University of Stavanger

FACULTY OF SCIENCE AND TECHNOLOGY

MASTER'S THESIS

Study programme/specialisation: Industrial Economics Finance and Risk Management	Spring, 2020 Open
Author: Simen Mosgren Bekkedal	
Programme coordinator: Reidar Brumer Bratvold	
Title of master's thesis: The Outside View: Correcting Biases in the Oil- and Gas Industry	
Credits: 30	
Keywords: Reference Class Forecasting Bias Production Estimates	Number of pages: 77 + supplemental material/other: 28 Stavanger, 14.07.2020 　　　　　date/year

Acknowledgement

Abstract

When companies want to operate and produce petroleum on the Norwegian continental shelf (NCS) they are generally required to make an extensive plan prior to the construction work. This is known as the plan for development and operation (PDO). The PDO addresses a number of decision relevant topics, including the estimated production rate on a year-by-year basis for the entire expected lifetime of the field. These estimates are reported as P10-, mean-, and P90 values. As will be demonstrated in this work, these estimates tend to be severely biased. The extent of this bias and procedures to reduce it will be illustrated and discussed.

The source material for this thesis is a data set containing estimates for 51 fields on the NCS. Analyzing the estimates, we find that the P10-P90 range, i.e., the 80 percent confidence intervals that the companies are requested to report in the PDO, only capture 40 percent of the true production outcomes. Furthermore, about 41 percent of these outcomes falls below the P10 values, on average. This suggests that assessors are generally overly precise and optimistic in their forecasts.

In an attempt to alleviate the bias, a modified reference class forecasting model has been applied to the data set. This methodology is a new adaptation of the classical reference class forecasting model, introduced by Kahneman and Tversky in the 1970s. The modified model calibrates the assessed P10-, mean- and P90 estimates, whilst the classical model uses solely the mean estimate and creates a new distribution from this single value. Since the modified model utilizes more information from the original forecast it was hypothesized that it would outperform the classical model. However, even though both models significantly reduced the bias in the estimates, the acquired results showed that the modified model is performing worse than the classical model.

Table of contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Goal

The goal of this thesis is to analyse petroleum production estimates and develop
and implement methods for debiasing the forecasts. The idea is to apply an ex-
post calibration technique. This means that the estimates are calibrated without
tweaking the models that have been used to produce them. The method that will
be presented will focus on debiasing the assessors. This is possible by comparing
historical estimates and the associated actual production values.

Mohus applied the classical reference class forecasting method to debias production
estimates in his 2018 thesis work [32]. In this thesis, however, a modified reference
class forecasting method will be used. The hypothesis is that this new method can
improve on the already good results of the classical method.

1.2 Background

1.2.1 Kahneman and Tversky

> If i had one wish, it is to see organizations dedicating some effort to
> study their own decision processes and their own mistakes, and to
> keep track so as to learn from those mistakes

The above quote [39] is an excerpt from an interview with Nobel Prize winner
Daniel Kahneman. Together with Amos Tversky he made many pioneering con-
tributions to the field of cognitive bias and judgement under uncertainty. In their
1977 paper "Intuitive predictions: biases and corrective procedures" [20] they in-
troduced the concept of reference class forecasting (RCF). The methodology allows
for a higher degree of objectivity, which Kahneman and Tversky referred to as tak-

1

ing an outside view. This framework has since been used and further developed in several scientific publications (Flyvbjerg, 2008 [11]; Mohus, 2018 [32]; Clemen et al., 2020 [5]).

1.2.2 Mohus

Mohus delivered a Master's thesis in Industrial Economics at the University of Stavanger in 2018 [32]. His paper focused on several topics:

- Estimate the value lost as a direct consequence of biased estimates in the PDO - including cost- and schedule overruns as well as overly optimistic production forecasts
- Discussion of why the industry fail to meet their budgets and what factors affect experts when they forecast future outcomes
- Implement a procedure - based on RCF - to debias the estimates

This thesis is in many ways a continuation of Mohus' thesis in that it aims to further improve debiasing of experts' assessments. Moreover, the data set used in his work is similar to the data set used in this paper, only this paper includes two more years of data. This allows for a comparison between the two models. Thus, it may be determined whether the new methodology is more accurate than previous iterations.

1.2.3 Clemen

Clemen is a professor of Decision Sciences at Duke University [44]. Clemen has a particular interest in the psychology of judgment and assessing expert probabilities. He has been the author of several scientific papers concerning these topics. Clemen and Lichtendahl [38] introduced models to debias experts using Bayesian methods and past performance data. The paper focused on maximizing the likelihood of the posterior distribution of the experts' bias parameters. More recently, Clemen has been the co-author of another paper which is now open for peer-review [5]. The new paper is based on the work by Clemen and Lichtendahl [38], but instead focuses on expanding the assessed confidence intervals. The main advantage of the new method is the ease of use. The methodology presented in this thesis is based on the new paper by Clemen et al.

1.3 Structure

This paper is made up of seven chapters.

In Chapter 1, the goal of the research is introduced, as well as the background material.

In Chapter 2, the procedures of petroleum exploration and production on Nor-

wegian soil is explained. Furthermore, uncertainty and decision analysis will be explained in this chapter.

In Chapter 3, the data, the method and the limitations of it will be briefly explained.

In Chapter 4, a closer examination of the data will be conducted.

In Chapter 5, the modified reference class forecasting methodology will be applied to the data and compared to the classical methodology.

In Chapter 6, a discussion about the data and the results of the RCF procedures is included.

In Chapter 7, the thesis is concluded.

Chapter 2

Theory

2.1 The Norwegian continental shelf

All coastal nations have right to the continental shelf 200 miles off the coast [37]. This can be considered the baseline juridical continental shelf. Norway also reigns beyond this 200 mile area in the areas around Svalbard and Jan Mayen. In total, the Norwegian continental shelf (NCS) spans more than two million square kilometers; almost six times the mainland area [24].

The NCS is illustrated in Figure 2.1. It spans three seas - the North Sea, the Norwegian Sea and the Barents Sea. It is divided into quadrants, which are squares of one longitudinal degree by one latitudinal degree. The quadrants are again divided into 12 square blocks. This partitioning is an effective way of systemizing the NCS. It is also important in the process of awarding licenses.

2.2 Chain of events

Developing an oil- and gas field is a long and demanding process. However, Norway has more than 50 years of experience as a petroleum nation and has thus been able to standardize a lot of procedures. The main phases of petroleum activities can be summarized as follows:

1. Initiation
2. Licenses
3. Exploration
4. Project development
5. Production
6. Plug and abandonment
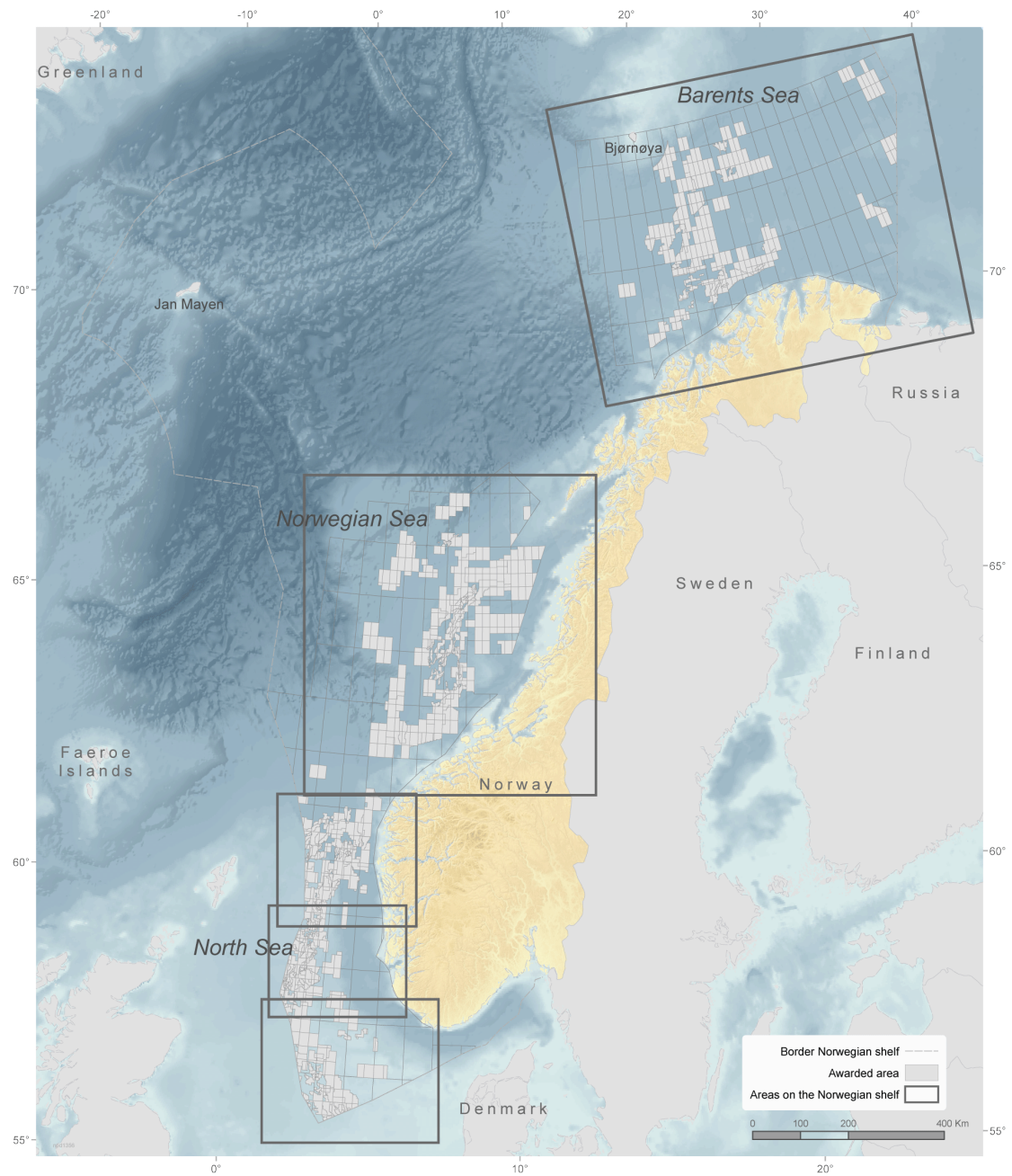
Figure 2.1: 2019 edition of the map of the Norwegian continental shelf [24]

2.2.1   Initiation

The first step in opening a new area for petroleum production is to evaluate the impact such activities would have on the area [8]. In this evaluation, many aspects should be considered, including environmental consequences, economic benefits and social effects.

After an assessment has been prepared it is put before different organizations [8]. This democratization of the process enables both supporters and adversaries to share their views on the issue.

2.2.2   Licenses

The NCS is by law the property of the Norwegian government. Before a company can engage in any endeavours on the soil a license has to be awarded. Licenses are awarded yearly, and any company that is pre-qualified to be a licensee may apply [25]. A license may be awarded for a part of a block, a whole block, or several blocks. The awarding process is administered by the Ministry of Petroleum and Energy (MPE).

There are two different kinds of licenses - survey licenses and production licenses. A survey license grants the right to explore for petroleum [8], whilst a production license grants the right to produce the reserves of the area. The hydrocarbons produced is the property of the licensees. The licensees need not be the operator on the block it has been awarded. The operator is appointed or approved by the MPE when the license is granted.

2.2.3   Exploration

In the exploration phase, the goal is to identify prospects, quantify the hydrocarbon volume and evaluate potential risk factors [35]. Typical exploration methods include conducting and analysing gravimetric-, magnetometric-, and seismic surveys. If the surveys give indications of a technically- and economically feasible prospect, an exploration well will be drilled. This allows for physical confirmation of petroleum accumulations as well as further analyses with, e.g., wireline tools and core samples.

