oil industry on the Norwegian Continental Shelf (NCS) and public projects from various public entities. Although public entities and the oil industry have different incentives, it is still in the interest of both to minimize cost overruns. A key problem regarding cost overruns, which applies to both public entities and the oil industry, is the selection of the most economically feasible projects. For example, underestimation of costs and overestimation of profits and social benefits causes the decisions on which projects to be executed are done on the wrong premises, as inaccurately estimated projects may seem more attractive than what they really are. For the oil industry, this problem may impair the companies' goal of being profit-maximizing, making their portfolio of projects suboptimal in terms of profits. Consequently, this will directly affect a company’s competitiveness in an international industry. For public entities, one main incentive for minimizing cost overruns is to maintain the people’s confidence in the spending of public assets. In Norway we have a generally high level of taxation, and thus, public projects are largely funded by taxpayers' money. Unlike the oil industry, the goal of public entities is to maximize the social benefits a project implies. Ultimately, the selection of public projects to be executed is a tradeoff between estimated costs and social benefit.

Despite the importance of reducing cost overruns, as demonstrated above, both oil projects and public project continue to experience cost overruns. In recent years, several studies on cost overruns in public transport infrastructure projects have been conducted, identifying both why these cost overruns occur and measures that can be done to reduce them. Flyvbjerg, Holm and Buhl published a study in 2002 claiming to be the first statistically significant study of cost overruns in transportation infrastructure projects, consisting of over 250 projects. They found that cost overruns occur on a regular basis regardless of project type, geographical location and when the projects were executed. Several other studies conclude the same, for example Lundman (2011). However, the amount of quantitative research on cost overruns the in the Norwegian oil industry is rather limited. There are, however, several in-depth case studies of oil projects on the NCS, for example the Norwegian Petroleum Directorate (2013) and NOU (1999). Despite the fact that the two were published several years apart, they both find substantial cost overruns in projects on the NCS. In recent years, some quantitative research on cost overruns in Norwegian oil projects have been published, such as Dahl et al. (2017), Lorentzen et al. (2017) and Oglend et.al (2016)

My interest for the topic arose after reading about substantial cost overruns in public projects in the media. After discussing the topic with my supervisor, Professor Roy Endré Dahl at the University of Stavanger, we came up with the idea of conducting a quantitative study comparing cost overruns in the oil industry and public entities in Norway. As far as my supervisor and I are aware, there are no quantitative studies on cost overruns across industries in Norway. This thesis is therefore an attempt to improve upon this, and gain insight in how the oil industry and public entities differ in terms of cost overruns.

The aim of this thesis is therefore to analyze cost overruns in Norwegian projects across industries. More precisely, the objective is to perform a quantitative analysis based on projects approved estimated and actual costs, comparing cost overruns in oil projects with public projects. This is achieved by analyzing descriptive statistics and performing univariate and multivariate econometric analysis using the data management software R Studio. We consider both project specific factors and macroeconomic factors that may contribute to cost overruns in a positive or negative matter.

This leads us to the following research questions to be discussed in this thesis:

1. **Is there a difference in cost overruns between oil projects and public projects?**
2. **How are cost overruns affected by macroeconomic factors?**
3. **Are the sectors' cost overruns affected by each other?**

1.2 Limitations

This thesis only investigates cost overruns with respect to estimated and actual costs of projects, where estimated costs are dated to the decision to invest (or as close to this as possible for projects where this information is not available). Any changes in scope during project execution, which may be a cost driver relative to estimated costs, are disregarded due to availability of data. A key question is whether the dataset is representative of the population. As the criterion for sampling was data availability, especially a problem for public projects, the dataset is not a complete list of executed projects. Thus, the results should be interpreted with caution.

1.3 The Norwegian Continental Shelf

Petroleum reserves were discovered on the NCS in the late 60s, and production of oil and gas began in the early 70s. The production had its peak in 2004 with approximately 264 000 million standard m3 (Sm3) of oil equivalents (o.e). Since then, the production on the NCS has been declining; indicating that the NCS has reached a mature phase. Although the production levels stagnated in 2004, the accumulated investments on the NCS continued to grow. As a reaction to falling production levels, the Norwegian government introduced incentives for further development. This resulted in tax relief on exploration expenses. In addition, the oil price increased dramatically in this period, which also stimulated further development on the NCS. As of 2015, the oil industry accounted for 15% of Norway's GDP, 20% of government revenues and 39% of Norwegian exports (Olje- og energidepartementet, n.d.). The oil industry is therefore an important source of income for the authorities and the Norwegian society in general. Large reductions in the oil price in recent years have led to major ripple effects in the Norwegian society, and as a result, the cost focus in the oil industry has increased. Now that the profit margins in this sector is greatly reduced, cost performance is therefore more important than ever, including the reduction of cost overruns in projects. The figures below shows the historical production levels on the NCS and the accumulated investments for the time period covered by the data set.

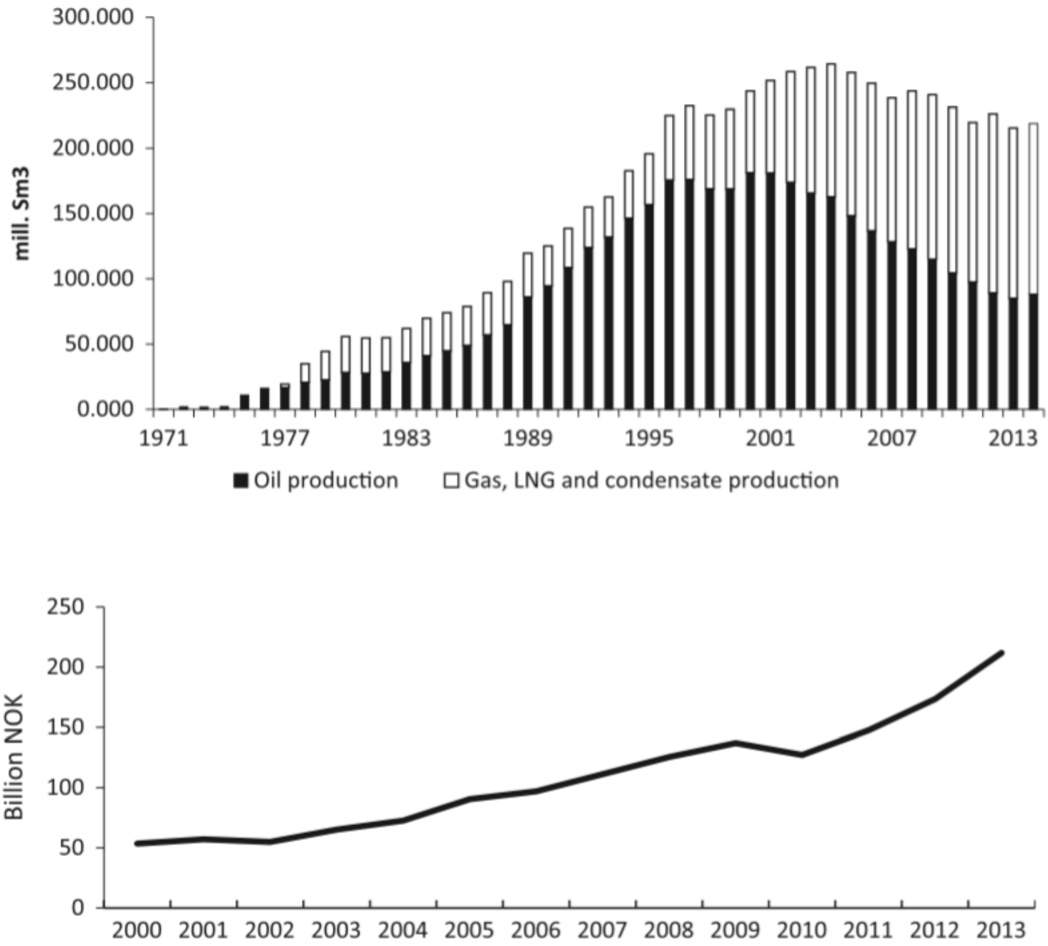

Figure 1.1: Historical production and investment levels on the NCS (Dahl et al., 2017)

1.4 Public sector

The public sector is the part of the economy providing governmental services such as military, police, infrastructure, education and health care. Unlike the private sector, whose purpose is to make profits for the owners, the target of public sector is to provide services with social benefits. As public services are largely funded by taxation, it is important that public funds are managed in an efficient and responsible manner. In order to ensure this, large public projects with estimated costs above 750 mNOK are obliged to follow a specific project model including external quality assurance. This is the so-called KS scheme (Norwegian abbreviation for quality assurance). In recent year, the focus on infrastructure development has increased substantially in Norway, and the investment budgets are increasing rapidly. The graph below shows the cumulative investment budgets for Statens Vegvesen, Jernbaneverket and Statsbygg, the public entities executing public road, rail and building projects. Since year 2000, the budgets have more than doubled, emphasizing the government's focus on infrastructure development.

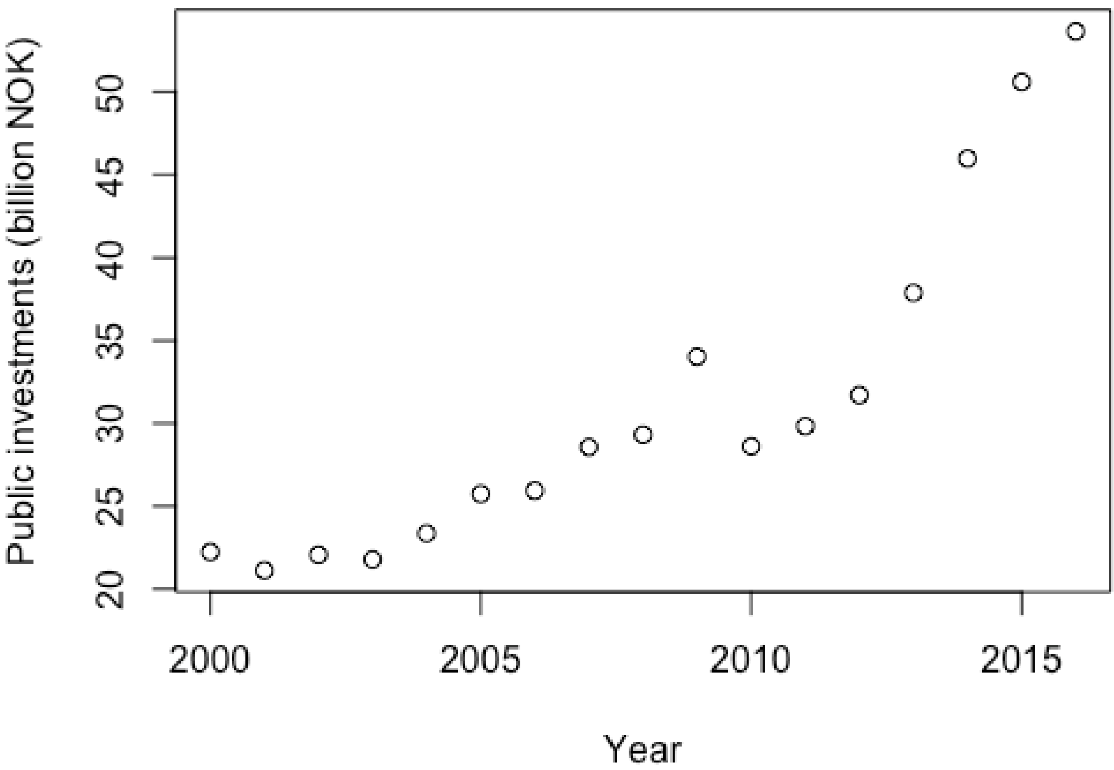

Figure 1.2: Public investments (2015 NOK-values)
Data source: Annual reports from Jernbaneverket, Statens Vegvesen and Statsbygg.

1.5 Literature review

The following section will review some relevant publications and results regarding cost overruns in projects.

Merrow (2011)

Merrow (2011) studies a dataset of 315 projects scattered across the world, whereas 130 of these are related to oil and gas. All projects had estimated costs exceeding 1 billion USD, so-called megaprojects, and were implemented between 1995 and 2010. Marrow defined the project outcome as a binary value, defining projects as unsuccessful in terms of cost performance if projects experienced cost overruns exceeding 25% of its estimated costs, and vice versa. In the study, he found that 78% of oil and gas projects to be unsuccessful (cost overrun > 25%), with an average cost overrun of 33%. For the whole dataset, he found 65% of projects to be unsuccessful, averaging a cost overrun of 40%.

Ernst & Young (2013)

A report by Ernst & Young (2013) provides a similar study consisting of 365 oil and gas megaprojects on a global basis. By evaluating the cost performance of selected projects, they find that cost overruns occur in 51% of projects, disregarding geographical location. The average cost overrun found was 64%. For European projects, 53% experience cost overruns, averaging a cost overrun of 57%. Ernst & Young also focus on the frequency of delays in projects, which may be a significant cost driver.

Location	Cost overrun frequency	Delay frequency	Average cost overrun
North America	58%	55%	51%
Latin America	57%	71%	102%
Europe	53%	74%	57%
Africa	67%	82%	51%
Middle East	89%	87%	68%
Asia/Oceania	68%	80%	57%
Total	**51%**	**73%**	**64%**

Table 1.1: Project outcome of oil and gas projects (Ernst & Young, 2013)

A major deficiency in the studies of Merrow (2011) and Ernst & Young (2013) is their focus on average cost overruns in projects. Therefore, we gain no insight into the underlying factors that contribute to cost overruns. The following two studies provide statistical analysis of the underlying causes, and are therefore of great interest when analyzing the performance of Norwegian projects.

Flyvbjerg et al. (2002, 2004)

As opposed to the findings of Merrow (2011) and Ernst & Young (2013), whose studies only considers descriptive statistics and the statistical distribution of the datasets, Flyvbjerg et al. (2002) published what they claim to be the first statistically significant study on cost overruns in transport infrastructure projects. Their study considered 258 projects implemented in the year interval 1910-2000 from 20 different countries. They found that 9 out of 10 projects experienced cost overruns, exceeding the estimated costs by 28% on average and with a standard deviation of 38%. The study distinguishes between project types (rail, road and bridge/tunnels) and geographical location (Europe, North America and others). The statistical analyzes show some differences in cost overruns between types in transport infrastructure projects, however, they do not appear to be more prone to cost overruns compared to other types of large projects. They also find that cost overruns seem to be a global phenomenon, as well as no reduction in cost overruns was observed over the 70 years the study examines.

Project type	Number of observations	Average cost overrun	Standard error
Rail	58	45%	38%
Bridge/tunnel	33	34%	62%
Road	167	20%	30%
Total	**258**	**28%**	**39%**

Table 1.2: Cost overruns in transport infrastructure projects (Flyvbjerg et al., 2002)

Sovacool et al. (2014)

The study investigates the frequency and magnitude of cost overruns in electricity related projects between 1936 and 2014. Considering 401 projects in hydroelectric dams, nuclear reactors, wind farms etc. in 57 countries, the average cost overrun found was about 66%, with nuclear reactors being the most prone to cost overruns with an average of 117%. The study finds that increasing project duration may inflate cost overruns due to unpredictability, delays etc. They also found that bigger projects experienced increased cost overruns, however, the trend couldn’t be said to be significant for all project types. There was also no clear statistically significant evidence of a present learning effect in projects – that is reduced cost overruns over time due to accumulated learning and experience. However, the findings correspond with that of Flyvbjerg et al. (2002), that cost overruns were observed regardless of project type and geographical location.