2.2.4   Project development

Depending on the results of the exploration, a decision must be made whether production should be pursued in the area. Figure 2.2 outlines the entire development process. The planning phase is typically finalized by a PDO submission to the MPE.
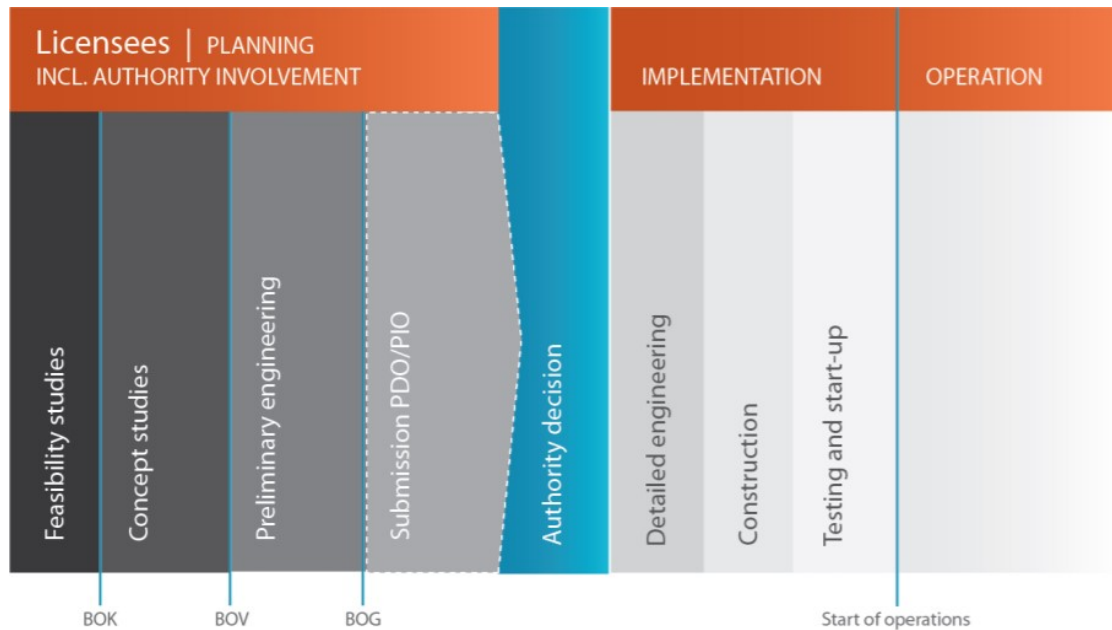
Figure 2.2: Roadmap of the necessary steps in the pre-operation phase [31]

A PDO - Plan for Development and Operation - is a formal document that should include [8]:

- Economic
- Resource

- Technical
- Safety

- Commercial
- Environmental

MPE may also require additional information to be included in the PDO [8].

Upon receiving a PDO, the MPE forwards it for consideration to the Ministry of Local Government and Regional Development, which addresses safety matters, and the NPD, which addresses resource related aspects [34]. These institutions will then submit the result of their consideration back to the MPE. Simultaneously, the MPE forwards the PDO to other parties entitled to voice their opinion, including affected municipalities, organizations and interest groups. Projects with an investment ceiling above five billion NOK need to have their PDO approved by the Norwegian Parliament.

Unless otherwise authorized by the MPE, the PDO must be approved prior to the development of a field. Construction and development usually commence shortly after PDO approval, making the reported estimates on cost, schedule and production highly important for the dimensioning of the field. Low production estimates in the PDO will lead to a construction meant for smaller volumes. If the field turns out to be large, it would entail a large loss in efficiency and profitability. Similar is true for high estimates and low actual volumes.

2.2.5 Production

During the production phase it is important to control production and the injection of pressure stabilizing fluids, as well as monitor crucial data about the reservoir and the stability of the well [36]. It is common to graph a time-series of production volumes to get an overview of a well's performance over time. This production profile will generally follow a pattern, outlined in Figure 2.3. Production volumes are ascending in the first phase of production. Production then increases until a plateau is reached. The plateau production rate is maintained for as long as possible. As fluids start to deplete and pressure is lowered, production decreases. Finally, production will have become so low that it is no longer economically viable to continue production.
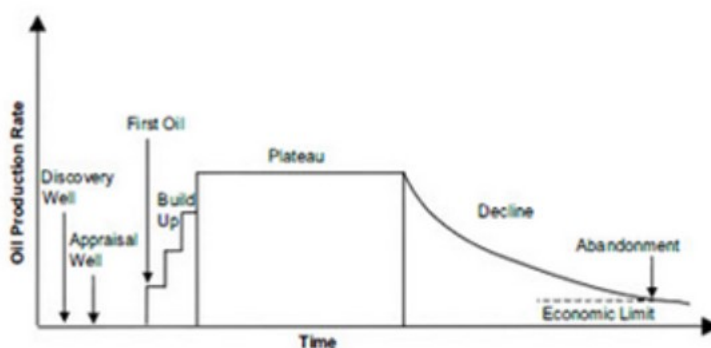
Figure 2.3: A generalisation of production profiles [36]

2.2.6 Plug and abandonment

If a field is no longer profitable a decommissioning plan will have to be filed to Norwegian authorities. This submission must include a disposal plan and an impact assessment [8]. The disposal plan must include information about the field, the production history, and possibilities for continued production. The latter is particularly important, as Norwegian authorities want to maximize the utility from petroleum assets. In the impact assessment, commercial- and environmental impact of cessation must be considered. The main goal of plugging a well before abandonment is to ensure long-term well integrity and protect the environment from the threat of pollution caused by petroleum activities.

2.3 Resources

All recoverable oil and gas are collectively coined resources. Resources are further categorized into reserves, contingent resources and undiscovered resources [28]. Figure 2.4 gives an overview of these categories.
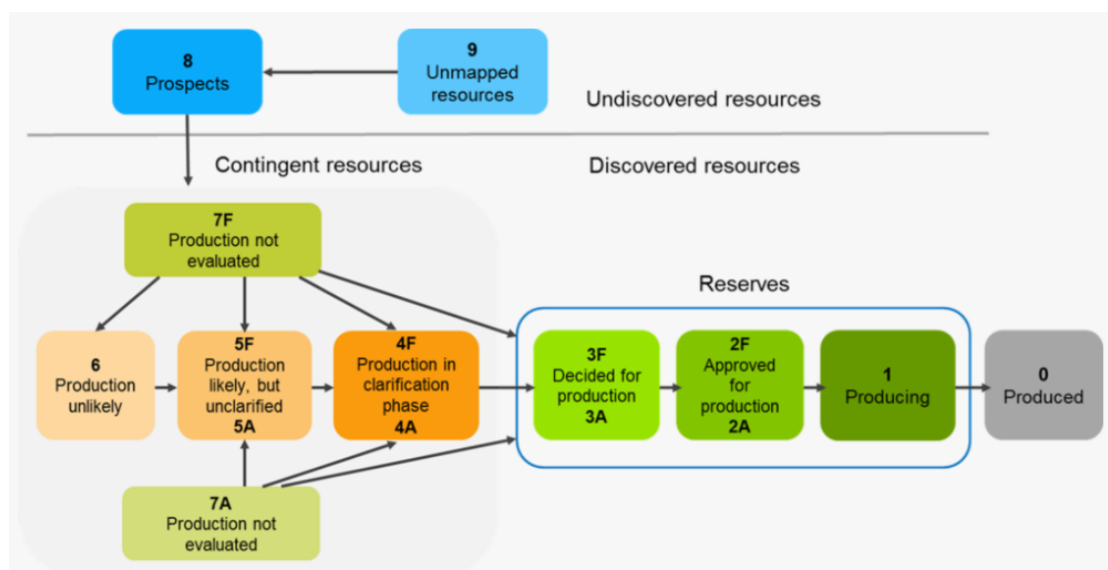
Figure 2.4: Classification of resources [28]

2.3.1 Reserves

Reserves comprise the petroleum volumes which the licensees have decided to produce [29].

Table 2.1: Original reserves on the NCS [$MMSm^3$][30]

Område		Oil	Condensate	NGL	Gas	Sum o.e.
Barents Sea		120.50	24.26	16.06	216.05	376.87
North Sea		4739.16	70.17	410.72	3010.58	8230.63
Norwegian Sea		738.84	41.40	171.48	889.02	1840.74

Table 2.1 illustrates the amount of *original reserves*, i.e., all the petroleum resources which have been planned for production up until this date, including that which have already been produced.

Remaining reserves is regarded as the difference between original reserves and produced volumes [30]. Table 2.2 gives an overview of the remaining reserves on the NCS as of December 31, 2019. Only counting oil, there is 1737 Sm^3 remaining - roughly 31 percent of the original reserves. By assuming an oil price of USD 49, the remaining oil has a value of about USD 535 billion.[1]

---

[1]Simple moving average of Brent crude oil for 2020 as of May 21, 2020

Table 2.2: Remaining reserves on the NCS $[MMSm^3]$[30]

Område	Oil	Condensate	NGL	Gas	Sum o.e.
Barents Sea	109.03	15.85	10.45	154.68	290.01
North Sea	942.75	0.00	103.06	1071.02	2116.83
Norwegian Sea	116.29	4.81	52.10	318.57	491.77

2.3.2 Contingent resources

Contingent resources are proven, but there has not yet been made a decision to produce [28]. This includes resources in improved oil recovery projects and resources that are unlikely to be produced.

2.3.3 Undiscovered resources

Resources which have not yet been proven by drilling are called undiscovered resources [28]. However, these resources are assumed to be there and have been estimated by some means. Due to the limited amount of data behind the estimates, they must be considered rough estimates.

2.4 Taxation and distribution of value

The previous section gave an indication of the value of the resources on the NCS. In terms of added value, government revenues and investments it is Norway's largest industry [26]. A key principle, which has influenced the development of the financial and legal framework governing the sector, is that the wealth provided by the industry should benefit the whole society. The goal of petroleum production on the NCS has always been to maximize value creation for the good of the society. Figure 2.5 illustrates the net government cash flow from petroleum activities between 1971 and 2020. A large portion of the revenue is attributed to the taxation system.

Oil companies operating on the NCS are subject to Norway's ordinary tax rate of 22 percent. Additionally, they are subject to a special tax rate of 56 percent, which gives a total tax rate of 78 percent [27]. The taxation system is intended to be neutral, i.e., a profitable project before tax should also be profitable after the tax burden. To ensure this, only net profit is taxable.

The taxation system ensures that the majority of the generated revenue from petroleum activity on the NCS accrues to the Norwegian government. However, it also ensures that the majority of the companies' costs are tax deductible. The Norwegian government (indirectly the Norwegian tax payers) pays for 89.2 percent of all petroleum related costs [41].
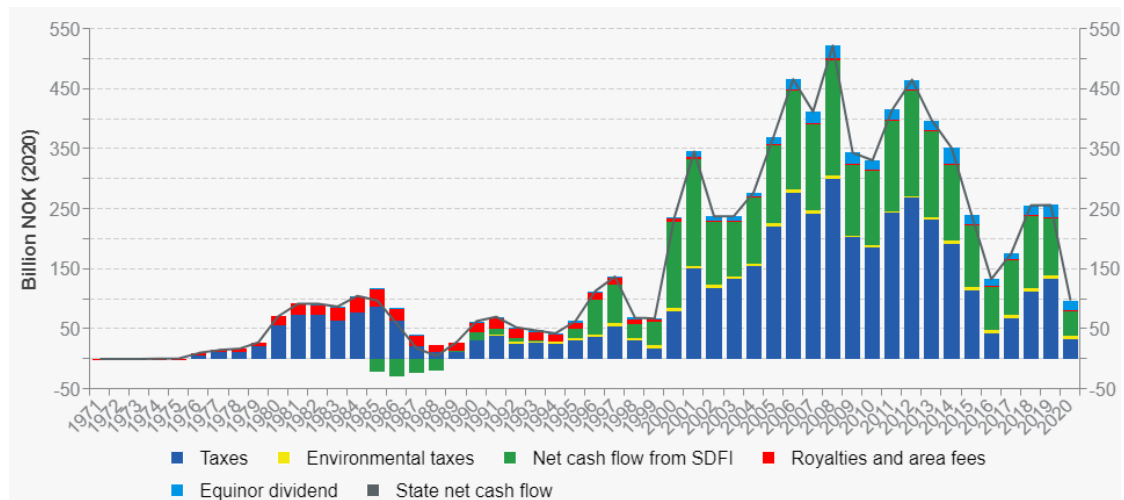
Figure 2.5: Net government cash flow from petroleum activities, 1971-2020 [26]

2.5 Uncertainty

Oil- and gas fields are very expensive projects. Many of which are considered megaprojects [23], with a price tag of more than 10 billion NOK [10]. The enormous size, combined with the complexity of subsea development, often causes a high degree of uncertainty.