Project type	Number of observations	Average cost overrun	Standard error
Nuclear reactors	180	117%	152,1%
Hydroelectric dams	61	71%	111,7%
Thermal power plants	36	13%	33,5%
Wind farms	35	8%	13,1%
Transmission projects	50	8%	40,4%
Solar farms	39	1%	17,8
Total	**401**	**66%**	**NA**

Table 1.3: Cost overruns in electricity projects (Sovacool et al., 2014)

Dahl et al. (2017)

This study by Dahl et al. (2017) offers a quantitative econometric analysis of cost overruns in the Norwegian petroleum industry, a study consisting of 80 petroleum related projects executed in the time period 2000-2013. They find several variables related to project characteristics and the business cycle of the petroleum industry to significantly affect the extent of cost overruns. The variables include changes in employment levels, project duration and project size to significantly contribute to the extent for cost overruns. They also found that the largest cost overruns typically occur late in the project execution time. Some of the variables causing cost overruns may be related to project management, as their analysis points toward project planner not paying enough attention to planning during boom periods, as suggested by the finding of this study.

Lorentzen et al. (2017)

The study is based on projects collected from a number of sources, including Norwegian public registers and other published articles. Equal to Flyvbjerg et al. (2002), they fail to find any statistically significant reduction in cost overruns over the last 30 years. The findings also correspond with Dahl et al. (2017) that cost overruns increase with project size. By comparing the dataset with the studies of Merrow (2011), they also find that the cost overruns in Norwegian oil projects do not differ significantly from other geographical areas, as proposed by Flyvbjerg et al. (2002).

From the above literature review, we have good indications that cost overruns are a phenomenon occurring regardless of project type and location. The findings of Flyvbjerg et al. (2002), Sovacool et al. (2014), Dahl et al. (2017) and Lorentzen et al. (2017) suggest that various project specific and macroeconomic variables such as project size and duration, commodity prices and business cycles may incur cost overruns. Thus, similar analyzes will be conducted on Norwegian projects in this thesis.

2 Data

Norwegian law requires that licensees on the NCS to submit and gain approval for a plan for development and operation of the petroleum deposit (PDOs) and special permits for installation and operation (PIOs), as declared in the Petroleum Act (Lorentzen et al., 2017). The actual costs used in this thesis are the official actual cost the licensee has reported to the ministry. Estimated and actual cost data from 94 projects from the petroleum industry were gathered, with approval dates ranging from year 2000 to 2013.

However, access to the estimated and actual costs of public projects is limited and difficult to obtain, making data collection a rather time consuming task. Data was collected from various sources. This includes annual reports of public sector entities, the national budgets of relevant ministries, through personal contact with employees of public entities and through cooperation with the research program Concept at NTNU. The dataset of public projects consists of 814 projects, consisting mainly of infrastructure projects such as construction, rail and road, but also some from the defense sector and public IT-projects.

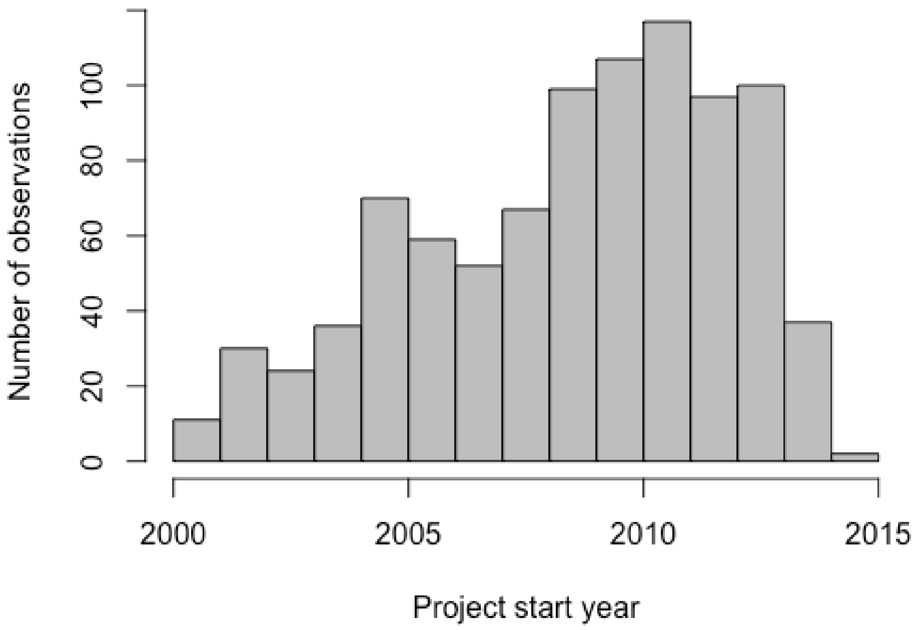

Figure 2.1: Dataset distribution of project start year

Due to availability data, scope changes, which may be a significant cost driver in projects, is not included in the analysis. The collection of data from projects over several years makes it possible to statistically analyze the development in cost overruns for both oil and public projects, using multivariate regression models to identify how changes in the macroeconomic environment contribute to cost overruns in both oil projects and public sector projects. It also makes it possible it identify the drivers of cost overruns, as proposed by Flyvbjerg et al. (2002).

The tables below outlines the distribution of projects in the final data set. Because of the availability of various project variables, the dataset was limited to projects where the year of start and completion, estimated and actual cost were available. All cost numbers are inflation adjusted to year 2015 NOK values using the Norwegian index of inflation. In addition, projects with estimated costs less than 10 mNOK are removed to increase the comparability of the different sectors. All projects were started in year 2000 or later, and completed no later than 2016.

Entity	< 100 mNOK	< 1000 mNOK	> 1000 mNOK	Total
Forsvaret		4	3	7
Forsvarsbygg		2	1	3
Jernbaneverket	8	17	6	31
NAV		1	2	3
Statens Lånekasse		1		1
Statens pensjonskasse		1		1
Statens Vegvesen	550	109	31	690
Statsbygg	25	47	5	77
UDI		1		1
Total public projects	**583**	**183**	**48**	**814**
Oil industry		**7**	**87**	**94**

Table 2.1: Number of projects per category

2.1 Cost Overruns

Cost overrun, sometimes referred to as cost escalation or cost increase, is an important measure when evaluating the implementation of a project. Note that cost performance is one of many Key Performance

Indexes (KPIs) for evaluating the success of a project, meaning that cost overruns in projects do not necessarily imply unsuccessfulness. In this context, the term is used informatively about the project owner's ability to comply with the agreed-upon estimated project costs. It should be noted that, in addition to the ability to perform on budget, the term might also say something about the quality of the project execution plan. That is, large cost overruns may not be due to low cost performance or lack of efficiency, but rather a result of a poorly planned project.

However, little quantitative research exists on the topic. For this thesis, the following definition of cost overruns was used: inflation adjusted actual cost minus estimated cost, as a percentage of estimated cost. See equation 1 below. Actual costs are defined as total project costs incurred by the project's end. Estimated costs are defined as the forecasted costs of a project at the time of decision to invest in a project (Flyvbjerg et al., 2002). For projects where estimated costs at the time of decision are not available, the closest available estimate was used; typically estimated costs at the time of project start. This is the same definition as used in most quantitative research on the topic.

$$Cost\ overrun = \frac{Actual\ cost - Estimated\ cost}{Estimated\ cost} \times\ 100\%$$

Equation 1: Cost overrun

For this thesis, the dataset consist of 94 oil projects and 814 public projects. To my knowledge, no studies comparing cost overruns in oil and public projects in Norway exist. As the project sizes of the dataset vary from 10 mNOK to several billion NOK, using relative cost overrun opposed to absolute cost overrun is necessary. If absolute cost overrun is used, the larger projects will dominate the regression models, effectively reducing the weighting of smaller projects in the analysis. Therefore, using relative cost overrun is more appropriate as it provides a more comparable basis for analysis. Instead, to include the effects of project size on cost overruns, this may be added as a separate variable in the regression model if it proves to be significant.

Using estimated costs from the time of decision to implement is somewhat controversial. According to Flyvbjerg et al. (2002), project promoters sometimes object to this method as cost estimates are continuously updated during project execution. During project execution, greater detail of designs, materials needed etc. are obtained, making cost estimates more accurate over time. As expected, the initial cost estimate is highly uncertain, and is therefore arguably unfair to use in terms of evaluating the implementation of a project. However, for the purpose of this thesis, the accuracy of initial cost estimates are of primary interest, as this directly affects the decision-making. According to basic microeconomic assumptions, companies are profit maximizing and capital is fixed in the short run (Levin and Milgrom, 2004). Given that companies have a portfolio of potential projects, inaccurate estimation of costs may lead to an inefficient allocation of scarce resources, as the expected value of projects, e.g. in terms of net present value or social utility, gives an incorrect ranking of projects. Thus, as proposed by Lorentzen et al. (2017), cost overruns may directly cause financial loss as a direct result of inaccurate cost estimation. From an economical perspective, this makes cost overruns unwanted, given that no changes in scope is made during the project execution. Scope changes may increase the productivity, social utility or even the net present value of projects, and will be discussed further in the Regression Analysis chapter of this thesis.

2.2 Statistical analysis

According to Flyvbjerg et al. (2002), the drivers of cost overruns may be divided into three main categories. The categories have distinct characteristics in terms of expected statistical distribution of a dataset, and therefore suggest which drivers that affect cost overruns in the dataset. The categories are as follows: forecasting error, optimism bias and strategic misrepresentation. The following table summarizes the statistical characteristics of each category.

Category	Mean (μ)	Skewness (γ)	Convergence
Forecasting error	= 0	= 0	$\lim_{t\to\infty} \mu_t = 0$
Optimism bias	> 0	> 0	$\lim_{t\to\infty} \mu_t = 0$
Strategic misrepresentation	> 0	> 0	No convergence

Table 2.2: Cost overrun drivers by category and their respective characteristics (Flyvbjerg et al., 2002)

Forecasting error is based on the assumption that stochastic processes always will be present during project execution. Hence, the actual cost of a project is a function of the estimated costs and stochastic processes. As a result, the actual costs of a project portfolio should be symmetrically distributed around the estimated costs, such that the actual costs equal the estimated costs for the portfolio ($\mu = 0$). Furthermore, as estimation techniques are continuously improved, cost overruns should converge towards zero over time.

Optimism bias is based on that the complexity and the probability of unlikely events in projects are underestimated. For a project portfolio, this would result in a mean cost overrun greater than zero as a result of insufficient planning and estimation of uncertain events. Opposed to forecasting errors, one would also expect a positive skewness of the cost overrun distribution due to optimism bias. Over time, as estimation techniques are improved and experience from projects are gained, optimism bias should be reduced and cost overruns converge towards zero.

Unlike the abovementioned categories, strategic misrepresentation is not based on uncertainties and the sub consciousness of project planners. Factors such as political and organizational pressure causes strategic overestimation of benefits and underestimation of costs in order to increase the chances of their project is approved and funded, and not the competitor's. In addition, if the project is big, it is difficult to cancel after the project has started due to large investments. (Flyvbjerg, 2007) This may systematically promote high cost overruns instead of projects with more realistic estimations of costs and benefits. Due to the nature of this category, no convergence is expected over time, given that the political and organizational pressure remains (Lorentzen et al., 2017).

2.3 Descriptive statistics

With this in mind, Flyvbjergs theory of cost overrun drivers may be applied to the dataset by considering its descriptive statistics. Table 2.3 below summarizes the average cost overruns of oil and public projects with respect to estimated costs, breaking down public projects into their respective public entities. Table 2.4 shows the general descriptive statistics of the dataset.

Public entity	< 100 mnok	100-1000 mNOK	> 1 000 mNOK	Total
Forsvaret		-7,93 %	3,34 %	-3,1 %
Forsvarsbygg		-0,18 %	-1,42 %	-0,59 %
Jernbaneverket	8,93 %	10,09 %	3,14 %	8,4 %
NAV		9,46 %	1,15 %	3,92 %
Statens Lånekasse		-0,01 %		-0,01 %
Statens pensjonskasse		46,15 %		46,15 %
Statens Vegvesen	11,81 %	5,34 %	5,03 %	10,48 %
Statsbygg	-0,86 %	-3,87 %	12,98 %	-1,8 %
UDI		12,04 %		12,04 %
Total public projects	**11,23 %**	**3,32 %**	**5,22 %**	**9,1 %**
Oil industry		**-9,85 %**	**16,81 %**	**14,82 %**

Table 2.3: Average cost overrun

	Number of projects	Cost overrun			Standard deviation	Level of significance, p*
		Minimum	Maximum	Average		
Public	814	-69,37 %	200 %	9,1 %	26,25 %	<0,001
Oil	94	-69,13 %	218,76 %	14,82 %	38,44 %	<0,001
All projects	908	-69,37 %	218,76 %	9,69 %	27,76 %	<0,001

Table 2.4: General descriptive statistics

*Level of significance from zero, one sided t test

We see from the above statistics that there are cost overruns in both oil and public projects on average. In public projects, the average cost overrun is 9,1 % with a strong significant difference from zero ($p < 0{,}001$). The cost overruns in public projects range between -69,37% and 200%. In oil projects, the average cost overrun is 14,82% with a strong significant difference from zero ($p < 0{,}001$). The cost overruns in oil projects range between -69,13% and 218,76%. For oil projects, the largest absolute cost overrun belongs to the Ormen Lange project, exceeding estimated costs by 38248 mNOK (38 billion NOK). For public projects, the E18 Bjørvika project had the largest absolute cost overrun of about 1686 mNOK. Note that the standard deviation of oil projects is higher than for public projects, 38,44% and 26,25%, respectively. In total, public projects tend to have a cost overrun of almost 6% less than oil projects. However, for projects with estimated costs greater than 1000 mNOK, public projects perform better, with an average cost overrun of 11,59 % less than for oil projects. The opposite pattern is observed for projects with estimated costs between 100 and 1000 mNOK; however, the number of observations is limited in this interval for oil projects.

According to Flyvbjerg s breakdown of cost overrun drivers, the distribution of cost overruns suggests presence of optimism bias and strategic misrepresentation regarding estimated costs, as the mean cost overrun is greater than zero. Note that scope changes are not included in the dataset, which may explain some of the cost overruns relative to the initial estimated costs. According to Merrow (2011), a project experiencing large cost overruns typically overruns its schedule as well due to the discovery of more work to be done. The limited availability of such information therefore makes it impossible to conclude on the drivers of cost overruns in Norwegian projects, as defined by Flyvbjerg.

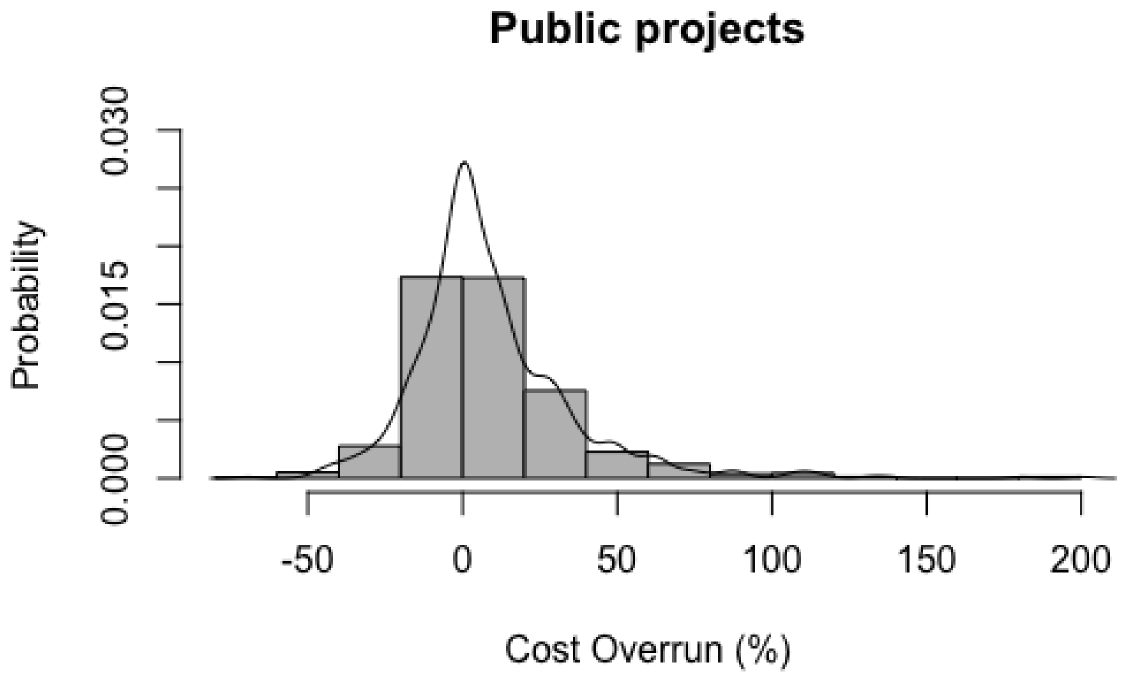

Figure 2.2: Histogram and density plot of cost overruns in public projects

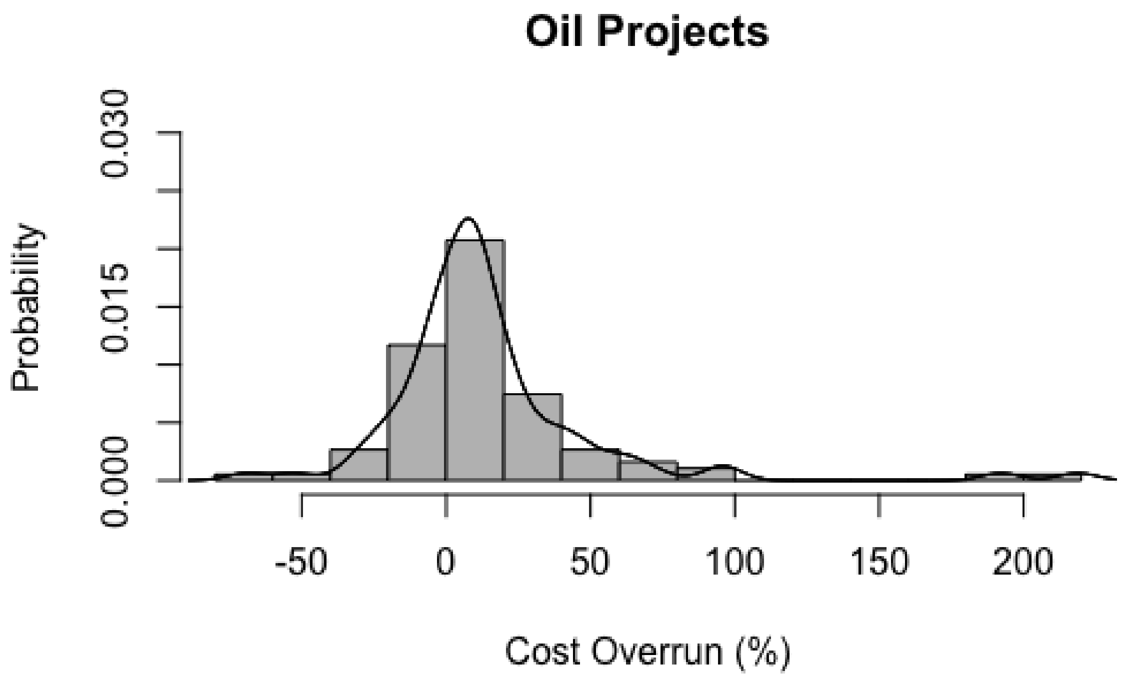

Figure 2.3: Histogram and density plot of cost overruns in oil projects

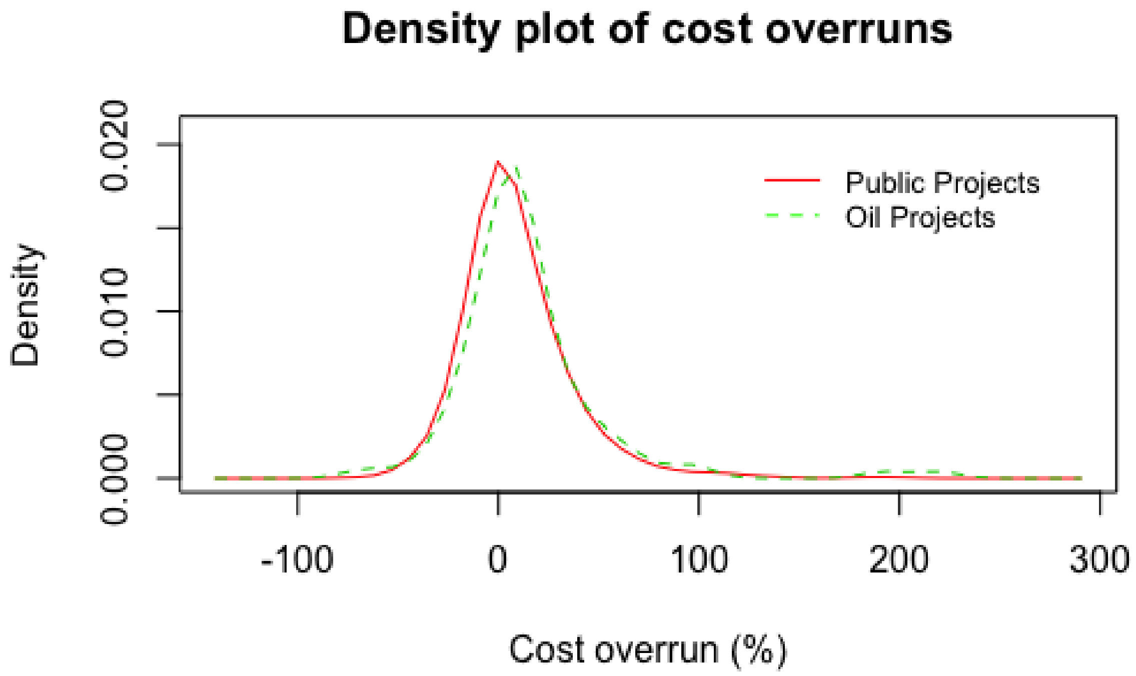

Figure 2.4: Comparison of cost overrun density plots of public projects and oil projects.

The figures above shows the distribution of cost overruns in public projects and oil projects. The skewness of the distributions is 2,71 for oil projects and 1,9 for public projects. Skewness is calculated by the formula below, where x_i is the cost overrun of observation i, $\bar{x}$ is the average cost overrun, n is the sample size and s is the sample standard deviation.

$$Skewness = \sum_{x=1}^{n} \frac{(x_i - \bar{x})^3}{ns^3}$$

Equation 2: Skewness of a sample distribution

Positive skewness indicates that the right tail of the distribution is thicker and longer than that of the left side, in this case indicating that both public and oil projects are more likely to experience cost overruns than not. Skewness of 0 would indicate a symmetric distribution about zero. With the positive skewness of oil projects being larger than public projects, it may indicate that oil projects are more vulnerable to extreme cost overruns. Evaluating the kurtosis of the distributions emphasizes this. As written by Wooldrigde (2009), the kurtosis of a distribution is a measure of the weight of the tails relative to the rest of the distribution, with high values indicating thicker and longer right tails.

Kurtosis is given by the following formula:

$$Kurtosis = \sum_{x=1}^{n} \frac{(x_i - \bar{x})^4}{ns^4}$$

Equation 3: Kurtosis of a sample distribution

The kurtoses of the distributions are 11,01 and 14,57 for public projects and oil projects, respectively. Again we find that oil projects are more prone to extreme cost overruns. This will be investigated further in the part 4 Empirical Analysis. Regarding Flyvbjergs definition of cost overrun drivers, this emphasizes the presence of optimism bias and strategic misinterpretation as suggested above. Whether optimism and strategic misrepresentation is the case cannot be concluded without further analysis. However, it should be noted that Flyvbjergs theory might not provide a complete picture of cost overrun drivers. Given that cost overruns is, by definition, limited to the interval between -100% to positive infinity, it may be unreasonable to expect a cost overrun distribution without positive skewness due to so-called black swan events (Dahl et al., 2017). It should also be noted that, practically, the event of cutting costs by 100% relative to the estimated costs is highly unrealistic. By comparison, there are no limitations for positive cost overruns, which may exceed the estimated costs by several hundred percent. (Lorentzen et al., 2017) This is emphasized by the dataset, as the cost overruns ranges from -69% to 218%. Hence, this suggests that a somewhat positive skewness and a thicker right tail in the distribution are to be expected.

3 Method

The following section discusses the methods used for univariate and multivariate analysis of the dataset. All regressions are performed in R Studio.

3.1 Univariate regression

Univariate regression is used to estimate a relationship between two variables, resulting in a modeled function of the dependent variable, in this case cost overruns. The output is then interpreted in terms of how the variables are related, as well as its statistical significance. Statistical significance is used to describe the likelihood of something being a result of random events. A significance level of 5% is typically used, meaning there is a 5% chance of rejecting the null hypothesis given that is was true, hence reaching the wrong conclusion. This way, by running statistical analysis in R Studio, one may establish the relationship between selected variables and how they affect the magnitude of cost overruns. One should also consider the R2, the explanatory power of the variable. This gives how much of the variability of the observations may be explained the selected variable, providing insight on the relationship between the two.

Another powerful measure in regression is the use of dummy variables. If the purpose is to distinguish between two groups, and not to observe how changes in a variable affect the dependent variable, creating a dummy variable may solve this problem. As written by Wooldridge (2009), the dummy variable takes the value 0 for one group and 1 for the other, meaning that the estimated coefficient on the dummy variable estimates the ceteris paribus effect between the two groups. For example, in this thesis, a dummy variable is used to distinguish between public and oil projects.

3.2 Multivariate regression

Although some variables may be found significant during univariate regressions, the result may change when performing multivariate regressions. The goal of multivariate regression is to capture ceteris paribus effect of a variable, that is, capturing the effect (in this case, a variables effect on cost overruns) of a variable while keeping other variables constant. This is not the case for univariate regression, as only one variable is examined at the same time, hence we are unable to control that the other variables stays constant.

One should be aware that all multiple regression methods suffer from the assumptions of regression. Hence, no method is necessarily complete in terms of modeling a phenomenon. Firstly, both methods considered for this thesis are linear multivariate regression methods. Both methods can be interpreted as simplifications of a complex phenomenon that is not necessarily affected linearly by the selected variables. The other assumptions of multiple regressions, according to the Gauss-Markov theorem, include that we have a random sample of observations, variables are normally distributed, variables are independent of each other and that the variance of error terms are constant across the independent variables (Wooldridge, 2009). Thus, regardless of method used for creating a multivariate regression model, certain assumptions are made. Selection of multivariate regression method therefore depends on which method yields the greatest benefits for each individual case. Below follows a discussion of potential methods considered appropriate for modeling cost overruns.

3.2.1 AIC versus exhaustive multivariate regression

AIC is a method for model selection. Do to so, the method measures the relative quality of various models for a given dataset. Assuming data is generated by an unknown process *f*, the AIC model is a take on estimating the information loss when trying to model *f*. However, as *f* is unknown, one may only estimate the relative differences in information loss between selected models, whereas one would choose the candidate model that minimizes the information loss as the optimized model. In order to interpret which model predicts the unknown process the best, each candidate is given a calculated AIC value, as calculated by the formula below.

$$AIC = 2k - 2ln(\hat{L})$$

Formula x.x: AIC values (Akaike, 1974)

Where $\hat{L}$ is the maximized values of the likelihood function for the selected variables in the model, and *k* is the number of selected variables. As shown, increasing number of variables will sharply inflate the AIC-value of the model. Through the $\hat{L}$ –variable, the AIC method awards goodness of fit the model provides modeling the unknown process *f*. The thought behind AIC is parsimony as selection criteria, hence the *k*-variable in the AIC formula inflating the AIC value by 2 for each variable added. Given a selection of candidate models, the model with the lowest AIC-value is considered the optimized model. Hence, adding many variables relative to their contribution to model fit is damaging to a model.

However, there are several problems with using AIC when selecting a multivariate regression model. Firstly, this method contradicts "conservative" statistics by ranking models rather than evaluating their statistical power and thus eliminating hypotheses and theories. One of the most striking problems is the idea of ranking models by their AIC value. This way, one could argue that the AIC method ignores the possibility that none of the candidate models are particularly good, instead ranking the models relative to each other without respect to their overall quality. As a result, when using this method, one may simply identify “the best of a bad bunch”. And, recognizing the badness of the candidate models may be just as important as identifying the best among them. As a result, the possibility of utilizing logical reasoning to supplement the selected model is removed, as this method does not produce intuitive statistics, such as p-values and explanatory power.

AIC has also been criticized for being too generous despite the idea of the method being parsimonious. According to Kadane and Lazar (n.d.), despite the penalty factor for including more variables in a model, AIC is too liberal and often prefers more complex, wrong models to a simpler and “true” model. By complicating a model unnecessarily, the readability and interpretability of the model may be greatly weakened.

The other method for model selection considered in this thesis was using an exhaustive research method. This is an alternative to stepwise regression, such as AIC, whereas instead of ranking models based on AIC-values, models are evaluated based on their explanatory power and well-established statistics such as p-values and explanatory power. This means that variables are selected sequentially based on their respective marginal contribution to the overall explanatory power of the models, conditional that the variables add a given marginal explanatory power. For this thesis, a marginal contribution to explanatory power of 1% per variable is considered fair. Utilizing the LEAPS function in R Studio does this. The LEAPS function is an exhaustive iterative process where all possible combinations of variables are examined; telling what combination of variables yields the highest R2 for different number of variables in the model. The result is the optimized models for a multivariate model with *n* independent variables. This is done for both public projects and oil projects to examine the different cost overrun drivers between the two.

The strengths of using exhaustive research for model selection are many. As opposed to AIC, this method utilizes established statistical methods such as p-values and explanatory power (as given by R2). Although these are conservative concepts, these are still established as standard statistical measures, and thus easier to interpret relative to other publications.