Galbraith defined uncertainty in his book *"The Age of Uncertainty"* [17] as "the difference between the amount of information required to perform the task and the amount of information already possessed by the organization". This definition has later been recognized by several independent papers (Daft and Lengel 1986 [7]; Brun 2011 [4]). Bratvold and Begg further elaborate that the uncertainty is a subjective aspect of our knowledge. There is no single, correct uncertainty. The uncertainty depends on the knowledge of the assessor.

Adapting the presented definition, petroleum reservoirs cannot be objectively uncertain. The volumetric size is an inherent property of the reservoir. It is only in our subjective reality that the values are uncertain.

2.5.1 Percentiles and forecasting

The use of percentiles is common practice in many industries. In particular, it is common to analyse the results of college admission tests, such as the SAT, using percentiles. It is also common for financial institutions to measure the resilience of a portfolio (Value at Risk) using percentiles. The convention ensures both an intuitive interpretation and statistical coherence.

If a weather forecast claimed that a hurricane would hit Houston (TX), but would not reach Dallas (TX), a resident in Dallas would believe that she was safe from the hurricane. The single-value forecast does not leave any room for uncertainty,

which may be confusing and misleading. Assume that the probability distribution that the forecast was based on indicated the following:

  (i) P10 = 10 km/h wind speed (light wind)
 (ii) P50 = 40 km/h wind speed (strong wind)
(iii) P90 = 140 km/h wind speed (hurricane)

where P10 is the *low esimate*, P50 the median of the distribution and P90 the *high estimate*.[2] The high estimate indicates that a hurricane will hit Dallas. This information was not disclosed in the arguably misleading single-value forecast.

Most forecasts are based on computer-generated models. The models use a set of inputs defined by an assessor to generate useful information related to the possibilities of different outcomes. This information is processed by an assessor, with a probability distribution and final forecast as a result. The forecasts are typically represented by a P10-, P50- and P90 value. These are three numbers meant to summarize the assessed distribution. The P10 value indicates that only 10 percent of the data points in the distribution falls below this value. Similar is also true for P50 and P90.

Consider a hypothetical fair die with 10 possible outcomes; 1 through 10. Table 2.3 illustrates the generated probability distribution of this die. The PDF is the probability density function, indicating the probability of getting exactly that outcome. The CDF is the cumulative distribution function, indicating the probability of getting exactly that outcome or lower. Since the die has 10 faces, the probability of getting 10 is 10 percent, whilst the cumulative probability is 100 percent. If the die is tossed a great number of times, the outcome is 10 about 10 percent of the time. A *perfectly calibrated judge*, i.e., an assessor without bias, will correctly assess the P10 as 2 and the P90 as 9. This ensures that the outcome of the die is within the assessed P10-P90 range 80 percent of the time. The P50 of the distribution is 5.5 - the mean of the middle values 5 and 6. Figure 2.6 illustrates how the P10, P50 and P90 are distributed in a distribution.

In the guidelines for the PDO, the NPD requests that estimates must include a P10-, mean- and P90 value [43]. It is important to stress that the NPD does expect the companies to provide them with P50 estimates. There is a very clear distinction between the two concepts which is of great significance when regarding the data from a statistical point of view.

2.5.2   Median vs mean

The median and the mean are two different ways of averaging.

---

[2]The definition of P10 and P90 might differ from other papers, but this will be the definition for the remainder of the thesis.

Table 2.3: Probability distribution of the toss of a 10-faced die

Outcome	PDF	CDF
1	10%	10%
2	10%	20%
3	10%	30%
4	10%	40%
5	10%	50%
6	10%	60%
7	10%	70%
8	10%	80%
9	10%	90%
10	10%	100%
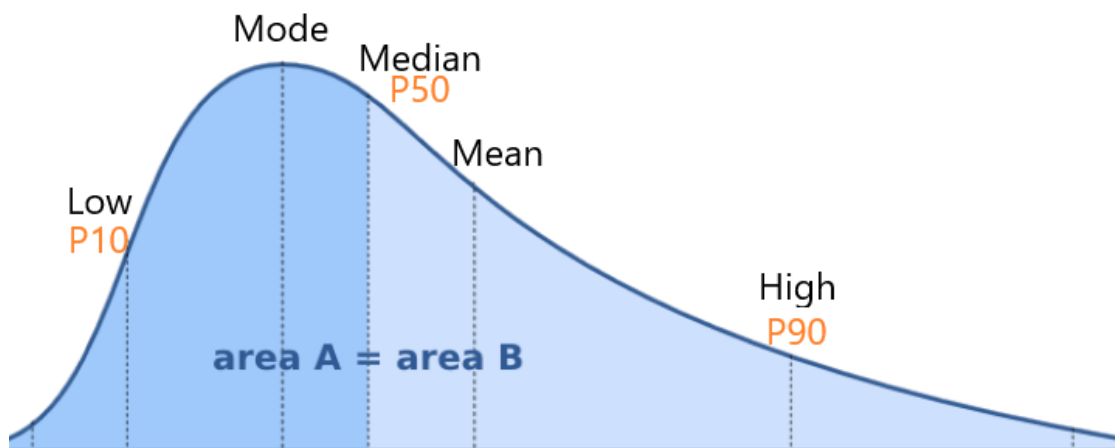
Figure 2.6: A skewed distribution with the associated P10-, P50- and P90 values
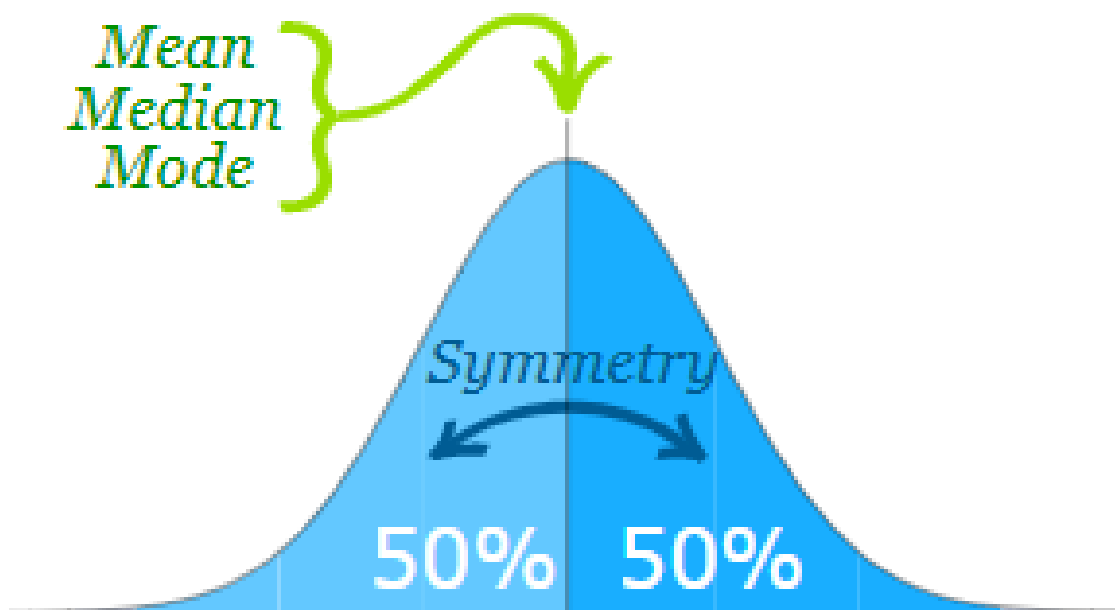
Figure 2.7: A typical normal distribution [22]

Figure 2.7 illustrates a normal distribution. Here, the median and the mean are the same value. This is due to the symmetry of the distribution. However, given a skewed distribution, the median and mean are two different values. A distribution is considered skewed when the majority of data points cluster on one side of the distribution, leaving the other side with a long tail. Figure 2.6 illustrates a positively skewed distribution. This is a typical distribution of possible production volumes in the petroleum industry. Production volumes are lower bound to 0, but may yield large values, albeit with a low probability of occurrence. Given positive skew, the mean will most often be larger than the median, while the opposite is usually true with negative skew.

2.5.3   Uncertainty modelling

Modern commercialization of petroleum dates back more than 150 years [42]. For the majority of this time, decision makers in the industry have spent less than adequate time contemplating inherent uncertainties [2]. It is only in the last couple of decades that effort has been made to increase the understanding of uncertainty. The increase in the number of scientific publications, conferences and meetings in the last years is a testament of this.

A decision can be defined as a conscious, irrevocable allocation of resources to achieve desired objectives [2]. This implies deliberation and the awareness that resources are lost in case of revocation. There are three fundamental elements of every decision: objectives, alternatives and information.

Bratvold and Begg [2] give the following reasons, among others, why decision analysis is important:

1. It is a process which allows for the identification of the most important factors that could influence the decision
2. It helps uncovering unexpected scenarios
3. Professionals have an obligation to present not just answers, but also the reasoning, limitations and assumptions that support the answers.
4. Careful consideration is necessary in cases where it is difficult to understand what the experts are trying to communicate. Even more so when different experts communicates different things
5. Consider historical data and experiences and learn from previous mistakes

Point five is especially important and requires the acknowledgement of the distinction between good decisions and good outcomes [2]. A good outcome is "a future state of the world that we prize relative to other possibilities" [2], while a good decision is "an action we take that is logically consistent with our objectives" [2]. People in general do not usually make this distinction. Bad outcomes are generally considered the result of bad decisions. This is a detrimental mindset which prevents decision makers to learn from past experiences and mistakes.

Chapter 3

Data and method

This chapter contains a description of how data was gathered and the methods that were used to systemize and analyze it. An overview of the limitations of this approach is also presented.

3.1   Data

As discussed in Chapter 2.2.4, companies are required to provide the authorities with estimates of future production rates prior to PDO approval. This is generally company confidential information, but it has been disclosed for this thesis under an NDA-agreement. Field- and company names will not be mentioned when production estimates are discussed. The actual production data from Norwegian oil- and gas fields are readily available online at the NPD website.

The data set is comprised of production estimates and the associated actual production outcomes from 56 oil fields on the Norwegian Continental Shelf from 1997 to 2019. For each field, the P10-, mean-, P90- and actual production values are provided for each year. Due to inconsistencies and irregularities in the data set, some data have been left out of the analysis. This include forecasts where (i) some, or all the estimates are 0, (ii) some, or all the estimates are equal, (iii) the P10 value is greater than the mean, (iv) the P90 value is smaller than the mean and (v) some data is missing. This is either caused by wrongful reporting or poor assessments. For five of the fields there were no estimates without one or several of these issues, excluding them from the analysis. This left 51 fields.

3.2   Method

Since both estimates and actual production outcomes are available it is possible to evaluate how well the estimates align with the actual outcomes. Studying

this comparison for several fields over several years will create a picture of the average statistical quality of estimates. The historical data accumulated across the NCS will be used as input to the approach described in this thesis, allowing for the calibration and debiasing of future estimates. This method is based on the invaluable contribution to the field of judgement and biases by Kahneman and Tversky [19]. The approach, which has been coined the *modified reference class forecasting* (MRCF), calibrates three assessed percentile values based on the accuracy of similar previous assessments. This will be further explained in Chapter 5.

3.3 Limitations

Statistical analysis is meaningless if it is not based on statistically meaningful data. To be able to draw any meaningful conclusion out of a statistical analysis both a sufficient amount of data and reliability of said data is crucial.

The data in this analysis constitutes a large percentage of the total number of fields on the NCS [9].[1] Moreover, Flyvbjerg and COWI's paper on RCF [12] included 172 road projects, 46 rail projects and 34 bridge- and tunnel projects. Thus, relative to similar research, the amount of data should be sufficient.

There is also a question about the reliability of the data, and whether the forecasts truly reflect the assessors' assessments. The methodology in this thesis assumes that P10-, P50- and P90 estimates, with foundation in a probability distribution, is used as input. There are several challenges with this assumption:

1. The NPD request companies to provide P10, mean and P90 estimates
2. The data set that was provided by the NPD includes several data points that are either wrongfully reported or poorly assessed
3. Different companies have different methods of forecasting and it is not known whether the percentiles and mean are based on probability distributions. It is not uncommon to put the majority of the forecasting effort to build a "base case" [2]. The P10 and P90 are usually less prioritized

Furthermore, the results presented in this thesis are based on NCS fields and, thus, all conclusions drawn from this study are limited to NCS forecasts. However, as will be discussed later, optimism bias is not limited to NCS forecasts, nor to the petroleum industry. It is a reasonable assumptions that both the causes and the magnitude of the bias also applies to other geographical regions, affecting the entire petroleum industry. Nonetheless, it would be both interesting and useful to incorporate data from other parts of the world - either to expand the reference class or to create several new reference classes. This would yield a more complete debiasing model.

---

[1]As of July 2020, there are 88 producing fields on the NCS.

Chapter 4

Analysis

In this chapter, the data set, which forms the foundation of this thesis, is presented. The accuracy of the assessors' forecasts will also be evaluated.

4.1 Production estimates

Across the 51 fields included in the data set there are a total of 1320 forecasted values, including P10-, mean- and P90 estimates. Figure 4.1 gives an overview of the mean estimates. The y-axis represents the estimated production volumes in millions of standard cubic meters, while the x-axis represents the production year associated with the given estimates. All the mean estimates associated with a single production year are summed to get the total annual mean. This is an allowable operation for mean estimates. However, the same is not true for percentiles. The average of percentiles is not a percentile. Although the average might represent a value close to the true percentile value of a distribution, it is not statistically consistent to make this assumption. Thus, the average of the percentiles will not be presented and discussed.
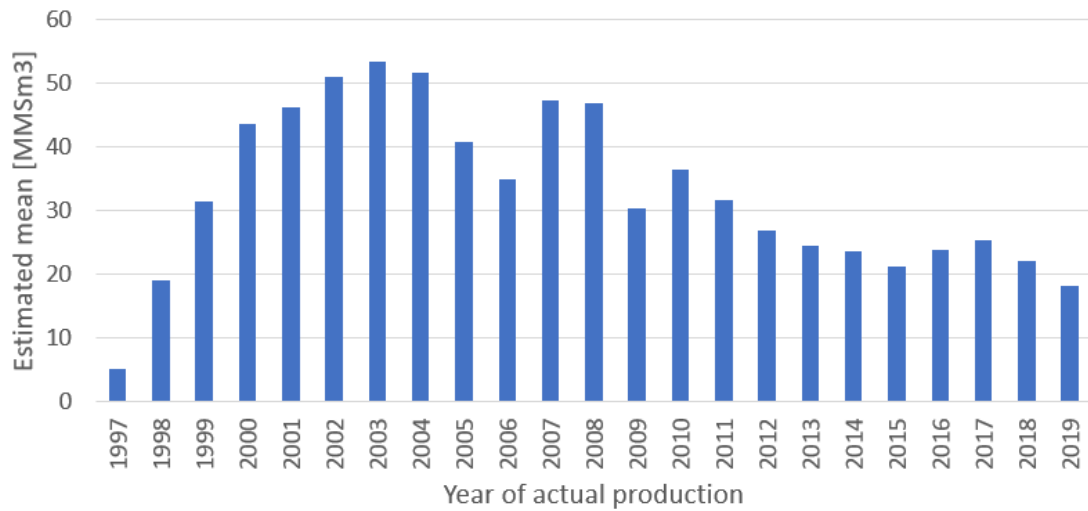
Figure 4.1: Sum of estimated means, year-by-year

All of the fields included in this analysis had their PDOs approved between 1995 and 2017. No estimates after 2019 are included, due to lack of actual production data after 2019. As a result, many fields have a limited number of years of production data. To illustrate, Figure 4.2 shows the number of fields that were producing at any given year after production start on the left axis, and the estimated production on the right axis. There is a greater number of fields for Year 1, 2, 3 and 4 than for Year 0 due to there being a great deal of data with poor quality for Year 0.
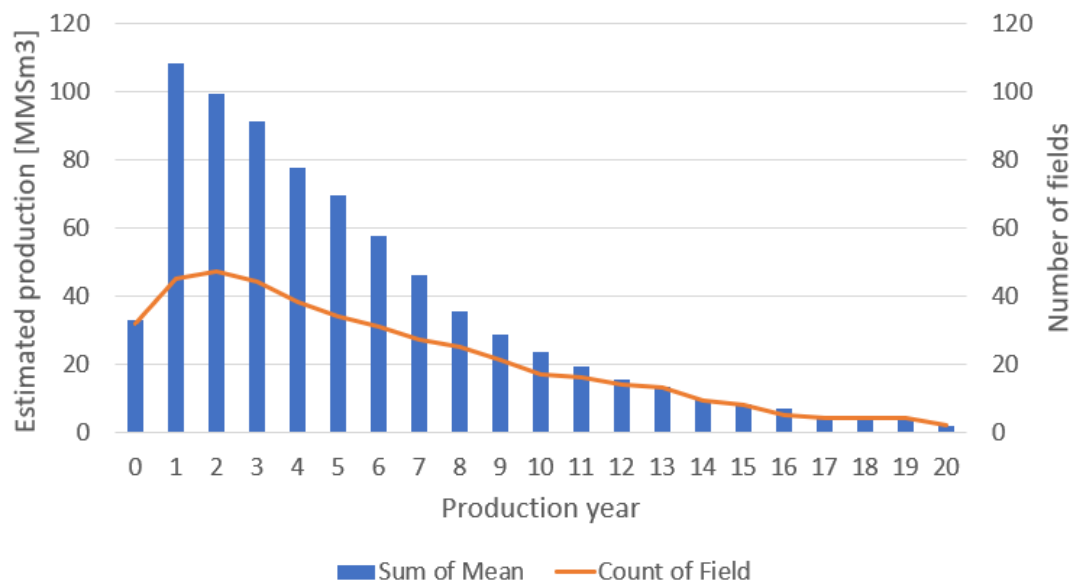
Figure 4.2: Estimated mean and number of fields in production, x years after production start

The declining production volumes, seen in Figure 4.2, is caused by both a de-

creasing number of producing fields and by physical limitations. As explained in Section 2.2.5, the production profile of a typical petroleum field is categorized by an initial buildup in the first few years, followed by a plateau phase before eventually decreasing towards zero. The described pattern can be seen in Figure 4.3, illustrating the average of the mean estimates. On average, the initial buildup lasts for about one year and the plateau phase lasts for about five years.
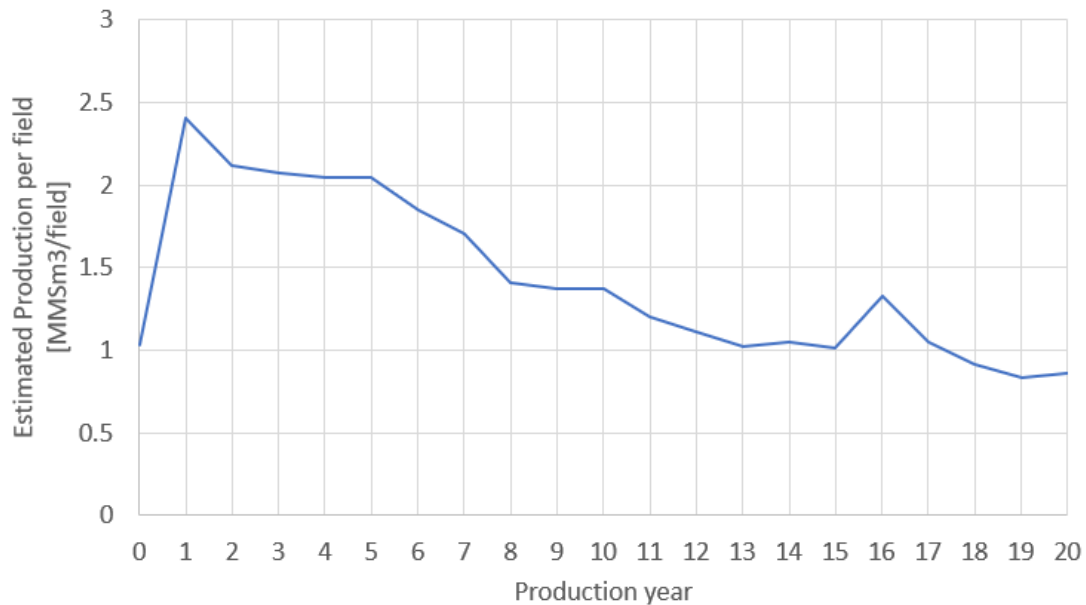
Figure 4.3: Estimated mean divided by number of fields in production, x years after production start

4.2 Actual production

Production estimates are estimated on a year-by-year basis from the year of estimated production start until the estimated cessation of the field. However, the estimated production start might not coincide with the actual production start. Out of the 51 fields that were included in the data set, 17 fields started producing at least one year after it was estimated. This thesis is a study of bias in production estimates and, thus, the delays should not complicate this analysis. To reduce the effect of delays, the first year of production will be compared to the first estimate. Although this reduces the effect of delays, it is not completely eliminated. It is likely that there is a difference between the month in which the actual production started and the month in which the production was estimated to start. Since the Norwegian authorities do not request month-by-month data, this information is not available. Therefore, delays have been removed to the extent possible given the lack of time granularity in the data.

Figure 4.4 is an illustration of the differences between the estimated total annual mean and total annual production outcomes. The Year 0-estimates are 64 per-
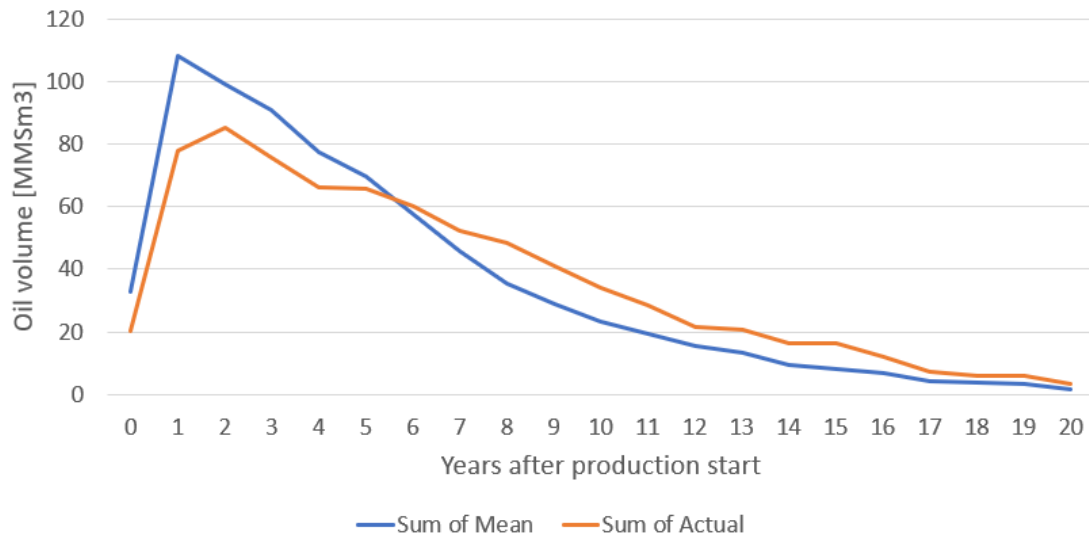
Figure 4.4: Direct comparison between mean estimate and actual production

cent larger than the actual outcomes. For Year 1 and 2 this number is 39- and 16 percent, respectively. For each of the five first years, the total annual mean estimates are higher than the total annual production. Figure 4.5 illustrates the accumulated mean estimates and the accumulated production for each year after production start. The accumulated mean estimates are higher than the true production outcomes for the first 16 years of production.