However, this method relies greatly of the pool of variables up for selection. Of course, the variables must be relevant for the phenomenon that is being modeled, which may not be the case when examining all possible combinations of variables. With this in mind, introducing logical reasoning to complement this method is necessary, and, a great strength of this method. When doing research, one may have ideas about which candidate variables make sense to include. This allows for manipulation of the variable pool by taking into account the correlation coefficients to avoid multicollinearity. In addition, if variables in the selected models may not be logically explained, these can easily be sorted out to yield logically interpretable models. E.g. having a model with high explanatory power does not necessarily mean one would understand the modeled process well. This is considered a great strength which AIC does not allow.

Based on the above discussion, the exhaustive method is selected, as the possibility of logical manipulation of the model is considered very important when modeling cost overruns.

4 Empirical analysis

This chapter contains the empirical output obtained from the conducted analysis in R Studio. The chapter is structured as follows: firstly, the results from univariate regressions are presented, presenting both project specific variables and macroeconomic variables. Secondly, the results from the exhaustive multivariate regression follow.

4.1 Univariate regression

This section contains the univariate regression outputs for both public and oil projects.

4.1.1 Project specific variables

The table below summarizes the results from univariate cost overrun regressions, which was run for a number of project specific variables. All regressions are run with cost overrun as the dependent variable and the following variables as independent variables.

Variable	Beta	Std. error	T	Level of significance, *p*
All public projects				
Year started	0,09568	0,31286	0,306	0,76
Year finished	0,2316	0,3396	0,682	0,495
Budget size	-0,0034	0,0017	-2,017	**0,044**
Post crisis	4,176	2,236	1,867	0,0622
Post crisis (budget > 1000mNOK)	6,047	4,532	1,334	0,189
Post crisis (budget > 100 mNOK)	7,506	2,498	3,004	**0,00296**
Project duration	0,3282	0,6528	0,503	0,615
Project duration (budget > 100 mNOK)	0,9521	0,6389	1,49	0,138
Project duration (budget > 1000 mNOK)	1,7466	0,9902	1,764	0,0844
KS projects vs public				
All projects	-7,534	3,1163	-2,418	**0,0158**
Budget >= 1000 mNOK	-8,847	6,902	-1,282	0,2063
Budget >=100 & <1000	-4,77	3,54	-1,347	0,17961
Budget >= 100 mNOK	-2,159	2,461	-0,877	0,38124
Oil projects				
Year started	0,1288	1,0596	0,122	0,903
Year finished	1,2102	0,8583	1,41	0,162
Budget size	4,63E-04	2,87E-04	1,6141	0,1099
Post crisis	16,7	7,782	2,146	**0,0345**
Project duration	10,004	2,369	4,223	**5,66E-05**

Table 4.1: Univariate regression results of project specific factors

Oil projects versus public projects

There are two major groups of projects in the dataset, oil projects and public projects. The oil industry has been criticized in media for a high level of spending on equipment, wages etc., and top executives in the oil industry has pointed out a lack of focus on costs in development projects. (Bertelsen, 2017) Thus, it is natural to expect that there are statistically significant differences between cost overruns in oil projects and public projects, with oil projects experiencing larger cost overruns. To test this, a linear regression model with a dummy variable was created to differentiate between the two. Cost overrun was set to the dependent variable, with the dummy variable taking the value 1 for oil projects and 0 for public projects. The null hypothesis is that there are no differences in cost overruns between oil projects and public projects. Hence, public projects is the base group, making the beta equal to the mean cost overrun of the project size category, and the oil coefficient the increase in cost overruns for oil projects relative to public projects.

Budget size		Estimate	Std. error	T	Level of significance, *p*
All	Beta	9,0959	0,9725	9,353	<0,001
	Oil	5,725	3,0227	1,894	**0,0585**
>100 mNOK	Beta	4,714	1,675	2,217	0,027287
	Oil	11,107	3,114	3,566	**<0,001**
>1000 mNOK	Beta	5,222	4,592	1,137	0,2575
	Oil	11,584	5,721	2,025	**0,0449**

Table 4.2: Effect of oil projects dummy variable

As shown in the table, the null hypothesis was not rejected at 5% level for all project sizes (p = 0,0585). This may be because that there are no oil projects in the interval 10-100 mNOK, hence inflating the variance of public projects relative to oil projects. Note that the dummy variable is nearly significant for all project sizes (p = 0,0585). However, when removing projects in the interval 10-100 mNOK, the level of significance shows that the null hypothesis was rejected for project sizes exceeding 100 mNOK at a very high significance (p < 0,001). Hence, the assumption of no differences in cost overruns between the two groups has been violated. This is also shown when considering project sizes exceeding 1000 mNOK, which is also significant at 5% level (p = 0,0449). Therefore, this analysis shows that oil projects systematically experience higher cost overruns than public projects when considering projects with estimated costs exceeding 100 mNOK. A study by Welde (2017) concludes that the cost control in large public projects is good, which corresponds with my results that the public sector performs significantly better than the oil industry in terms of accurate cost estimates. The box plot below outlines the distribution of cost overruns in both public projects and oil projects. The box represents the middle 50% of observations in terms of cost overrun, while the thick line across the boxes is the median observation. We see from the heigth of the box that public cost overruns are distributed more tightly around the median than that of oil projects. From the length of the whiskers in the box plot, we also see that oil projects have a thicker right tail (the top whisker), as previously found in the density plots.

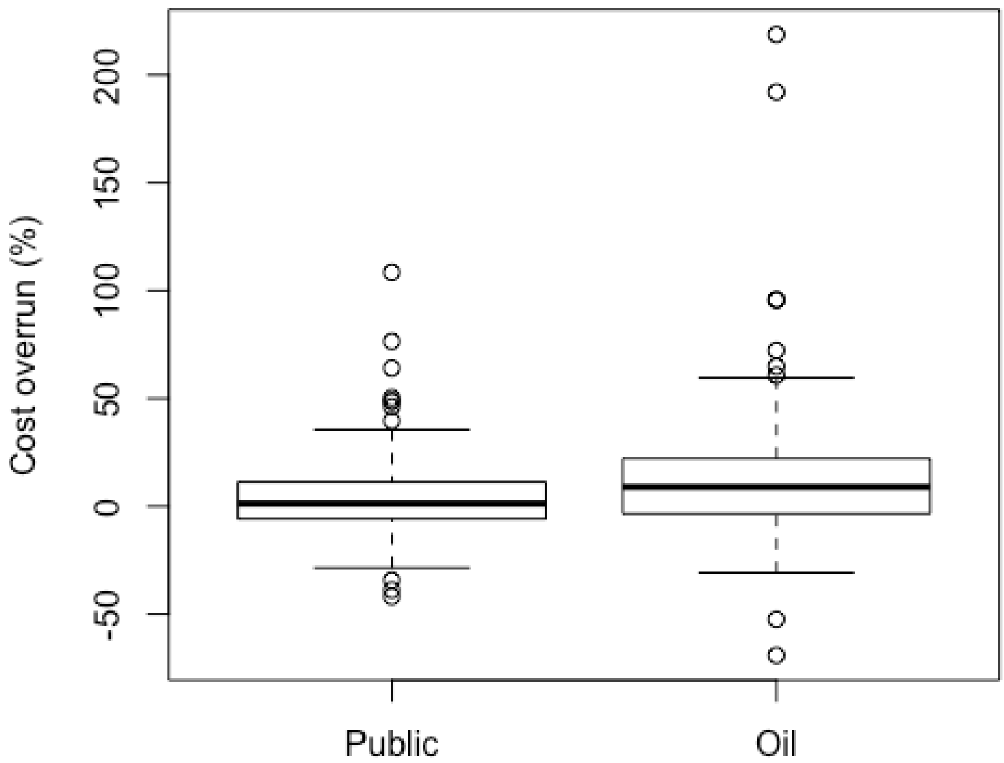

Figure 4.1: Box plot of cost overrun distributions for budget size >100 mNOK

Execution year

For both oil and public projects, we see that execution year, considering both year of project start and end, does not significantly contribute to cost overruns. The positive beta values indicate an increase in cost overruns over time, however this is not found significant. With respect to the previous discussion regarding Flyvbjergs categorization of cost overrun drivers, it is logical to expect a decline in cost overruns over time. Due to accumulated experience from various project types, improvement of estimation techniques and technological development, cost overruns should converge towards zero over time for cases of forecasting error and optimism bias. These results correspond with the conclusion of Flyvbjerg et al. (2002) that cost overruns have not decreased the past 70 years. As no decrease was found, no effect from accumulated project experience is observed. This points towards strategic misrepresentation of estimated costs being a driver of cost overruns. However, a conclusion regarding cost overrun drivers cannot be made without further analysis of other contributing factors, e.g. scope changes.

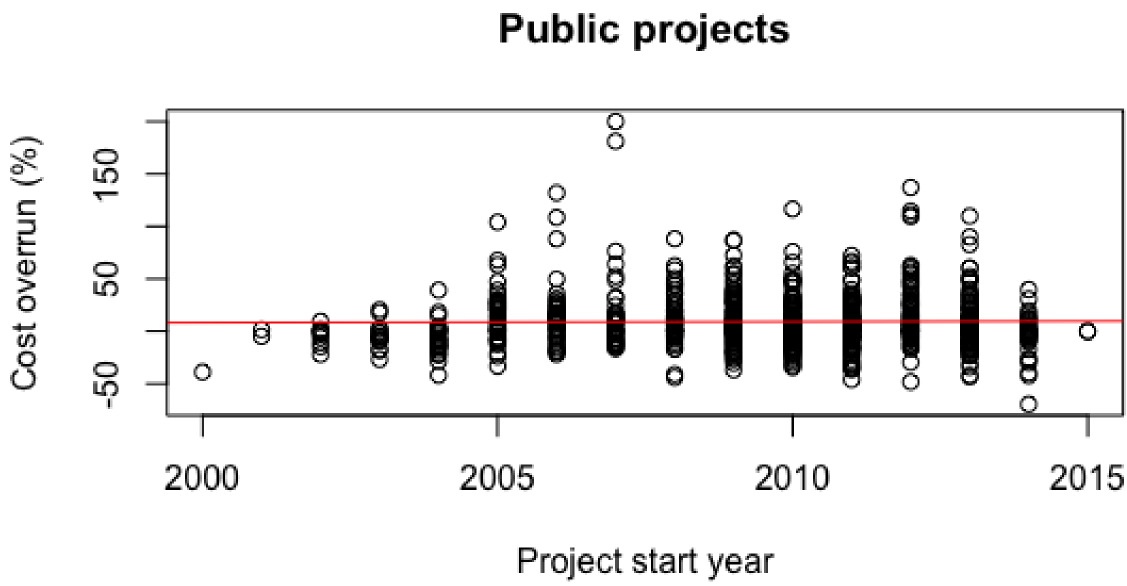

Figure 4.2: Public projects start year univariate regression

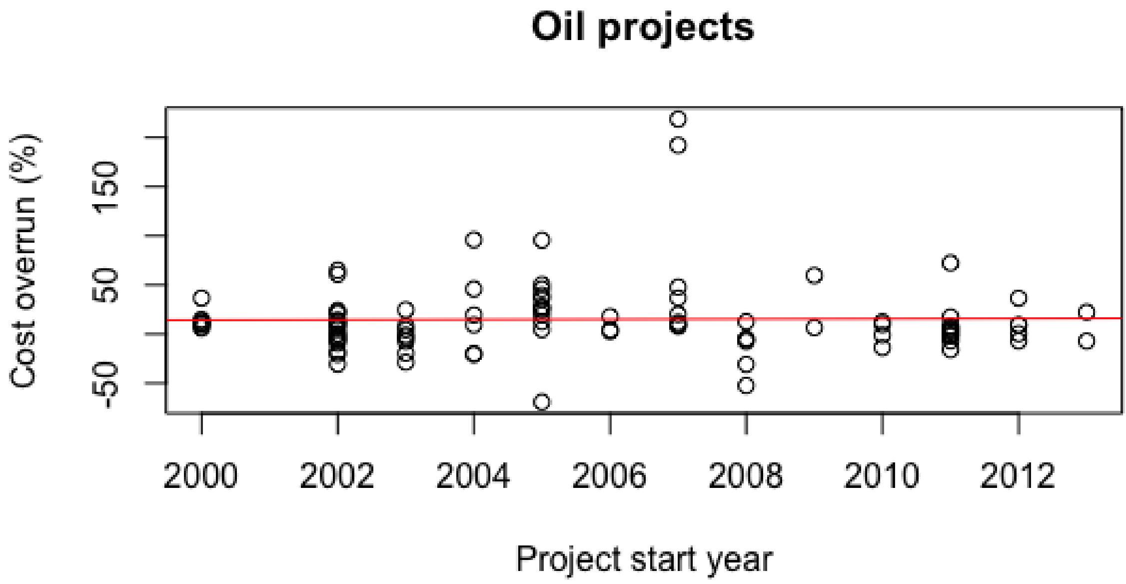

Figure 4.3: Oil projects start year univariate regression

Before and after the financial crisis

To further analyze the positive trend of cost overruns increasing over time (see above discussion), the dataset was further divided into two groups based on a projects year of completion is before or after the financial crisis. Projects completed in 2009 and later were put in the post financial crisis group by assigning a dummy variable taking the value 1 for projects completed in 2009 and later. As the financial crisis led to a sharply reduced oil price, and consequently sharply reduced revenues for oil companies, one would expect cost overruns to decrease, as an increased focus on cost performance seems to be a logical response to decreasing revenues. As the oil industry contribute largely to Norway's GDP, relative to other industries, an economic downturn in the oil industry may lead to ripple effects throughout society, consequently affecting the public projects in the dataset. Thus, by intuition, one would expect a decrease in cost overruns if the dataset were divided with respect to project completion before and after the financial crisis.

However, the univariate regression finds that cost overruns are significantly higher for projects completed after the financial crisis for oil projects and public projects with budget > 100 mNOK, with p values of 0,0345 and 0,00296 respectively. This seems to contradict logical reasoning, pointing towards other factors as more important for cost overruns. For the oil industry, is should be noted the NCS is maturing, meaning that more complex and remote oil fields are developed in order to maintain production levels, consequently demanding more complex projects which may increase cost overruns.

It is also possible that there is no linear relationship between accumulated experience and cost overruns, alternatively a decreasing marginal effect of more experience and not a cost overrun converging towards zero over time. One should also be aware of the motivation to lower the cost overruns. The assumption of project cost estimators trying to rationally maximize benefits may be naive. Projects may be delayed or accelerated without regards to cost overruns, as such decisions are a function of costs and benefits. E.g. accelerating projects to exploit a higher oil price may lead to increased cost overruns, but of negligible size compared to the possible increased profits margins gained from acceleration, as proposed by Lorentzen et al. (2017) Commodity prices are therefore highly relevant explanatory factors for further analysis. Another possible explanation may be unrealistically low estimated costs resulting from increased cost focus, however, this may not be concluded from the this analysis.