For an unbiased assessor, the P50 forecasts should be assessed such that the number of actual production outcomes is equal above and below this value. This makes the interpretation of P50 simple. In Section 2.5.2, the differences between the median and the mean were discussed. Given skewed distributions, the two values will not be the same. Nonetheless, to allow for an easier interpretation of the data, the mean estimates will be assumed equal to the median. The ramifications of this decision will be discussed in Appendix A. Hereinafter, the mean, median and P50 will be used interchangeably.

Some important terminology when discussing bias include overprecision and optimism. The former is when assessors consistently produce forecasts with too narrow intervals, i.e., the P10-P90 range captures less than 80 percent of the true production values. The latter is when assessors produce forecasts where the true values, for the majority of the time, falls below the assessed P50.
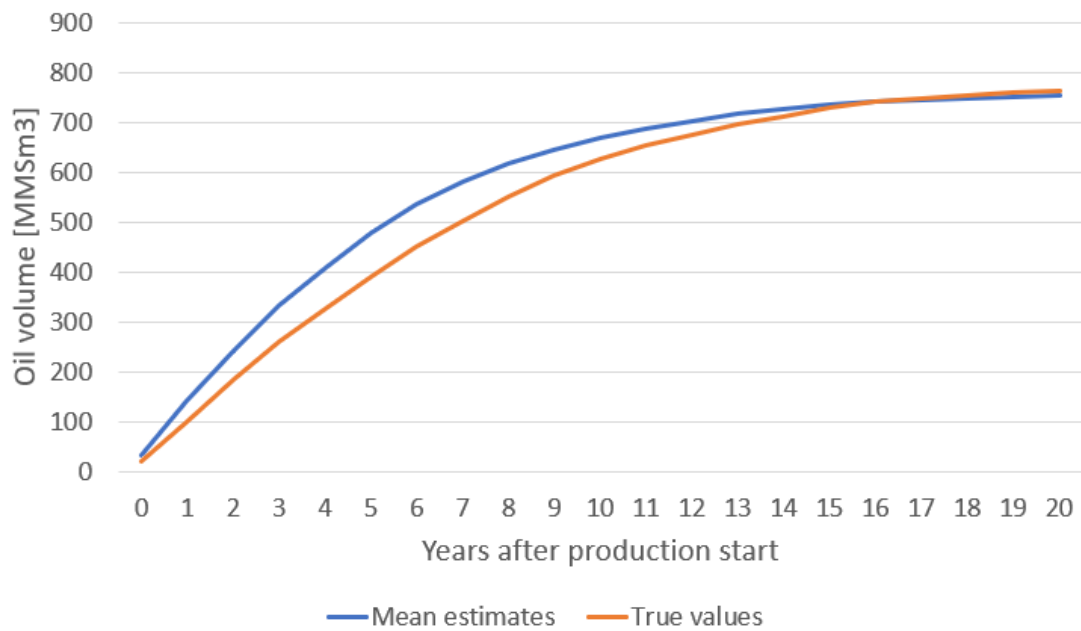
Figure 4.5: Direct comparison between mean estimate and actual production, accumulated production
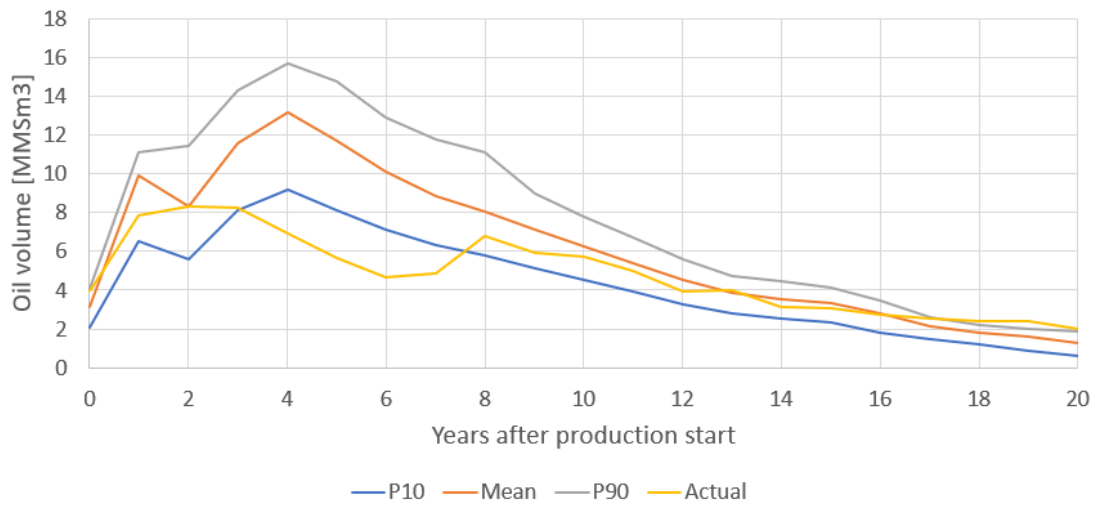
Figure 4.6: Direct comparison between estimates and actual production for Field A

Figure 4.6 illustrates the P10-, P50- and P90 estimates, as well as the actual production data for field A. 67 percent of the true production outcomes are lower than the estimated P50. 19 percent of the true values are lower than the P10 values and 14 percent are higher than the P90 values.
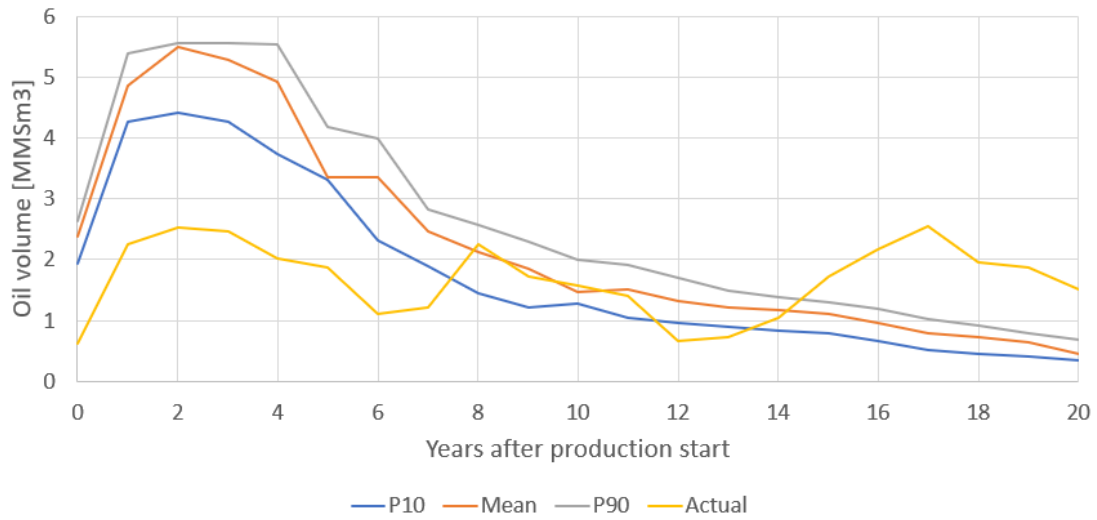
Figure 4.7: Direct comparison between estimates and actual production for Field B

Figure 4.7 illustrates the percentiles and actual production data for field B. 62 percent of the true production outcomes are lower than the estimated P50. 48 percent of the true values are lower than the P10 values and 29 percent are higher than the P90 values.
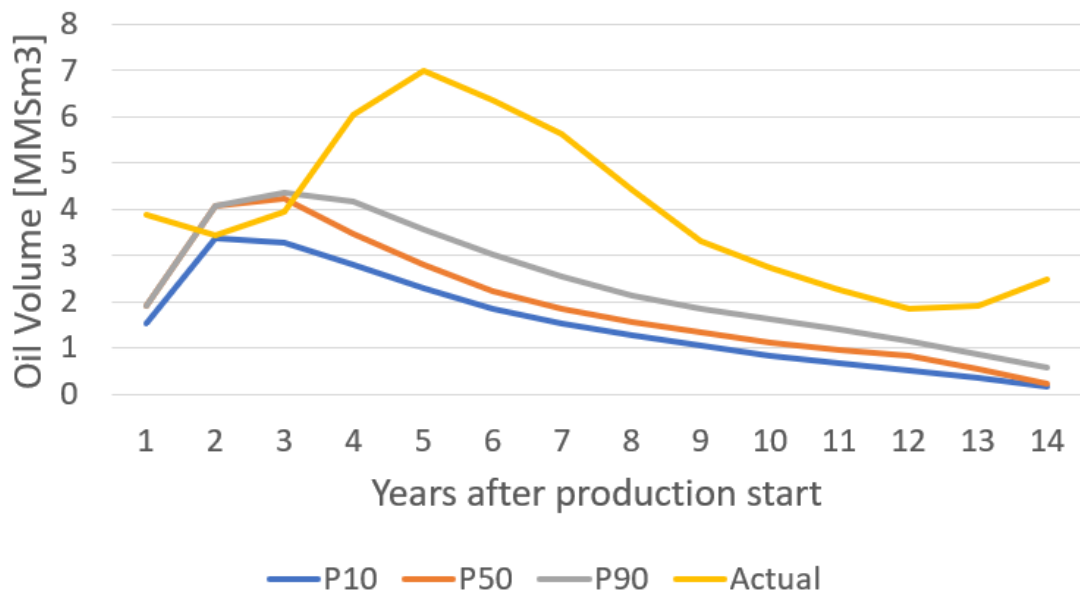
Figure 4.8: Direct comparison between estimates and actual production for Field C

Figure 4.8 illustrates the percentiles and actual production data for field C. 90 percent of the true production outcomes are *higher* than the estimated P50. 0 percent of the true values are lower than the P10 values and 57 percent are higher than the P90 values.

Figure 4.6 and 4.7 both indicate overly precise and optimistic forecasts. Figure 4.6, 4.7 and 4.8 are all evidence of overprecision. However, these figures do not provide substantial evidence that forecasts on the NCS are biased in general. Variations are expected in probabilistic forecasting. However, when the majority of the data indicates biased forecasts, it becomes a trend. This is further studied in Section 4.2.2.

4.2.1 Additional investments

As petroleum fields grow older, and production volumes diminish as a result of depleting reservoir pressure, the potential benefits of enhanced oil recovery-investments increase. Table 4.1 gives an overview of some of the additional investments on the NCS. This adds another level of complexity to the analysis. It is not known whether these investments are accounted for in the PDOs and, thus, only the first four years of production data will be discussed for the remainder of this thesis.

4.2.2 Year-by-year

Figure 4.9 illustrates the estimated production versus actual production for the first year of production for every field. The orange markers represent the mean forecasts. The vertical lines intersecting the mean forecasts represent the assessed 80 percent confidence intervals. The blue, diagonal line is 45 degrees and indicates where the estimated production equals the actual production. By definition, a perfectly calibrated judge would see his 80 percent confidence intervals intersect the blue line eight out of ten times. The dotted line is a linear regression of the input data. For a perfectly calibrated judge, this line should align with the blue line. The assessed 80 percent intervals rarely cross this line, indicating overprecision. Furthermore, the regressed line indicates optimistic forecasts. Figure 4.10 shows the same distribution, but with lower values on the axes, highlighting smaller fields. There are no apparent differences between small and large fields for Year 0.

Table 4.1: Overview of additional investments in fields that are already producing [32]

Field	PDO Approved	Investment	Years after PDO
VALE	2001	2005	4
FRAM	2001	2007	6
VEGA	2007	2013	6
BALDER	1996	2003	7
GLITNE	2000	2007	7
OSEBERG SØR	1997	2004	7
TAMBAR	2000	2007	7
URD	2004	2011	7
VILJE	2005	2012	7
GULLFAKS SØR	1996	2004	8
VOLVE	2005	2013	8
ALVE	2007	2016	9
VARG	1996	2005	9
NORNE	1995	2005	10
KVITEBJØRN	2000	2013	13
OSEBERG ØST	1996	2011	15
SYGNA	1999	2015	16
VISUND	1996	2012	16
ÅSGARD	1996	2012	16
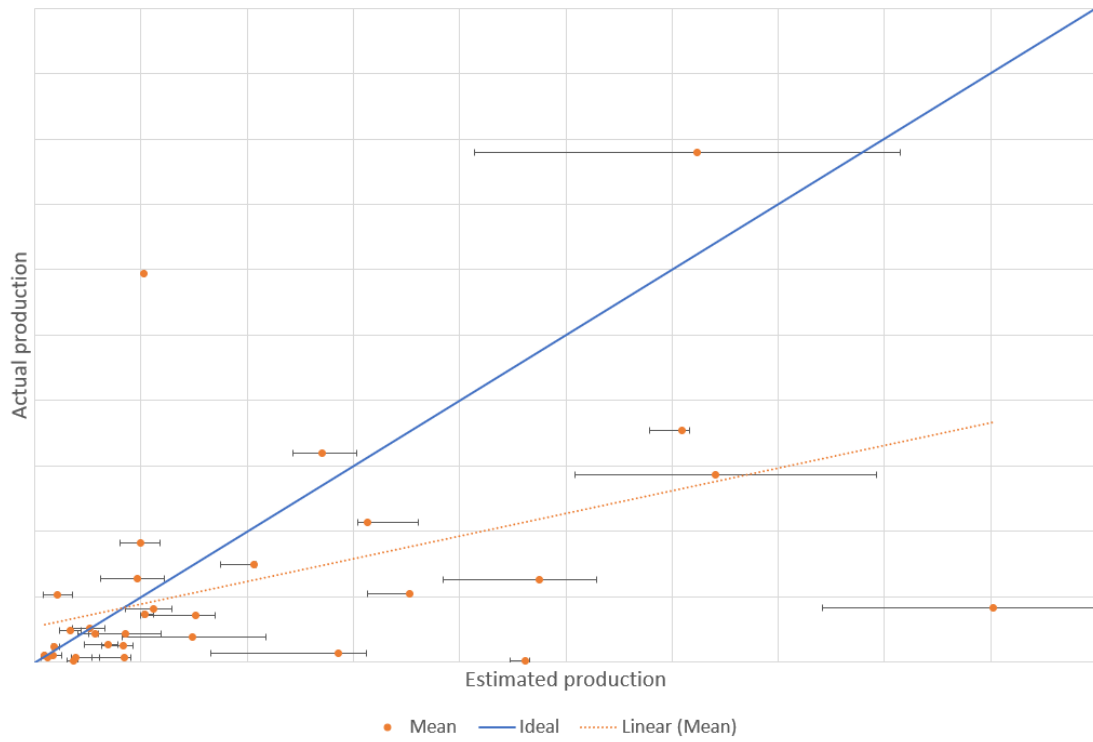
Figure 4.9: Confidence intervals for estimates versus actual production, first year of production, all fields
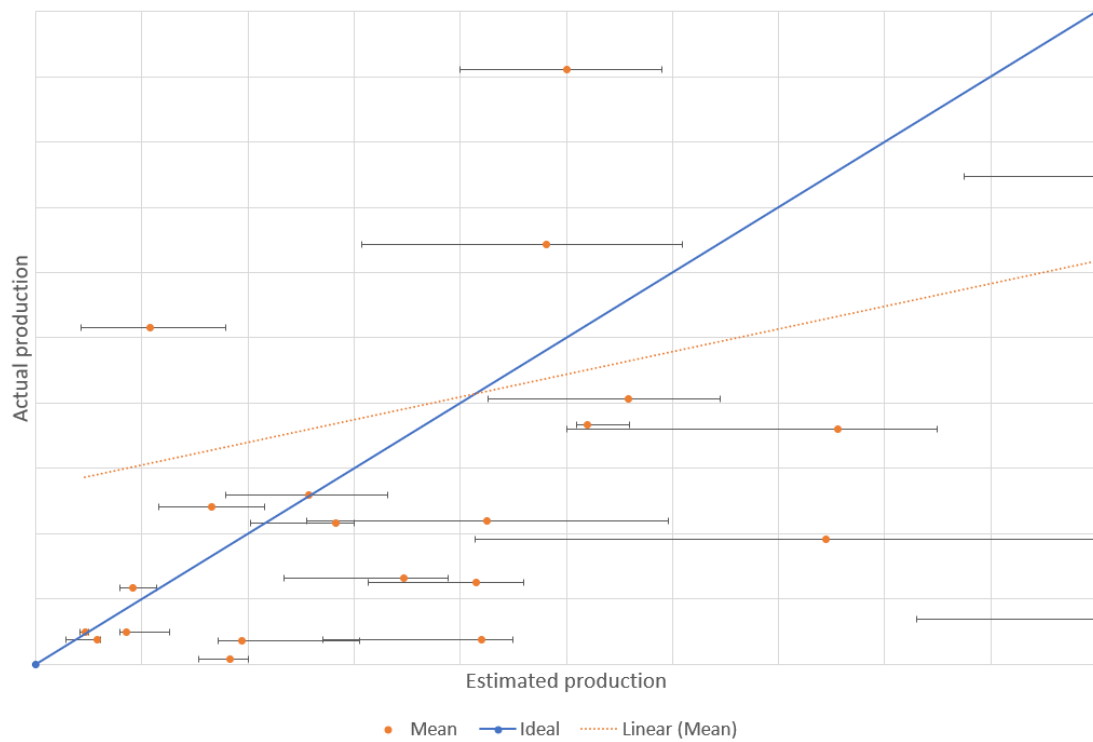
Figure 4.10: Confidence intervals for estimates versus actual production, first year of production, smaller fields
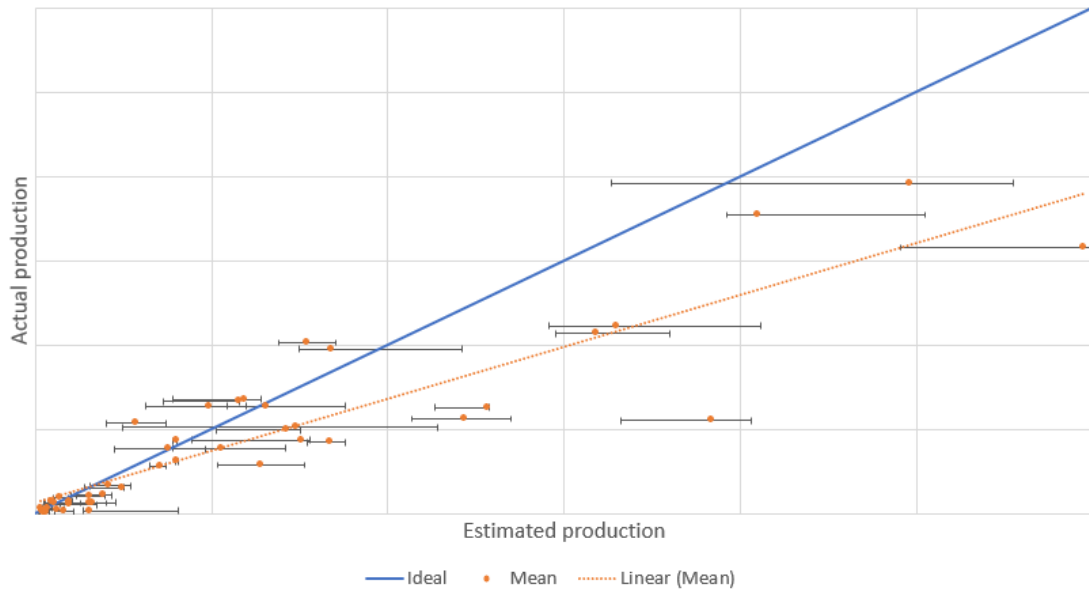
Figure 4.11: Confidence intervals for estimates versus actual production, second year of production, all fields
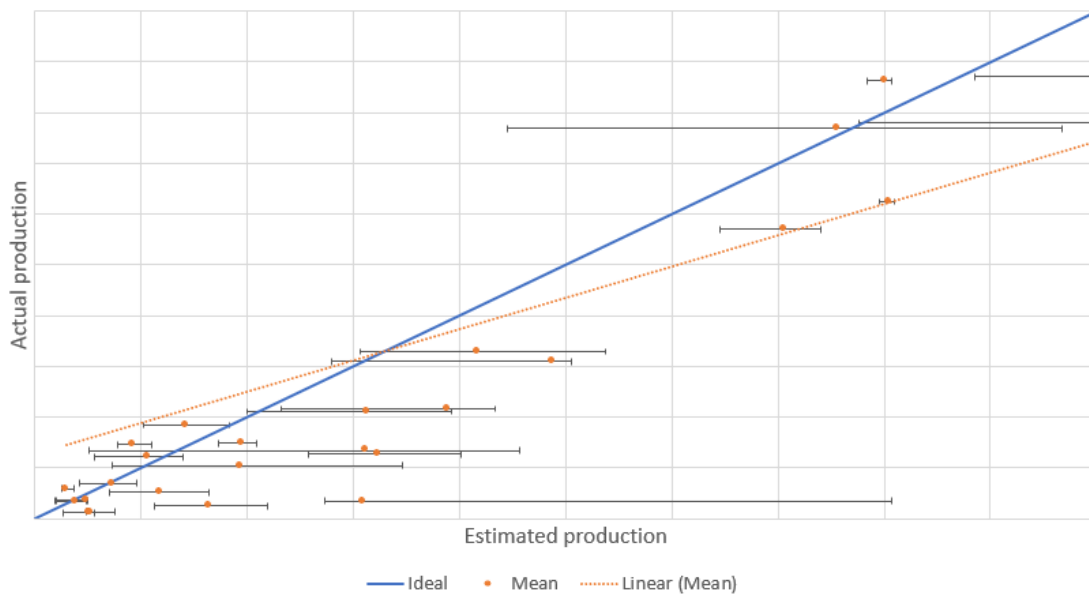
Figure 4.12: Confidence intervals for estimates versus actual production, second year of production, smaller fields

Figure 4.11 and 4.12 illustrates the same as Figure 4.9 and 4.10, but with data from Year 1. Even more such examples are provided in Appendix B.

Table 4.2 summarizes and gives a complete oveview of the first four years of data. It also illustrates the distribution that results from a perfectly calibrated judge. It is evident that the assessors' forecasts are not congruous with a perfectly calibrated

judge. For example, the assessors' 80 percent confidence intervals capture only 40 percent of the true values, indicating overly precise forecasts. 41 percent of the true values falls below the assessed P10, indicating optimistic forecasts.

Table 4.2: Actual production values vs estimates, first four years

	Inside interval	Below P50	Above P50	Below P10	Above P90
Perfectly calibrated judge	80%	50%	50%	10%	10%
Assessors' forecasts	39.29%	66.07%	33.93%	42.26%	18.45%

4.3   Are there signs of improvement?

The analyses in this chapter have investigated the overall quality and precision of forecasts. However, there have been no analyses on whether the forecasts have improved in the past two decades.
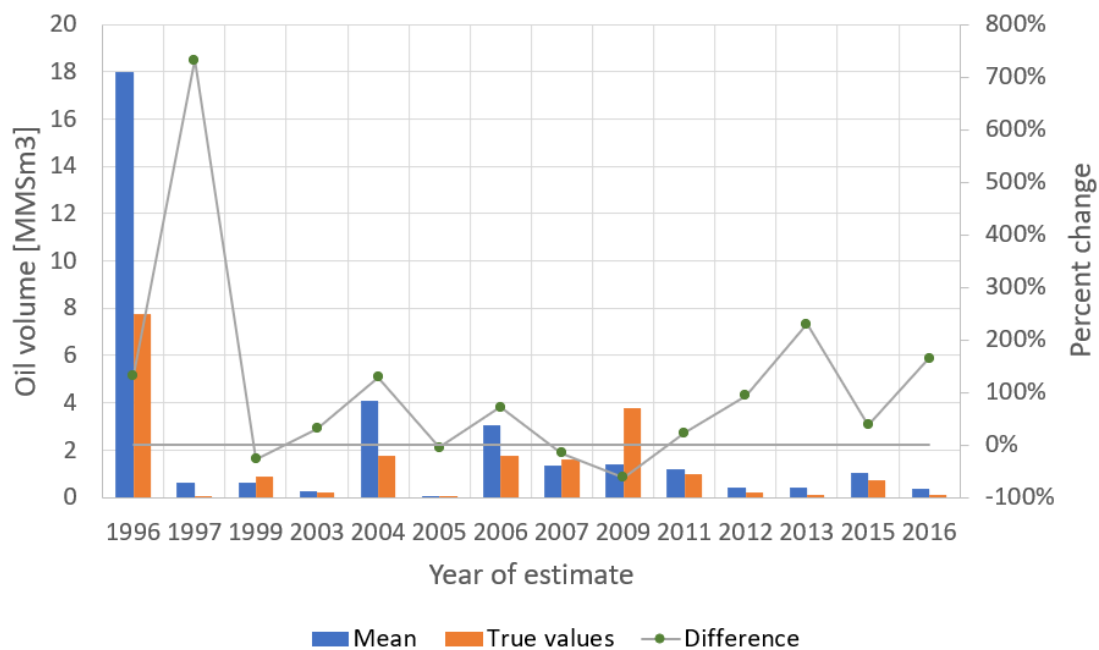
Figure 4.13: Estimates versus actual production outcomes, by the year in which the estimates were made
Year 0