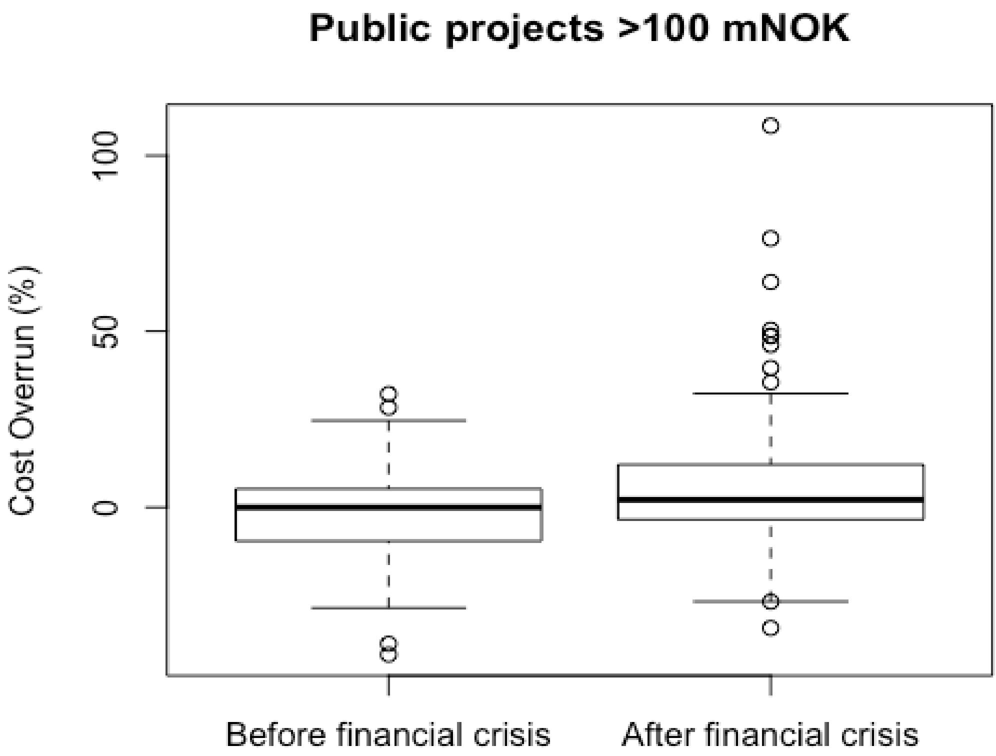

Figure 4.4: Box plot of cost overruns of public projects >100 mNOK before and after financial crisis

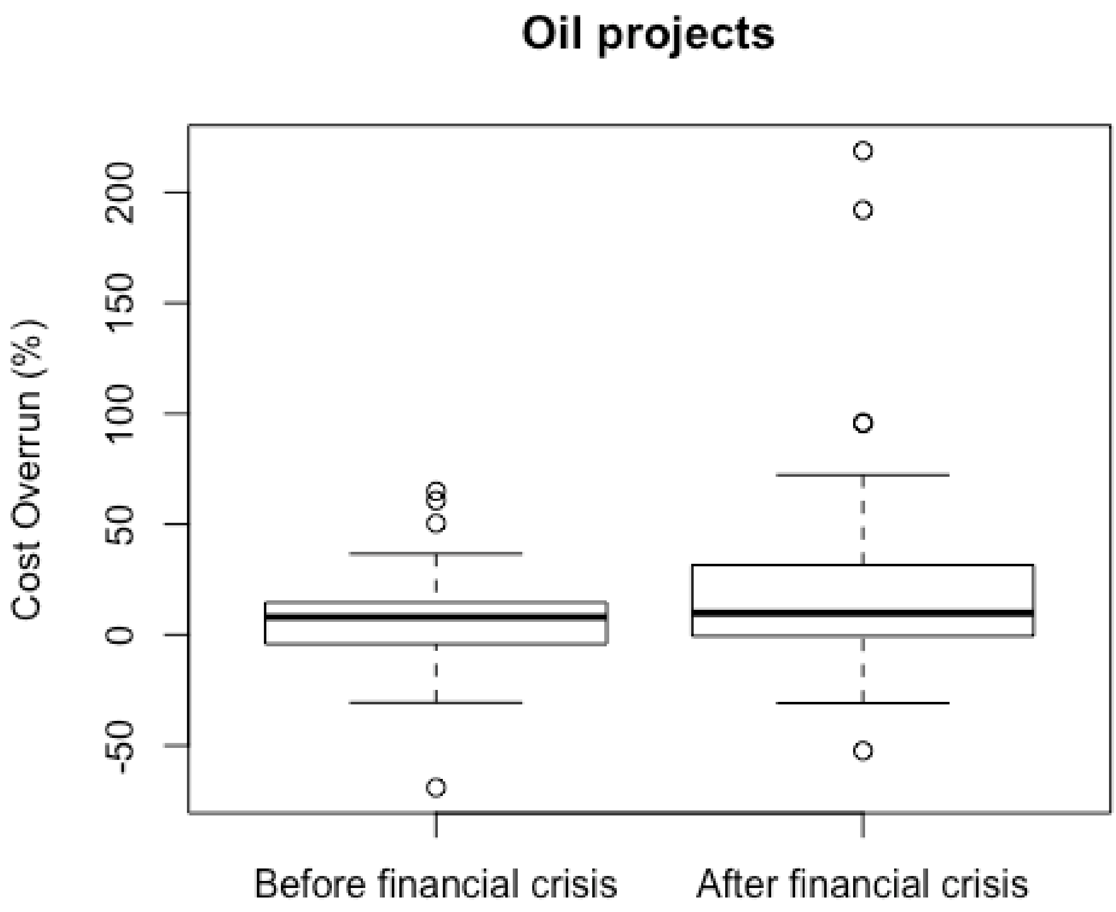

Figure 4.5: Box plot of cost overruns in oil projects before and after financial crisis

Budget size and project duration

For both budget size and project duration, a positive relation with cost overruns is expected. Larger budgets and longer duration indicates more complex projects, consequently increasing uncertainty, chances of delays and extra costs.

When considering all budgets for public projects, the budget size is significant. Although with a relatively small negative beta, the analysis indicates that smaller projects have a higher cost overrun, pointing towards too little emphasis on planning for small projects. However, it should be noted that slight cost underestimations of small projects would make more of an impact in terms of relative cost overrun. Consequently, it seems that public projects perform better in terms of cost overrun with increasing budgets. For oil projects to opposite relationship was found; slightly increasing cost overruns with increasing projects budgets, however this was not found to be significant.

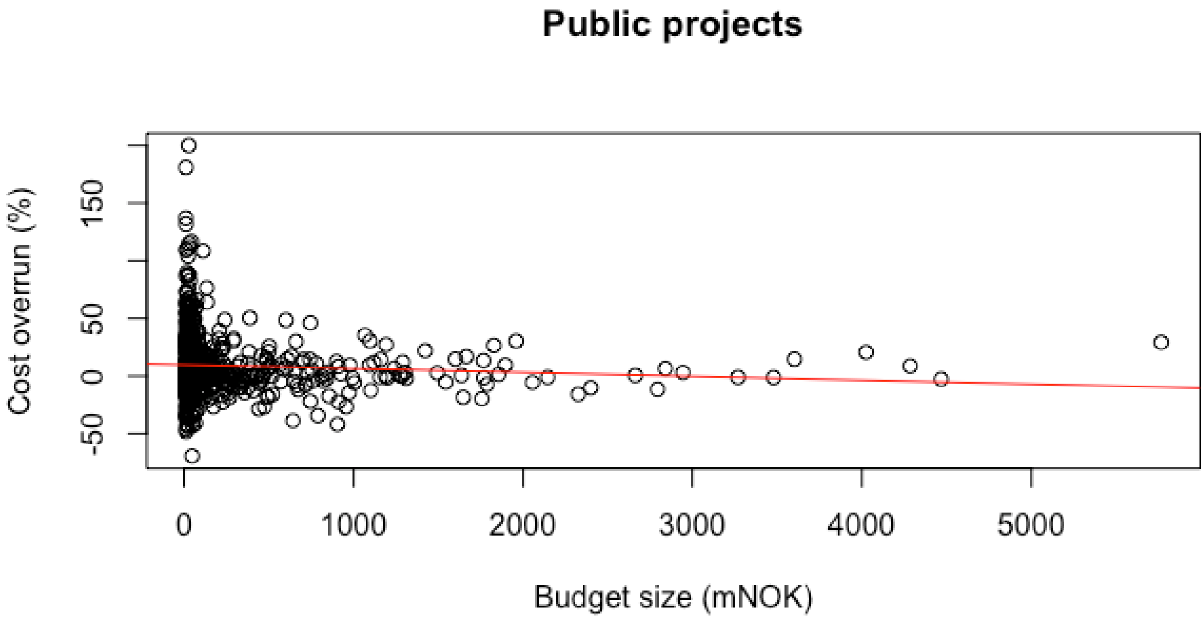

Figure 4.6: Public projects budget univariate regression

For project duration, the univariate regressions indicate that cost overruns in the oil industry are caused by lengthy projects, which is logical as discussed above. This complies with Flyvbjerg et al. (2004), where it was found that cost escalations in transport infrastructure projects were strongly dependent on the length of the project implementation phase. The same trend can be seen for public project, although not significant. This may be due to the average length of public projects are substantially shorter than for oil projects, averaging 1,45 years and 3,1 years respectively. The dataset contains limited observations of lengthy public projects, making it difficult to demonstrate a significant increase in cost overruns with increasing project length.

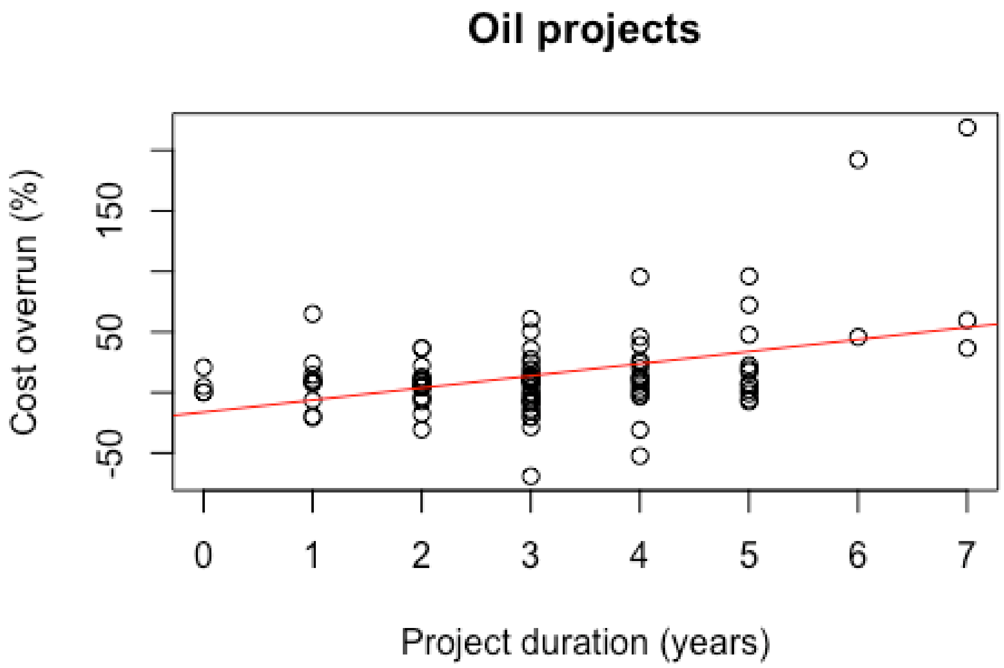

Figure 4.7: Oil projects duration univariate regression

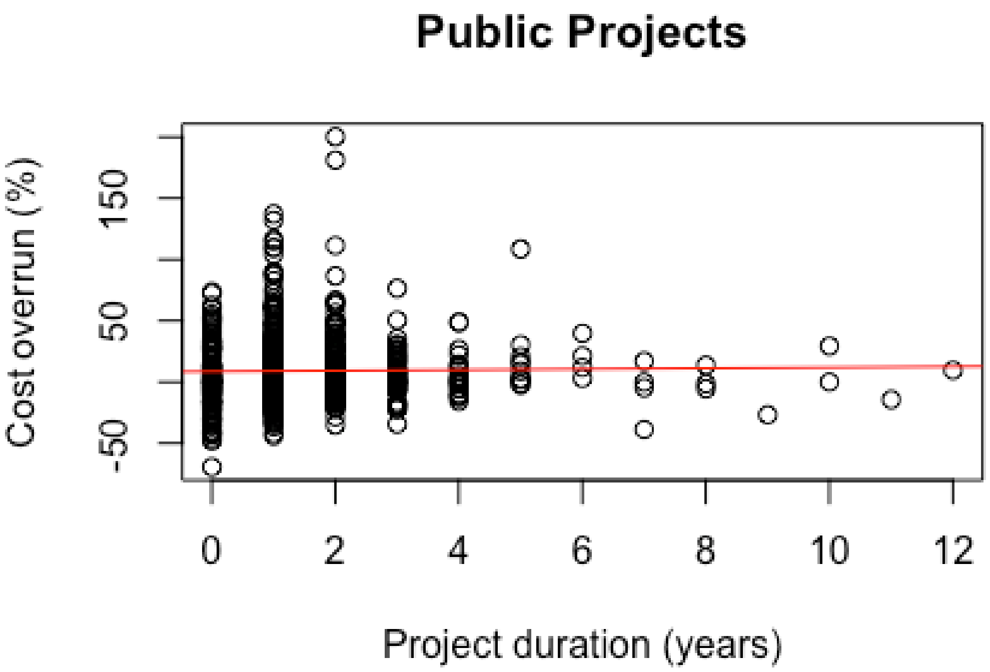

Figure 4.8: Public projects duration univariate regression

External quality assurance

As previously discussed, public projects with estimated costs exceeding 750 mNOK are required to undertake external quality assurance (shortened to KS, from Norwegian “kvalitetssikring”) as part of the government project model. (Regjeringen.no, 2014) However, some projects with lower estimated costs undertake the external quality assurance, with the lowest estimated costs of a KS-project in the dataset being 370 mNOK. The aim of this quality assurance is to increase the cost performance of public projects as a result of numerous public projects experiencing large cost overruns. Thus, one should expect the increased focus on cost estimation would result in a reduction in cost overruns.

By creating a dummy variable, taking the value 0 for-non-KS projects and 1 for KS-projects, the following analysis was done. The regression was run using cost overrun as the dependent variable and the KS dummy variable as the independent variable. By removing the smallest projects (estimated costs <100 mNOK) as these are too small in order to comparable with KS-projects, we fail to find a significant reduction in cost overruns. Although a negative trend is observed, which correspond to decreasing cost overruns with increasing estimated costs for public projects, we find no statistical evidence of the KS-process reducing cost overruns. The box plot below shows no significant differences in the cost overrun distribution for the two groups.