Figure 4.13 illustrates the first year of production for the entire set of data, with the horizontal axis indicating the year in which the estimates were made, allowing interpretation of the development of Year 0 estimates over time. Figure 4.14 and 4.15 extends the analysis to also show the results for Year 1 and Year 2. The green line indicates the percentage difference between the mean estimates and the associated true values. When the green line is above the grey line, the forecasts are higher than the true production values. There are natural fluctuations from year to year, but there are no indications of decreasing levels of bias.
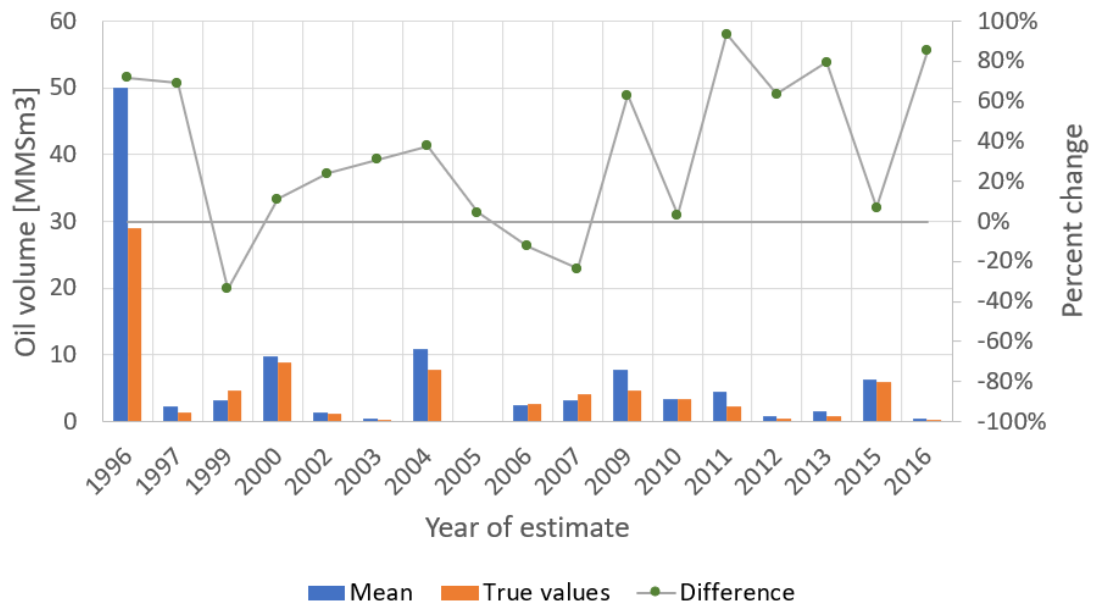
Figure 4.14: Estimates versus actual production outcomes, by the year in which the estimates were made
Year 1
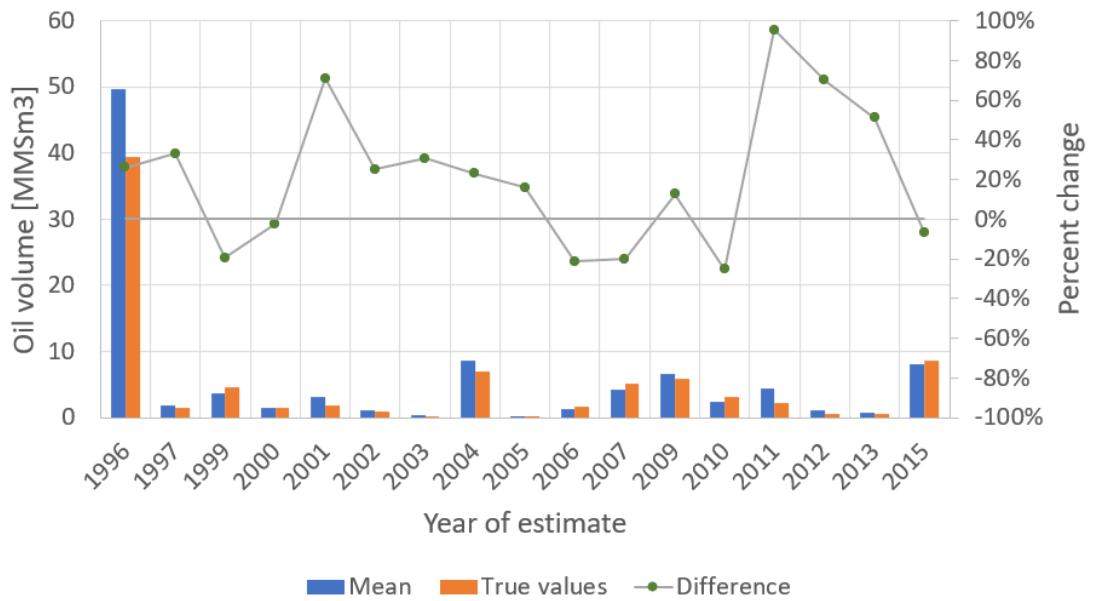
Figure 4.15: Estimates versus actual production outcomes, by the year in which the estimates were made
Year 2

Chapter 5

Reference class forecasting

5.1 Introduction to the modified reference class forecasting approach

In Chapter 4, forecasts from the NCS were compared to the historical production outcomes in order to evaluate the quality of the assessments. The evidence showed that the forecasts were biased, both in terms of being optimistic and overly precise. The following chapter will introduce the modified reference class forecasting (MRCF) approach and apply it to the data. This is done to see whether the methodology can correct the assessments and reduce the bias. Moreover, the MRCF approach will be compared to the classical reference class forecasting (CRCF) approach. MRCF is a more complicated model, using more inputs than the CRCF model. The hypothesis is that the modified model will perform better at debiasing the estimates than the classical model.

In the MRCF method, there are two sets of data: the reference class and an application class. These classes consists of data with common characteristics, pre-determined by the person or organization performing the calibration. This is the *class characteristic*. Examples include: similar geological structures, the same assessor, the same company, the same production year or the same year after production start. It is important that the expected amount of biases within and between the classes are homogeneous. In the reference class, both the historical true outcomes and the estimates must be known. In a realistic calibration scenario, the application class would consist of forecasts made today, i.e., where the true values are not yet known. However, for the purpose of this thesis, the application class constitutes a small sample of the reference class, in order to allow for the documentation of the effectiveness of the method. To ensure consistency, it has been chosen to always have 10 forecasts in the application class.

In this thesis, the class characteristic is "data of the same year after production start". Since it has been chosen to study only the first four years of production (see Section 4.2.1), the classes are: the "Year 0"-class, the "Year 1"-class, the "Year 2"-class and the "Year 3"-class.

In order to explain the calibration methodology of the MRCF method and the results that are gathered, some background information is necessary. First, the data is fitted to a Bi-Normal distribution. This is a flexible distribution capable of fitting any three percentile values. Then, an evaluation of the forecasts in the reference class is performed. By comparing these forecasts to the historical true values the calibration coefficients can be measured. Afterwards, the coefficients are applied to the application class and the results of the analysis is evaluated. Following this methodology, an out-of-sample analysis is conducted. This is done to measure the accuracy of the procedure over a large number of iterations, providing more statistical evidence.

Table 5.1 gives an overview of the nomenclature that is used in this chapter.

Table 5.1: Nomenclature for Chapter 5

$j \in J_R$	the set of reference class values
$j \in J_A$	the set of application class values
x_j^T	historical production outcomes
$x_j^{B,0.1}$	assessed (biased) P10 values
$x_j^{B,0.5}$	assessed (biased) P50 values
$x_j^{B,0.9}$	assessed (biased) P90 values
$x_j^{U,0.1}$	calibrated (unbiased) P10 values
$x_j^{U,0.5}$	calibrated (unbiased) P50 values
$x_j^{U,0.9}$	calibrated (unbiased) P90 values

5.2   Bi-Normal distribution

The Bi-Normal distribution is a distribution attributed to Clemen et al. [5]. As the name implies, the distribution consists of half a normal distribution from two different distribution. Mathematically, the probability density function (PDF) of the concatenated function is described as follows:

$$PDF = \begin{cases} \frac{1}{\sqrt{2*\pi*\sigma_L^2}} * exp(-\frac{(x-m)^2}{2*\sigma_L^2}); & \text{if } x < m \\ \frac{1}{\sqrt{2*\pi*\sigma_R^2}} * exp(-\frac{(x-m)^2}{2*\sigma_R^2}); & \text{if } x > m \end{cases} \tag{5.1}$$

where m is the median, σ_L is the standard deviation for the left side of the distribution and σ_R for the right. An example the Bi-Normal distribution is illustrated in Figure 5.1. It shows the left and right halves of two normal distributions. The two distributions have the same median, but different standard deviations.
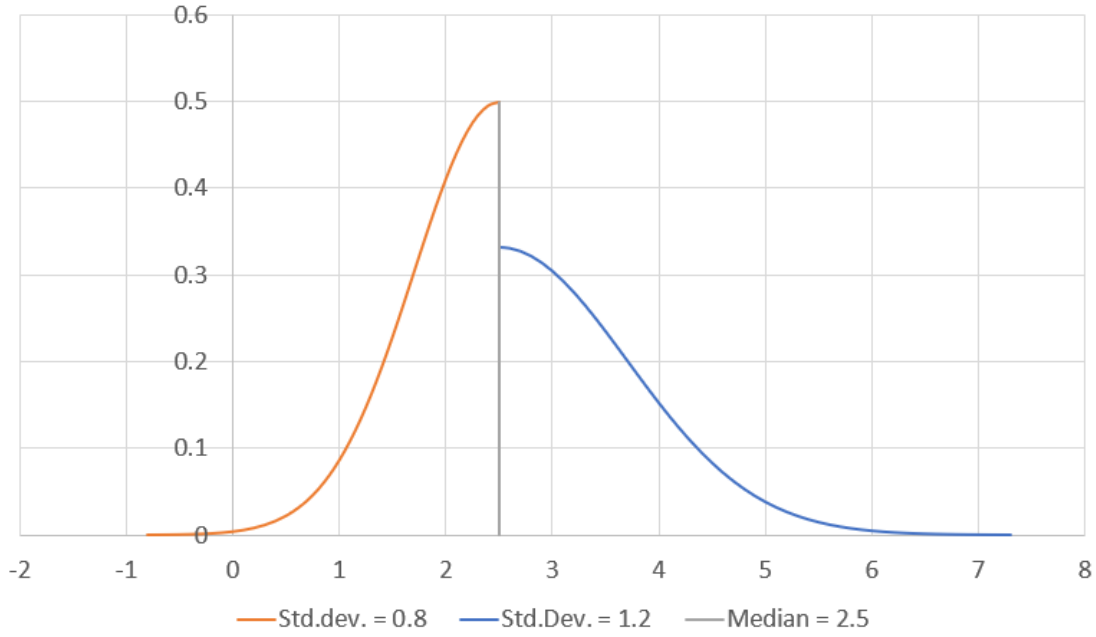
Figure 5.1: Example of a Bi-Normal distribution

The Bi-Normal distribution is used to fit a distribution to the assessed percentiles $x_j^{B,0.1}$, $x_j^{B,0.5}$ and $x_j^{B,0.9}$. The standard deviation for the left side of the distribution is estimated using:

$$\frac{x_j^{B,0.5} - x_j^{B,0.1}}{1.282} \tag{5.2}$$

where 1.282 is the standard z-score for an 80 percent confidence level, whilst the standard deviation for the right side of the distribution is estimated using:

$$\frac{x_j^{B,0.9} - x_j^{B,0.5}}{1.282} \tag{5.3}$$

The Bi-Normal fitting is repeated for every $j \in J_R$.