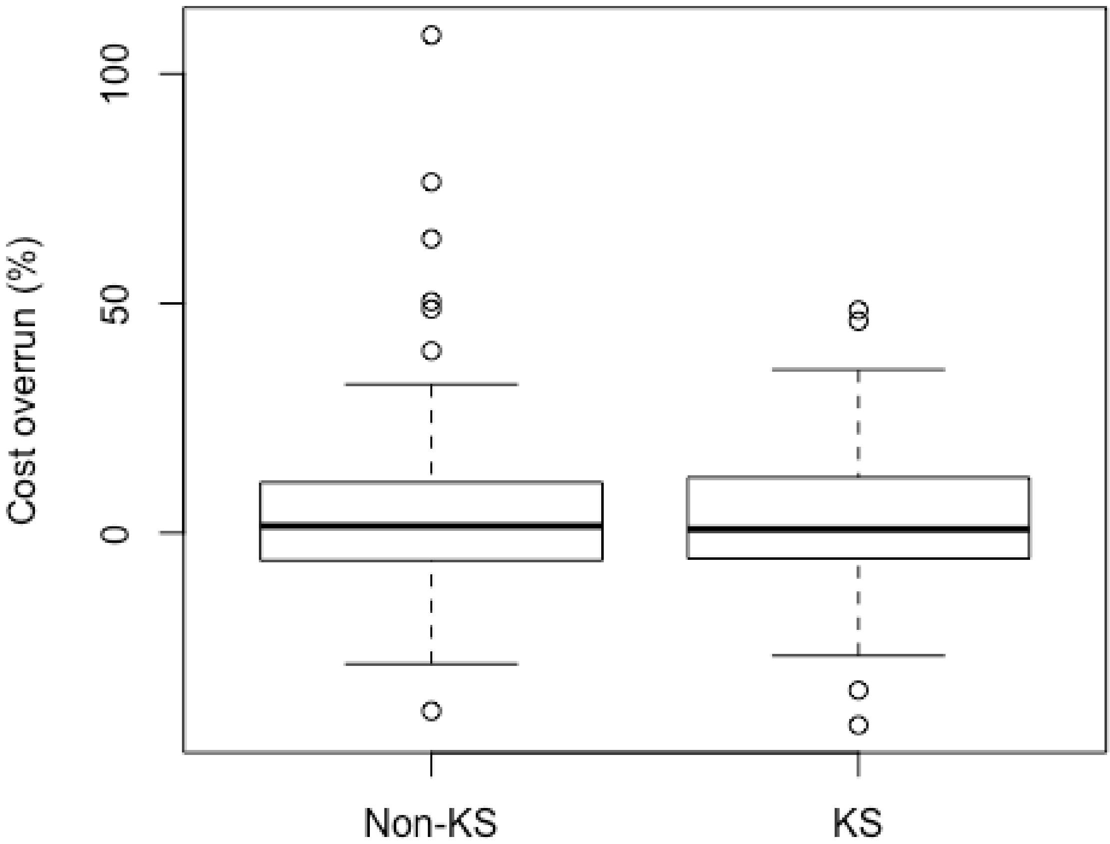

Figure 4.9: Box plot of cost overrun distribution of KS and non-KS projects

4.1.2 Macroeconomic variables

The estimated costs used in this thesis are estimated at the time of decision to invest or as close to this as possible. It is fair to assume that these estimates are based on the macroeconomic environment in order to produce realistic cost estimates. However, as projects are typically executed over several years, the macroeconomic environment is subject to change. A study by Olatunji (2010) studying the impact of oil price regimes on construction costs in Nigeria found several macroeconomic variables to have significant effects on construction costs, including inflation rate and GDP growth. In this analysis, we focus on macroeconomic variables that are volatile and somewhat logically related to our dataset, and thus may lead to surprises that in turn lead to cost overruns (Dahl et al., 2017). Some variables are related to the respective business cycles of oil projects and public projects, such as investment level and commodity prices. The other variables are related to the macroeconomic environment in Norway. All macroeconomic variables used are as follows: Oil Price, Steel Price, Norwegian CPI, Oil industry employment level, national employment level, Key Policy Rate, Pension Fund transfer rate, GDP growth rate, NCS investment level, Public investment level, total Norwegian exports and GDP from oil and marine activities. Annual averages are used for oil price, steel price, Key policy rate, Pension fund transfer rate, employment levels and GDP growth. Investments levels, Export and GDP from oil and marine activities are given as accumulated values for each year, while CPI is indexed. Changes in these variables may result in increased cost level and affect the availability of skilled workers etc. However, it should be noted that many of the macroeconomic variables included in this analysis are highly correlated, and one must therefore be careful when including these variables in a regression model to avoid multicollinearity. The public investment level is limited to the accumulated investment allocations for Jernbaneverket, Statsbygg and Statens Vegvesen, as these public entities dominate the dataset of public projects. A table of correlations between macroeconomic variables follows.

	Oil price	Steel price	CPI	Oil empl	Empl	KPR	Pension fund	GDP Growth	NCS invest	Public invest	Export	GDP oil/sea
Oil price	1											
Steel price	0,51	1										
CPI	0,51	0,73	1									
Oil empl.	0,67	0,79	0,91	1								
Empl.	0,56	0,75	0,96	0,96	1							
KPR	-0,56	-0,6	-0,78	-0,69	-0,66	1						
Pension Fund	0,09	0,01	-0,01	-0,01	-0,14	-0,3	1					
GDP Growth	-0,19	-0,2	-0,37	-0,37	-0,4	0,14	-0,21	1				
NCS invest	0,57	0,77	0,93	0,97	0,95	-0,73	-0,07	-0,29	1			
Public invest	0,22	0,6	0,92	0,96	0,87	-0,68	-0,12	-0,27	0,84	1		
Export	-0,31	0,08	0,28	0,07	0,2	-0,17	0,1	0,15	0,15	0,43	1	
GDP oil/sea	-0,7	-0,78	-0,91	-0,96	-0,97	0,67	0,16	0,39	-0,94	-0,76	-0,03	1

Table 4.3: Correlation matrix for macroeconomic factors

Data sources: Statistics Norway (SSB) and annual reports from public entities. Where CPI is the Consumer Price Index, Oil Empl is the employment level in the Norwegian oil industry, Empl is the Norwegian employment level, KPR is the Key Policy Rate, Pension fund is the transfer rate from the Pension fund to the Norwegian state budget, GDP Growth is the annualized growth rate of Norway's GDP (%), NCS Invest is the accumulated investments on the NCS, Public invest is the accumulated investments for public entities, Export is the total value of Norwegian exported goods and GDP oil/sea is the contribution to national GDP from oil and marine activities.

As shown in the correlation matrix, several variables are highly correlated as previously expected. We find the highest correlated variables to be the employment levels in the oil industry and on a national basis. This is to be expected, as the employees in the oil industry constitute a large percentage of Norwegian workers. The investment levels are also highly correlated as the oil industry has experienced strong growth while the government of Norway has increased the focus on developing infrastructure in the same period. Not surprisingly, oil related variables are strongly correlated, such as Oil industry employment level and investment level on the NCS, as well as the oil price and investment level/employment level. Logically, CPI and employment levels are highly correlated. High level of employment mean increased demand for goods and services, thereby increasing the general price level. There is also significant correlation between GDP oil/sea and most of the macroeconomic variables, which is expected due to the importance of the offshore industries to the Norwegian society.

To further analyze the effects on cost overruns by the selected macroeconomic variables, several univariate regressions were run. The dependent variable for all regressions is Cost overrun (%), while changing the independent variable between the selected macroeconomic variables. The results are presented in the table below. All variables were run for both public projects and oil projects in order to investigate the differences between the two. The value of the macroeconomic variables was set to the value they had at the respective projects' start year to simulate the macroeconomic environment at time of decision to invest.

Variable	Beta	Std. error	T	Level of significance, *p*	R2
Public projects					
Oil Price	0,08535	0,03437	2,483	**0,0132**	0,007538
Steel Price	-0,03656	0,02818	-1,298	0,19482	0,002069
CPI	-0,0023	0,1796	-0,013	0,99	2,019e-07
Oil Industry Employment Level	0,02245	0,08972	0,25	0,8024	7,713e-05
National Employment Level	0,006342	0,008421	0,753	0,452	0,000698
Key Policy Rate	0,08402	0,73252	0,115	0,909	1,62e-05
Pension Fund Transfer Rate	-2,309	1,098	-2,104	**0,0357**	0,005422
GDP Growth Rate	-0,08325	0,6447	-0,129	0,897	2,053e-05
NCS invest	0,001435	0,0217	0,066	0,94731	5,382e-06
Public invest	-0,1425	0,1728	-0,825	0,41	0.0008371
Export	3,546e-05	6,679e-05	0,531	0,596	0,0003471
GDP oil/sea	-2,349e-05	2,089e-05	-1,124	0,2512	0,001555
Oil projects					
Oil Price	0,1981	0,122	1,624	0,108	0,02788
Steel Price	-0,000807	0,1103	-0,007	0,994	5,80E-07
CPI	-0,2144	0,6663	-0,322	0,748	0,001124
Oil Industry Employment Level	-0,0427	0,3716	-0,115	0,909	0,0001435
National Employment Level	-0,005443	0,03286	-0,166	0,869	0,0002982
Key Policy Rate	-1,2	1,98	-0,772	0,4424	0,006429
Pension Fund Transfer Rate	-2,407	2,354	-1,023	0,3092	0,01124
GDP Growth Rate	8,3067	3,3242	2,499	**0,0142**	0,06356
NCS invest	0,01569	0,09789	0,16	0,873	0,000279
Public invest	0,739	1,007	0,734	0,465	0,005818
Export	3,67e-04	2,643e-04	1,391	0,168	0,02059
GDP oil/sea	-9,563e-06	7,437e-05	-0,129	0,898	0,0001797

Table 4.4: Univariate regression results of macroeconomic variables

As shown in the table above, we find the oil price and the Pension Fund transfer rate to be significant in terms of cost overruns for public projects, both significant at 5% level. The beta values indicates that each dollar the oil price increases, public projects experience a 0,085% increase in cost overruns on average. As for the Pension Fund transfer rate, a 1% increase in the transfer rate decreases public projects' cost overrun by 2,3% on average. As shown in the correlation matrix, the correlation coefficient between the two variables is 0,09, indicating a very low level of correlation. Increasing cost overruns with rising oil price could indicate lack of competent workers, as the oil industry may seem more attractive to workers due to higher wages. The beta of Pension Fund transfer rate is interesting. From the correlation matrix we see that GDP growth and Pension Fund transfer rate is negatively correlated (correlation coefficient -0,21). This means that a decline in the Norwegian economy leads to greater funding for public projects in order to stimulate the economy and secure jobs. The negative beta of Pension Fund transfer rate indicates that public entities estimate costs more precisely in times of recession.

For oil projects, we find the GDP growth rate to be significant at 5% level. For every percent the GDP grows, oil projects experience an increase in cost overruns of 8,3% on average. GDP is directly linked to the general level of activity in Norwegian industries, and may therefore be a good proxy for employment levels, pressure on supplier markets and general price level. This result is therefore logically intuitive, as increased GDP would result in cost-inflating factors.

The above analysis is based on the macroeconomic variables being set to their respective value of the year of project start. In order to eliminate the dependence on the start and end year of projects, the same univariate regressions were run, instead using the relative change in the macroeconomic variables during the project execution period. GDP is excluded from this analysis, as GDP is specified in annual growth rate instead of the annual value to prevent substantial correlation with the other variables. These variables are now called surprise variables, and represent changes in the macroeconomic environment. Changes in the macroeconomic environment may challenge the assumptions that were made when estimating costs. Hence, if there have been surprisingly large or unexpected movements in macroeconomic variables during project execution, these may affect the size of cost overruns. The regression results from this analysis in presented in the following table.

Surprise Variables	Beta	Std. error	T	Level of significance, *p*	R2
Public projects					
Oil price surprise	-0,02461	0,01921	-1,281	0,201	0,002017
Steel Price surprise	-0,008149	0,025598	-0,318	0,75	0,0001248
CPI surprise	0,4028	0,2167	1,272	0,204	0,001988
Oil Industry Employment surprise	0,03916	0,07181	0,545	0,586	0,000366
National Employment surprise	0,02372	0,35649	0,067	0,947	3,582e-05
Key Policy Rate surprise	-0,01904	0,02262	-0,842	0,4	0,000872
Pension Fund transfer rate surprise	0,10362	0,03971	2,609	**0,00924**	0,008314
NCS investment surprise	-0,02622	0,03684	-0,712	0,477	0,0006233
Public investment surprise	0,033	0,05277	0,625	0,532	0,0004812
Export surprise	-1,3256	0,4912	-2,699	**0,0071**	0,00889
GDP oil/sea surprise	-0,1073	0,1973	-0,544	0,587	0,0003641
Oil projects					
Oil Price surprise	-0,08338	0,05552	-1,502	0,136597	0,02393
Steel Price surprise	0,1136	0,1007	1,128	0,2622	0,01364
CPI surprise	5,1107	0,9791	5,22	**1,1e-06**	0,2285
Oil Industry Employment surprise	0,7296	0,2146	3,399	**0,001**	0,1116
National Employment surprise	3,665	1,094	3,35	**0,00117**	0,1087
Key Policy Rate surprise	-0,02831	0,06372	-0,444	0,657843	0,002141
Pension Fund Transfer surprise	0,2127	0,159	1,337	0,184357	0,01907
NCS investment surprise	0,4523	0,1468	3,081	**0,00272**	0,09352
Public investment surprise	0,3287	0,1666	1,972	0,0516	0,04056
Export surprise	-2,474	1,332	-1,858	0,0663	0,03617
GDP oil/sea surprise	-1,4270	0,5288	-2,699	**0,002828**	0,07335

Table 4.5: Univariate regression results of macroeconomic surprise variables

For public projects, the pension fund surprise variable turns out significant. This was to be expected due to the results in univariate regression, and the same reasoning applies. As for the Export surprise variable, the negative beta implies that cost overruns in public projects are reduced during periods which the export from Norway increases. As increasing exports during the periods which the dataset covers are strongly related to the upturn in the oil industry, this indicates that public projects perform better, in terms of cost performance, in times of high activity in the oil industry.

We see that for oil projects, the CPI surprise, Oil Industry Employment surprise and National Employment surprise are all highly significant at 5% level, They all have great explanatory power (22,85%, 11,16% and 10,87%, respectively). However, all these variables are highly correlated as shown in the correlation matrix. In times of high employment levels, people have greater purchasing power, which in turn increases demand and prices for goods and services as well as wage levels. CPI is a good proxy for overall inflation in Norway. However, CPI is adjusted by the Norwegian Central Bank and has been stable and predictable about the goal of 2% yearly inflation over long period of time. It therefore makes no sense to include the CPI surprise variable in the multivariate regression models, although the explanation power of the variable from the regression output is rather high. As CPI and Oil Industry Employment Level is highly correlated, this makes it possible to include the Oil Industry Employment surprise variable further in the multivariate regression analysis. This variable is given as the relative difference between numbers of employees in the Norwegian petroleum sector from the start to the end of the project. With increasing numbers of employees, one would expect the talent pool to be exhausted, ultimately leading to lower-skilled workers to be employed, in turn reducing efficiency and inflating cost overruns. This theory is backed by the univariate regression, with a 1% increase in numbers of employees leading to a 0,7% increase in cost overruns on average. As workers in the Norwegian petroleum industry accounts for a relatively large proportion of the national workforce, as well as about 0,5% higher explanatory power, the surprise variable for oil industry employment seems to be a better explanatory variable for cost overruns in oil projects. We also find the NCS investment and GDP from oil and marine activities surprise variables to be significant for oil projects, with NCS investment increasing cost overruns and GDP from oil and marine activities decreasing cost overruns. The opposite effects are as expected, as the two variables are almost perfectly negatively correlated with a correlation coefficient of -0,97. However, the negative beta of GDP oil/sea surprise may not seem intuitive. For the time period the dataset covers, we only see major changes in this variable at two occasions; the financial crisis and the downturn in the petroleum industry around year 2014. Hence, the dataset contains very limited number of observations during these times of recession, and one must therefore interpret this result with caution.

4.2 Multivariate regression

The following section presents the results from the multivariate regressions.

4.2.1 Oil projects

Based on the correlation matrix of the variables selected for examination, there is widespread correlation between the variables. Therefore, it is not feasible to combine all significant variables in multivariate regression models, as severe multicollinearity would be the result. A regression using exhaustive selection is used, allowing for logical reasoning to supplement the selection of variables. Given the described methodology for selection of the optimized multiple regression models, the following tables summarize the optimized specification of the multivariate regression models.

Number of variables	Variables					Model fit	
	Duration	Brent surprise	Pension fund surprise	GDP growth	NCS investment surprise	R2	AR2
1	X					0,1624	0,1533
2	X			X		0,232	0,2151
3	X	X		X		0,2627	0,2382
4	X	X		X	X	0,2997	0,2683
5	X	X	X	X	X	0,3244	0,286

Table 4.6: Multivariate regression by iteration process for oil projects

Where duration is the execution time of the project, Brent surprise is the relative change in brent oil price during project execution, Pension fund surprise is the relative change in pension fund transfer rate during project execution, GDP growth is GDP growth rate the year of project start, NCS investment surprise is the relative change in NCS investments during project execution.

The above table shows the results from the LEAPS function for oil projects. By adding more than 5 variables, the fit of the model (both R2 and adjusted R2) increases by less than 1%. By adding variables beyond this point, the model is deteriorated as more variables explain the cost overruns, without the explanatory power of the model increasing significantly. Which of R2 and adjusted R2 yields the most accurate result is somewhat discussed. The main difference is that AR2 penalizes more variables in the model; however, it does not necessarily provide more accurate numbers. Therefore, both will be reported in this thesis.

	Beta	**Std. Error**	**T**	**Level of significance, *p***
(intercept)	-29,251	9,54794	-3,067	0,00287
Duration	6,55128	2,75196	2,389	0,01902
Brent surprise	-0,15197	0,05488	-2,769	0,00686
Pension fund surprise	0,25112	0,14023	1,791	0,07676
GDP growth	7,12523	2,92377	2,437	0,01682
NCS invest. surprise	0,47093	0,1863	2,528	0,01326

Table 4.7: Multivariate regression specifications for oil projects

We see from the table above that all variables are statistically significant at 5% level except the Pension fund surprise variable. First, we find that cost overruns increases with the duration of projects equal to 6,5% per year of project execution time. This is consistent with the result of the similar univariate regression, and is also a logical effect since longer duration may hamper coordination, logistics and predictability in general. This is also the variable contributing the most to the model fit with an R2 and adjusted R2 of 16,24% and 15,33%, respectively. More interestingly, we find that changes in the oil price during project execution tends to reduce cost overruns by 0,15% for every 1% the oil price increases. This result contradicts the findings of similar research that often find that an increase in oil price also increases cost overruns, for example Dahl et al. (2017). This may be due to limited number of observations from the years of low oil price, especially during the financial crisis and the drop in oil price around 2014. With the limited dataset during these periods, one may simply not capture the true effect of oil price changes. The periods of recession are limited in duration, and are short relative to the time any restructuring of the industry may take. It may therefore be logical to moderate this model by removing this variable, as discussed in part 3 Method.

	Beta	Std. Error	T	Level of significance, *p*
(intercept)	-32,0573	9,8327	-3,26	0,00158
Duration	8,17	2,7774	2,942	0,00416
Pension fund surprise	0,2627	0,1453	1,807	0,07408
GDP growth	8,1773	3,0056	2,721	0,00783
NCS invest. surprise	0,2262	0,17	1,33	0,18686

Multiple R-squared: 0.2655, Adjusted R-squared: 0.2325

Table 4.8: Moderated multivariate regression specifications for oil projects

The table above shows the specifications of the multivariate regression for oil projects. As expected, the explanatory power of the model has decreased 5-6% (considering both R2 and adjusted R2) due to the removal of the oil price surprise variable. The moderated model is not necessarily inferior, on the contrary, since it has been moderated based on logical reasoning. Besides, the oil price and NCS investment variables are positively correlated with a correlation coefficient of 0,57. Thus, by removing the oil price variable, we also reduce the problems with multicollinearity in the model.

We now find the NCS investment surprise variable insignificant in terms of its p-value, however, it still contributes in terms of both logical and statistical explanatory power. One reason for this effect may be increased pressure on the supply market, as the increased investment level may lead to a scarcity of supply capacity. Previous studies, such as Osmundsen et al. (2015) find that times of economic boom in the Norwegian oil industry sharply increases expenditures such as rig rates and material costs, which may relate to the findings regarding the NCS investment surprise variable. The NCS investment variable is also almost perfectly positively correlated with the Oil industry employment variable. Increasing investments leads to more jobs, and correspondingly the demand for skilled labor increases. One explanation may be that as the number of employees in the oil industry increases, it becomes gradually harder to recruit the necessary amount of skilled workers, and therefore marginal productivity and efficiency per worker will decrease. One should also note that the wages in the Norwegian oil industry has increased significantly relative to other sectors with the economic upturn in the oil industry after the financial crisis until the downturn in 2014. Consequently, increasing investments may lead to a reduction in productivity and efficiency, as well as an increase in cost levels, resulting in inflated cost overruns in oil projects.

For the variables GDP growth and Pension fund surprise, the same reasoning goes as for the GDP growth variable from the univariate regression on macroeconomic variables.

4.2.2 Public projects

By performing similar analysis for public projects, problems arise. We have previously seen, in the univariate regression results, that small projects offer the greatest variation in cost overruns in public projects. The distribution of cost overruns as a function of budget size shows an inverted funnel-shape, indicating that the variability of the cost overrun variable is non-constant across the range of observations, so-called heteroskedasticity. The variability is clearly greater in low-budget projects than for large projects. Existence of heteroskedasticity is a major problem when applying regression analysis on a dataset. The non-constant variability makes it difficult to adapt a linear regression model on the dataset, and it may therefore invalidate statistical tests of significance. This breaks with one of the fundamental assumptions of classic linear regression that there is no heteroskedasticity present in the observation. As a result, the Gauss-Markov theorem does not apply, meaning that OLS-estimators are no longer BLUE (Best Linear Unbiased Estimators) (Wooldridge, 2009). This problem manifests itself when similar analysis was carried out on public projects. The table below shows the output from the LEAPS function for public projects.

Number of variables	Variables						Model fit	
	Pension fund surprise	Steel price	Export	GDP Oil/Sea	Budget	Duration	R2	AR2
1	X						0,008314	0,007093
2	X	X					0,01386	0,01143
3		X	X	X			0,02986	0,02627
4		X	X	X	X		0,03367	0,0289
5		X	X	X	X	X	0,03894	0,03299

Table 4.9: Multivariate regression by iteration process for public projects

The table shows the significant lack of explanatory power for the models with one to five independent variables, with the five-variable model explaining only about 3,3% of the variability in the dataset. As a result of non-constant variance, we fail to fit a multivariate linear regression model with high explanatory power, what is called a type II error. The type II error means that we fail to reject the null hypothesis; in this case we fail to find a relationship between cost overruns and the selected variables, when the null hypothesis was actually uncharacteristic of the population. As we can see from figure 4.6, we find that we may avoid the non-constant variability by removing projects with estimated costs less than 100 mNOK, as this interval experiences larger variability in cost overruns compared to the rest of the dataset. As a result, we avoid the problem of heteroscedasticity as the variability in cost overruns is now more constant, and one can thus fit a linear regression model that does not violate the basic assumptions the linear regression. In addition, projects with low estimated costs are likely to be executed over a relative short period of time. As the levels of macroeconomic variables are defined as yearly averages, it is likely that no effect can be observed on cost overruns due to changes in macroeconomic levels as the execution time of smaller projects is simply too short. Hence, it is reasonable to omit these projects from the following analysis. The table below summarizes the leaps output now that only projects with estimated costs >100 mNOK was considered.

Number of variables	**Variables**							**Model fit**	
	Pension fund surprise	Employment surprise	Employment	Export	GDP Oil/Sea	Steel price	Duration	R2	AR2
1	X							0,0306	0,02613
2	X	X						0,06467	0,05646
3			X	X	X			0,1104	0,09869
4		X		X	X	X		0,1459	0,1308
5			X	X	X	X	X	0,1624	0,1438

Table 4.10: Multivariate regression by iteration process for public projects >100 mNOK

By removing projects with estimated costs <100 mNOK, we find that the heteroskedasticity is greatly reduced as the models have greater explanatory power. We find that by adding more than 5 variables to the model, the explanatory power increases by less than 1%; hence, the 5-variable model is the optimal choice when trying to model the cost overruns in public projects. The model with 5 variables gives an R2 of 16,24% and adjusted R2 of 14,38%, indicating between 16,24% 14,38% of the variability of cost overruns in public projects may be explained the following variables: public employment level at the time of decision to invest, the total Norwegian export, GDP from oil and marine activities and project duration. However, with only 14,38% to 16,24% of the variability in cost overruns explained by the model, the proposed model fails to find a strong correlation between the selected variables and cost overruns in public projects. Considering the model contains five variables, the contribution to explanatory power from each variable is rather small. This indicates that, due to the wide selection of macroeconomic variables, more research is still required to capture the drivers of cost overruns in public projects. Such variables may be project specific variables, which was not possible to gather for this thesis. Logically, one would expect cost overruns to be a function of several project specific variables such as uniqueness of a project, technical complexity, access to necessary materials and services and owner as well as scope changes as discussed earlier. Besides, the low explanatory power may indicate that Flyvbjerg's categorized cost overrun drivers play a role, especially optimism bias and strategic misrepresentation as proposed earlier in this thesis.

	Beta	**Std. Error**	**T**	**Level of significance, *p***
(intercept)	4,683e01	1,616e02	0,29	0,77225
Empl	-0,1306	0,04521	-2,889	0,00424
Export	6,103e-04	1,126e-04	5,42	1,53e-07
GDP oil/sea	-5,37e-04	1,103e-04	-4,867	2,13e-06
Steel price	-0,1836	6,987e-02	-2,628	0,00918
Duration	1,553	0,6276	2,474	0,0141

Table 4.11: Multivariate regression specifications for public projects

However, with that being said, some of the variables selected for the multivariate regression model are rather intuitive to interpret, and points towards that the model explains some of the variability. The above table summarized the model specifications for public projects. All variables selected in the model are significant at 5% level. We see that the national employment rate variable has a small negative beta, indicating a small reduction in cost overruns in times when the unemployment rate in Norway is low. Opposed to oil projects, which experience the opposite effect, it seems that public entities have no problems with recruiting skilled labor for their projects.

The steel price variable may not seem intuitive with respect to its effect on cost overruns. Like the oil price, there have been limited drops in steel price for the time period covered by the dataset. Hence, the number of observations from times with low steel prices may be too limited to capture its true relationship with cost overruns, similarly with the reasoning behind the negative beta of oil prices on oil projects. Consequently, this variable may be removed from the model, as the logic behind including it is rather debatable.

	Beta	Std. Error	T	Level of significance, *p*
(intercept)	1,845e02	1,549e02	1,191	0,234795
Empl	-1,556e-01	4,477e-02	-3,476	0,000609
Export	4,663e-04	9,964e-05	4,68	4,95e-06
GDP oil/sea	-4,785e-04	1,095e-04	-4,371	1,89e-05
Duration	1,662	6,344e-01	2,62	0,009386

Multiple R-squared: 0.1367, Adjusted R-squared: 0.1214

Table 4.12: Moderated multivariate regression specifications for public projects

The above table summarizes the multivariate regression specifications for the moderated model for public projects. As for the moderated model for oil projects, the explanatory power is somewhat reduced. However, the model now contains one less variable, and is therefore more logically sound.

Like in oil projects, we find that project duration contributes to increased cost overruns, significant at 5% level. Unlike oil projects, which on average exceeds estimated costs by about 8% each year of the project execution, public projects are not as exposed to the duration of projects. For public projects, each execution year of a project increases cost overruns by 1,5% on average. This may be seen in connection with the explanatory power of the different regression models above. As the model for oil projects has higher explanatory power, oil projects are consequently more prone to cost overruns due to changes in the macroeconomic environment. Hence, longer project duration may therefore allow for more significant changes in the macroeconomic environment to occur, causing a higher beta for the duration variable for oil projects than for public projects. Besides, one should be aware of that the average duration of public projects is rather short compared to oil projects, as discussed earlier.

With this being said, the explanatory power of the public model is rather small, with an R2 and adjusted R2 of 13,67% and 12,14% respectively. Considering that the model contains four variables, the explanatory power of the model is rather low. Therefore, we fail to find a significant relationship between cost overruns in public projects and the available variables.

4.3 Comparison of public and oil projects

The findings of the previous analysis points towards similarities in cost overruns between the two sectors. Firstly, both sectors experience increasing cost overruns with increasing project duration. In addition, I find that cost overruns of public projects are affected by the activity in the oil industry, as indicated by the variable GDP oil/sea, although the beta of this variable is approximately zero. This confirms the hypothesis that the cost overruns in the different sectors are to some degree affected by each other. The next step in this analysis is therefore to compare the different sectors more directly. This is done by running the models found in the previous sections on the opposite dataset, e.g. running the oil project's model on the public project's dataset, and vice versa. This analysis opens up for discussing to what extent which variables affect different project types, that is, oil and public projects. The below tables summarizes the results when the models were run on the opposite dataset.

	Beta	Std. Error	T	Level of significance, *p*
(intercept)	-40,47	2,281e02	-0,177	0,8596
Empl	1,173	1,443	0,813	0,41835
Export	-2,935e-05	2,777e-04	-0,106	0,91605
GDP oil/sea	1,012e-04	7,735e-05	1,308	0,19418
Duration	9,772	3,276	2,983	0,00368

Multiple R-squared: 0,1899, Adjusted R-squared: 0,1535

Table 4.13: Multivariate regression for oil projects using variables of public projects model

	Beta	Std. Error	T	Level of significance, *p*
(intercept)	-0,95061	2,48633	-0,382	0,7026
Duration	3,76038	1,40524	2,676	0,008
Pension fund surprise	0,09559	0,03944	2,424	0,0161
GDP growth	0,40321	0,79083	0,51	0,6106
NCS investment surprise	-0,1468	0,070602	-1,931	0,0547

Multiple R-squared: 0,0644, Adjusted R-squared: 0,04784

Table 4.14: Multivariate regression for public projects using variables of oil projects model

As expected, since the duration-variable was selected for both models in the previous section, we find this variable to be significant when running the models on the opposite datasets. For oil projects, no other variable is significant using this method. And, the explanatory power of all the variables combined is weaker than the explanatory power of the duration variable alone (14,48% and 16,24%, respectively), indicating no casual relationship between cost overruns in public and oil projects. The same finding goes for public projects when using the oil project's model variables. However, the Pension fund surprise variable turns out significant at 5% level. Interestingly, this variable was found significant in univariate regression models, but not selected further in multivariate regression. From the correlation matrix, we see that this variable is hardly correlated with any of the other variables, indicating that both project groups' cost overrun is affected by this variable. However, the explanatory power of the model is rather negligible with and adjusted R2 of only 4,7%. Both models suffer from too many independent variables with no real impact on the dependent variable.

In summary, this analysis shows that there is some correlation between cost overruns in the two sectors, as the GDP oil/sea variable is significant for public projects. However, as we see from the results when the models were run on the opposite dataset, we fail to find a simple relation between cost overruns in the two sectors.

5 Conclusion

In this thesis, I study cost overruns in the Norwegian oil industry and public sector. Although there are limited publications on cost overruns in Norwegian projects, the findings in this thesis largely correspond with similar studies in other geographical regions. The findings support the studies of Flyvbjerg et al. (2002), Ernst & Young (2013) and Merrow (2011), which states that cost overruns is a global phenomenon without dependence on geographical regions. Norwegian projects have, equally to projects in other countries, a positive skewness and kurtosis in the distribution of cost overruns. Similarly to Sovacool et al. (2014) and Flyvbjerg et al. (2002), we find no evidence of a learning effect, as the magnitude of cost overruns does not decrease significantly over time, as one would expect, as accumulated knowledge and experience would logically reduce cost overruns over time. The findings in this thesis also correspond with other quantitative and econometric studies on the subject. Dahl et al. (2017) found that employment level in the oil industry and duration to significantly inflate the magnitude of cost overruns in oil projects. As employment level in the oil industry is strongly correlated to the investment level, which I found to significantly increase cost overruns, the findings of my thesis are consistent with other literature on the topic. I also find the duration of projects to increase cost overruns for public entities alike, as suggested by Flyvbjerg et al. (2002), who concludes that projects in general basis experience this effect.

As stated above, my findings clearly correspond with the existing publications on cost overruns. However, since there is a void in the current literature regarding comparative studies such as this thesis, my findings bring new considerations regarding cost overruns. Even though the oil industry and public entities have different incentives for cost performance, it is still in the interest of both to minimize cost overruns. Thus, one may argue that the two are comparable, although the incentives are somewhat different. When compared statistically, I find that the oil industry experiences significantly larger cost overruns, with a positive skewness in the distribution of cost overruns of 2,71 versus 1,9 for public projects. This emphasizes the claim in media, as well as the general perception of the people, of a lower cost focus in the Norwegian oil industry compared with other industries. However, one should be aware of possible pitfalls when comparing projects from different project owners. Firstly, as previously discussed, the time of estimation is crucial, as there may be differences in terms of available details and degree of uncertainty of a project at different estimation stages. Secondly, there may be differences in estimation methodology. This may vary from purely deterministic methods to a more stochastic approach when estimating costs and risk factors. And lastly, the different sectors may use different indices and methodology for price adjustments (Welde, 2017).

To further analyze the differences between oil and public projects, several multivariate regressions was run. Multiple regressions enable the study of the ceteris paribus effect of the variables, that is, observing the relative changes in cost overruns by varying one variable at a time. By utilizing an exhaustive method for multivariate model selection, I arrive at the following models for modeling cost overruns in the two sectors:

$$Cost\ overrun_{OIL} = \alpha + \beta_1 Duration + \beta_2 Pension\ fund\ surprise + \beta_3 GDP\ growth + \beta_4\ NCS\ investment\ surprise$$

Equation 4: Model of cost overruns in oil projects

$$Cost\ overrun_{PUBLIC} = \alpha + \beta_1 Duration + \beta_2 Employment\ level + \beta_3 GDP\ oil/sea + \beta_4\ Export$$

Equation 5: Model of cost overruns in public projects

Cost overruns in public projects haven proven to be harder to model as they seem to be less affected by the selected variables. I find the explanatory power of the models to be 23,25% - 26,55% for oil projects and 12,14% - 13,67% for public projects. The explanatory power of the selected models above indicates that more research is required in order to fully reveal the dynamics of cost overruns in both sectors. Especially for public projects, where the low explanatory power of the model for public projects indicates that I fail to find a simple model for estimating cost overruns in these projects. This corresponds with the findings of Welde (2017), who also concludes with the need of more project specific variables in order to provide better quantitative analysis on cost overruns. However, considering cost overruns as the unexpected inflation of costs during project execution, any explanatory power may be considered noteworthy. These finding lend support for two of the proposed research questions. Firstly, we find a significant difference in the magnitude of cost overruns between the two sectors. Secondly, the hypothesis that the macroeconomic environment affects cost overruns is confirmed by the fact that several macroeconomic variables are chosen by the exhaustive multivariate model selection. However, I fail to find strong evidence that the cost overruns in the two sectors are affected by each other. Although the variables GDP oil/sea and Export is selected for the public model, which indicates that public projects are somewhat prone to cost overruns due to the business cycle of the oil industry, the explanatory power of the models where the opposite dataset is run on the selected model are negligible.

The explanatory power of the models indicates that additional project specific variables are needed to further improve the models. Such variables may be including changes in scope and experienced delays during project execution and the experience level of project managers and workers. Other variables include the characteristics of the project owners, technical complexity and uniqueness of projects, contract form and supplier relations.

Both oil and public project experience increasing cost overruns with increasing project duration. Project size and duration are obviously good proxies for the complexity of projects, and thus the uncertainty associated with estimation of costs. Hence, long duration projects increases the likelihood of changes in the macroeconomic environment during project execution, which may break with the assumptions made when estimating costs. This emphasizes the importance of the project manager’s ability to focus on planning of complex projects. In order to improve the estimation of costs, it may be advantageous to break down large projects in a number of subprojects. This allows for more precise estimation and reduction of uncertainty in each subproject, consequently obtaining a more realistic overall picture of the uncertainty of a project. It also emphasizes the importance of proper sensitivity analysis with a realistic picture of the probability and impact of risk factors, as pointed out by Flyvbjerg et al. (2002). As suggested in this study, strategic misrepresentation and optimism bias are potentially major problems in project management. The findings in this thesis indicate the presence of strategic misrepresentation and optimism bias when estimating costs, and thus political and psychological issues in project management. For companies and project managers in general, the fact of carrying out a project represents an intrinsic value. A result of this may be strategic underestimation of costs and correspondingly overestimation of profits and benefits in order to get a project approved and executed. It may also interfere with the ability of forecasters to evaluate the probability and impact of risk factors. There are, however, measures that may be implemented in order to reduce these issues. An obvious solution is to provide external quality assurance of estimates, such as in the public sector, where independent consultants are hired to evaluate whether project plans and estimates are realistic before execution. This method may reduce the impact of both optimism bias and strategic misrepresentation. Other solutions may be implementing additional incentives and penalties related to the accuracy of estimates. Due to the complexity of oil projects, it may be difficult to establish a complete picture of risk factors and their respective impact, and it may therefore be discussed whether introducing larger buffers in cost estimates would compensate for underestimation of risk. However, as previously discussed, further research on the effect of contract forms and project complexity is needed to further evaluate potential measures.

As observed by Dahl et al. (2017), cost overruns oil projects are dependent on the business cycle of the oil industry. This is emphasized by the findings in this thesis, as the multivariate model for oil projects selected several variables related to the business cycles of the oil industry and the Norwegian economy in general. Both the Pension fund surprise and GDP growth variables relate to the economic climate in Norway. The NCS investment variable is a good indicator for the business cycle of the oil industry, as increasing investments is strongly correlated with the activity level in the industry. As discussed earlier, during times of economic boom in the Norwegian oil industry, rig and material costs are inflated, wages increases rapidly and companies may experience lack of available skilled personnel. Consequently, the prices increase while efficiency decreases, as shown by Osmundsen et al. (2015). This emphasizes the importance of analyzing macroeconomic conditions as a part of the project-planning phase. More specifically, the project owners must carefully monitor the business cycle and predict future trends. One should therefore be aware of the escalation of costs during boom periods. This also relates to recruitment patterns. During times of economic boom, recruitment frequency is sharply increased. In order to avoid scarcity of personnel, project owners must carefully assess their future need of personnel and recruit in times of lower activity. The companies should also be careful when downsizing in low activity periods to avoid scarcity of key personnel during future upturns. As availability of personnel is considered a local business cycle indicator, oil companies should therefore monitor the activity of other oil companies in order to estimate the future activity level. Thus, we find that macroeconomic analysis is an important part when planning projects, consequently increasing the accuracy of estimates and providing a better basis for investment analysis.

5.1 Areas of future research

The results in this thesis correspond largely with existing publications on the topic, but also provide new insight regarding differences in cost overruns in the public sector and oil industry in Norway.

As mentioned, a major limitation of this research is the lack of project specific variables. Such information is required in order to further reveal the full dynamics of cost overruns in both sectors. This includes variables as the degree of planning done before project execution, experience levels of the companies and their forecasters, contract and compensation formats and supplier relations. This way, one could further investigate the dynamics behind the variables that were selected for the multivariate models. Further analysis of these variables may give valuable insights as to where the greatest potential for improvement lies.

As discussed earlier, I chose to remove the variables of commodity prices form both the multivariate models. This was due to logical reasoning as there has been only two short periods in which the commodity prices have changed covered by my dataset. That is during the financial crisis in 2008 and the downturn in the oil industry in 2014. My analysis therefore contains a limited amount of observations from periods of decreasing commodity prices. These periods have had limited duration compared to the time needed for restructuring in the industry, and therefore makes it difficult to estimate to what extent the commodity prices affect the magnitude of cost overruns. It would therefore be interesting to study the effects of commodity prices in the future. Since 2014, the oil industry has experienced a greatly reduced oil price for a prolonged time with accompanying downsizing and organizational restructuring. The industry may therefore protect itself from uncertainty regarding business cycles in the future to a larger extent than previously observed. They also seem to have a more predictable need for labor, materials etc., due to politics and the buffer production from US oil production that is likely to stabilize the macroeconomic climate, thus leading to smaller fluctuations in investments and oil price. Consequently, it seems fair to expect a somewhat changed effect on cost overruns from the NCS investment surprise variable, as well as observing how long-lasting changes in commodity prices affect cost overruns.

Another exciting area of future research is to further investigate whether the sectors affect the cost overruns of each other. Although I have not been able to demonstrate this hypothesis, the findings suggest a certain dynamic between the two. As shown by the public sector investment allocation (figure 1.2), the accumulated investment level has increased considerably in recent year, and the current government in Norway has promoted further increases. It is logical to think that increased investments in recent years, which are not covered well in my dataset, will lead to increased pressure on available personnel, supplier markets and inflate the general price level of needed products and services, equal to what we have seen in the oil industry. Thus, it emerges as an interesting area to study how a more stable activity level in the oil industry, combined with increasing investments in the public sector, will affect the interaction of cost overruns between the two sectors.

6 References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), pp.716-723.

Bertelsen, M. (2017). Det var ikke noe vi prioriterte! [online] *Dagens Næringsliv*. Available at: http://www.dn.no/nyheter/2017/02/13/0756/Energi/-det-var-ikke-noe-vi-prioriterte) [Accessed 18 Mar. 2017].

Dahl, R., Lorentzen, S., Oglend, A. and Osmundsen, P. (2017). Pro-cyclical petroleum investments and cost overruns in Norway. *Energy Policy*, 100, pp.68-78.

EY (2013). Spotlight on oil and gas megaprojects. [online] Available at: http://www.ey.com/Publication/vwLUAssets/EY-spotlight-on-oil-and-gas-megaprojects/$FILE/EY-spotlight-on-oil-and-gas-megaprojects.pdf [Accessed 28 Mar. 2017].

Flyvbjerg, B. (2007). Decisions based on scant information. Challenges and tools during the front-end phase of projects. *Concept Report, 17,* pp.41-53.

Flyvbjerg, B., Holm, M., and Buhl, S. (2002). Underestimating Costs in Public Works Projects: Error or Lie?. *Journal of the American Planning Association*, vol. 68(3), pp. 279-295.

Flyvbjerg, B., Holm, M. and Buhl, S. (2004). What Causes Cost Overrun in Transport Infrastructure Projects?. *Transport Reviews*, 24(1), pp.3-18.

Kadane, J. and Lazar, N. (n.d.). Methods and Criteria for Model Selection. [online] Carnegie Mellon University. Available at: https://www.cs.cmu.edu/~tom/10-702/tr759.pdf [Accessed 5 May 2017].

Levin, J. and Milgrom, P. (2004) Producer Theory. [online] Available at: http://web.stanford.edu/~jdlevin/Econ%20202/Producer%20Theory.pdf (Accessed: 6 March 2017).

Lorentzen, S., Osmundsen, P. and Sandberg, F. H. (2017). Kostnadsutvikling på norsk sokkel. *Samfunnsøkonomene*, vol. 2, p. 77-81.

Lundman, P. (2011). Cost management for underground infrastructure projects: A case study on cost increase and its causes. Pd. D. Luleå University of Technology, Department of Civil, Environmental and Natural Resources Engineering.

Merrow, E. W. (2011). Industrial megaprojects: Concepts, Strategies, and Practices for Success. Wiley, Hoboken NJ.

NOU 1999:11. Analyse av investeringsutviklingen på kontinentalsokkelen

Oglend, A., Osmundsen, P., and Lorentzen, S. (2016). Cost Overruns in Norwegian Oil and Gas Projects: A Long-tailed Tale. IAEE International Conference Proceedings

Olatunji, O. (2010). The impact of oil price regimes on construction cost in Nigeria. *Construction Management and Economics*, 28(7), pp.747-759.

Olje- og energidepartementet, (2017). Olje og gass. [online] Regjeringen.no. Available at: https://www.regjeringen.no/no/tema/energi/olje-og-gass/id1003/ [Accessed 23 Mar. 2017].

Oljedirektoratet (2013). Vurdering av gjennomførte prosjekter på norsk sokkel. Rapporter 2013, Oljedirektoratet

Osmundsen, P., Emhjellen, M., Johnsen, T., Kemp, A. and Riis, C., (2015). Petroleum taxation contingent on counter-factual investment behavior. *Energy J. 36*, 1–20.

Regjeringen.no. (2014). Hva er KS-ordningen?. [online] Available at: https://www.regjeringen.no/no/aktuelt/Hva-er-KS-ordningen/id2001422/ [Accessed 14 Mar. 2017].

Sovacool, B.K., Gilbert, A. and Nugent, D. (2014). Risk, innovation, electricity infrastructure and construction cost overruns: Testing six hypotheses. *Energy 74*, pp. 906-917.

Welde, M. (2017). Kostnadskontroll i store statlige investeringer underlagt ordningen med ekstern kvalitetssikring. *Concept Report, 51*, pp.18-23, 67-69

Wooldridge, J. (2009). Introductory Econometrics - A Modern Approach. 4th ed. Mason, OH: South-Western Cengage Learning, pp.104-105, 254, 740-742, 840.