5.3   Assessing the quality of the forecasts

Prior to calibration, it is important to evaluate how the estimates in the reference class compares to the historical true production outcomes. As discussed in Section 2.5.1, 10 percent of the historical true values should fall below the assessed P10 estimates. Similarly, 10 percent of them should fall above the P90s. The number of true values should also be equal above and below the P50s. If the number of true values included in the reference class is not divisible by 10, then the amount of true values that should be below the P10- and above the P90 estimates is found through plotting probabilities, as discussed below.

To evaluate how many true values should fall below particular percentiles, a hypothetical perfectly calibrated judge is invented. The perfectly calibrated judge is an assessor whose cumulative probabilities of the true values perfectly align with a plotting position formula [6]:

$$p_j = \frac{(j - a)}{1 + J_R - 2a} \tag{5.4}$$

where J_R is the amount of true values in the reference class and a is a constant. a can take values between 0 and 0.5. Cunnane suggested in his research [6] that $a = 0.4$ would ensure minimum variance and, thus, this value has been chosen for this thesis.

There are 32 fields with assessments made for Year 0. 22 of these forecasts constitutes the reference class, whilst the remaining 10 constitutes the application class. Since 22 is not divisible by 10, a plotting position formula is needed to figure out how many data points should fall below P10 and P50 and below P90 and P50.

Table 5.2 illustrates how a perfectly calibrated judge would have assessed the cumulative probabilities (shown in the plotting probability column). Since all his forecasts are unbiased, his assessed percentiles are $x_j^{U,0.1}$, $x_j^{U,0.5}$ and $x_j^{U,0.9}$. With these three percentiles, an entire distribution can be fitted using the Bi-Normal distribution.

To better illustrate how a perfectly calibrated judge makes his forecasts, two hypothetical fields (field F and field G) are introduced.

Assume that the perfectly calibrated judge made an assessment for the Year 0 production for field F, illustrated in Table 5.3, and that the true production value for this field is 3.1. This is below the P50, placing it in the left side of the distribution. The cumulative distribution can now be calculated using the formula presented in Equation 5.1 (where $x < m$) and integrating it from 0 to the true the true value; 3.1. The result from this exercise is 0.117, which is found in Table 5.2.

Table 5.2: Plotting positions and the associated plotting probabilities, following the plotting position formula 5.4

Plotting position	Plotting probability
1	0.027
2	0.072
3	0.117
4	0.162
5	0.207
6	0.252
7	0.297
8	0.342
9	0.387
10	0.432
11	0.477
12	0.522
13	0.568
14	0.613
15	0.658
16	0.703
17	0.748
18	0.793
19	0.839
20	0.883
21	0.928
22	0.973

Further assume that the perfectly calibrated judge makes a forecast for Year 0 for field G, presented in Table 5.4, and that the true value for this field is 12.8. This is above the P50, meaning the right-side distribution will be used. Calculating the cumulative probability for this outcome yields 0.883 - equal to one of the other numbers in the plotting probability. This applies for every estimate the perfectly calibrated judge make, eventually matching all the values in Table 5.2.

For the perfectly calibrated judge, two data points falls below 0.1 (P10) and two data points lie above 0.9 (P90), as seen in Table 5.2. Moreover, 11 data points lies above 0.5 (P50) and 11 data points falls below. It is important to emphasize that the perfectly calibrated judge does not perfectly forecast future estimate. He is simply making forecasts without biases.

5.4   Estimating the calibration coefficients

In order to reduce forecasting biases, three calibration coefficients are calculated - the β-coefficient to calibrate the median, the α_L-coefficient to calibrate the P10

Table 5.3: Assessment of Year 0 for field F by a perfectly calibrated judge

Field F			
Year	P10	P50	P90
0	3	4	6

Table 5.4: Assessment of Year 0 for field G by a perfectly calibrated judge

Field G			
Year	P10	P50	P90
0	5	9	11

and the α_R-coefficient to calibrate the P90.

22 out of the 32 available data points for Year 0 are chosen at random. The estimates of the 22 fields are listed, and compared with the associated true values, in Table 5.5:[1]

1. First, β is estimated:
   Table 5.6 shows the distribution of true values above and below P50 from the data presented in Table 5.5. Not surprisingly, a majority of the true values are below the assessed P50.

   All of the median estimates are now multiplied with β, which initially is set to 1. β starts at 1 such that if the assessments are not biased, they are not calibrated. If unbiased data is calibrated, the calibration itself will introduce bias. β is then *reduced* in increments of 0.0001.[2] The rationale behind using this incrementation is that it emulates a continous distribution of β values. The process is repeated until the condition of Equation 5.5 met. In this particular case, the condition is met at $\beta = 0.6887$. The number of true values below P50 now aligns with that of the perfectly calibrated judge - 11 values.

$$Median(\beta x_j^{B,0.5} - x_j^T) = 0 \tag{5.5}$$

2. After calibrating the P50 values, the next step is to estimate α_L:
   With this method, the P10 values will be affected by the change of the P50 values that resulted from the multiplication with β. The new P10 values after the P50s are calculated using Equation 5.6.

$$x_j^{U,0.5} - (x_j^{B,0.5} - x_j^{B,0.1}) \tag{5.6}$$

---

[1]The fields are labelled from 1 to 56 due to there originally being 56 fields in the data set.

[2]In situations where there are less true values below P50 than that which the perfectly calibrated judge indicates, β is increased.

Table 5.5: Median estimates and the associated true values for 22 of the original 32 data points in Year 0

Field	P10	Median	P90	True
1	0.179	0.257	0.332	0.259
4	0.042	0.046	0.050	0.049
5	0.153	0.183	0.200	0.008
6	0.234	0.347	0.389	0.131
15	0.116	0.166	0.216	0.241
17	0.512	0.513	0.514	2.968
19	0.830	1.428	1.560	0.069
20	0.270	0.420	0.450	0.038
23	0.255	0.426	0.596	0.219
26	1.518	1.566	1.806	1.070
29	0.079	0.085	0.126	0.050
30	0.427	0.558	0.645	0.407
31	0.510	0.520	0.560	0.367
32	3.710	4.510	5.310	0.415
33	2.543	3.207	3.965	1.432
34	0.400	0.500	0.590	0.911
37	0.202	0.283	0.300	0.216
38	0.042	0.108	0.179	0.516
52	1.924	2.375	2.645	0.628
54	2.895	3.049	3.084	1.770
55	0.080	0.091	0.114	0.117
56	2.072	3.116	4.073	3.901

Table 5.6: Distribution of true values above and below P50 for Year 0

Below P10	Above P50	Below P50	Above P90
13	8	14	5

There are then eight true values below the P10. Since the P10 values still need to be calibrated, Equation 5.7 is used to estimate the α_L calibration coefficient:

$$x_j^{U,0.5} - \alpha_L(x_j^{B,0.5} - x_j^{B,0.1}) \tag{5.7}$$

Again, α_L starts out at 1 and is increased by 0.0001 until the amount of true values below P10 is aligned with the perfectly calibrated judge.[3] Table 5.5 indicates that two values should be below P10 (two values below 0.1). The value of α_L is 2.2340 after the calibration procedure. After the calibration

---

[3]In situations where there are less true values below P10 than that which the perfectly calibrated judge indicates, α_R is decreased.

of the P10 values, it is likely that some of the values are below 0. Since a negative production volume is irrational, a lower bound of zero has been imposed.[4]

3. Lastly, the same procedure used to calibrate α_L is used to calibrate α_R, where Equation 5.8 indicates the new P90 values after P50 has been calibrated and Equation 5.9 explains how α_R is estimated.

$$x_j^{U,0.5} - (x_j^{B,0.5} - x_j^{B,0.9}) \tag{5.8}$$

$$x_j^{U,0.5} - \alpha_R(x_j^{B,0.5} - x_j^{B,0.9}) \tag{5.9}$$

After the calibration procedure, α_R is 6.1795.

5.5 Applying the calibration coefficients

After the calibration procedure, the calibration coefficients are applied to the application class. The uncalibrated data set is presented in Table 5.7.

Table 5.7: An overview of the application class estimates before calibration

Field	P10	Median	P90	True
2	0.028	0.058	0.061	0.038
12	0.313	0.415	0.460	0.126
14	0.875	1.031	1.040	0.748
21	0.500	0.756	0.850	0.361
36	0.172	0.194	0.305	0.036
41	0.307	0.481	0.609	0.642
45	1.217	1.353	1.515	1.596
46	0.414	0.744	1.087	0.191
49	2.240	2.308	2.330	0.010
51	1.568	1.764	1.766	0.519

Table 5.7 is summarized in Table 5.9. Seven data points fall below P10 and two data points above P90. This means that the 80 percent confidence interval only capture one of the data points in the set, indicating a high level of bias.

To apply the calibration coefficients, 0.6887 is first multiplied with the P50 values. Second, 2.2340 and 6.1795 are inserted for α_L and α_R in Equation 5.7, 5.9, respectively. Table 5.8 shows the entire application class after calibration. The distribution of true values compared to the assessments is seen in Table 5.9. It

---

[4]This truncation could have been avoided, as suggested by Clemen et al. [5], by log-transforming the data prior to calibration.

becomes evident that, although not perfectly matching the perfectly calibrated judge, the calibrated data seems much less biased than the uncalibrated. For instance, with the calibrated data set, 70 percent of the true values falls within the 80 percent confidence interval of the assessors. More results, and discussion around the results of the debiasing technique will be presented later.

Table 5.8: An overview of the application class estimates before calibration

Field	P10	Median	P90	True
2	0	0.040	0.059	0.038
12	0.059	0.285	0.566	0.126
14	0.362	0.710	0.764	0.748
21	0	0.521	1.102	0.361
36	0.085	0.134	0.821	0.036
41	0	0.331	1.122	0.642
45	0.629	0.932	1.935	1.596
46	0	0.513	2.633	0.191
49	1.437	1.590	1.724	0.010
51	0.777	1.215	1.266	0.519

Table 5.9: Distribution of true values for the application class above P50 and P90 and below P50 and P10 for Year 0

	Below P10	Above P50	Below P50	Above P90
Uncalibrated	7	2	8	2
Calibrated	3	3	7	0
Perfectly calibrated judge	2	5	5	2

5.6 Out-of-sample analysis

The previous section showed that the calibration technique aided in debiasing the assessors' forecasts. The procedure split a data set into two parts - the reference class and the application class. The reference class was the set where the calibration coefficients were established and the application class was where these coefficients were applied. This emulates how the technique would be implemented in a real-world scenario, where the application class constitutes the forecasts where the true values are not yet known. This kind of analysis is called out-of-sample analysis.

For this thesis, out-of-sample analysis is used in order to objectively measure the debiasing effect of the presented methodology. However, instead of using just a single set of random values in the reference- and application class, 1000 iterations of randomly selected data sets in these classes are run through. This is done in order to evaluate whether the debiasing approach is robust across a variety of different data sets. The procedure presented earlier is used, randomly selecting new sets of reference- and application classes for every iteration. By storing the results of

every iteration, is is possible to average them at the end of the procedure. The computer code is provided in Appendix D. Appendix C also includes illustrations of the Excel worksheet that was used, in order to explain how the code and the worksheet function together.

The reason why 1000 iterations is chosen is because any additional iterations would provide an insignificant change in the results. This is evaluated by taking the cumulative moving average (CMA) of the calibration coefficients and study when it stabilizes. Figure 5.2 illustrates the results from the CMA exercise on data from Year 0. It is evident that 1000 iterations is more than enough to stabilize the calibration coefficients. Similar is also true for the other years (see Appendix B).
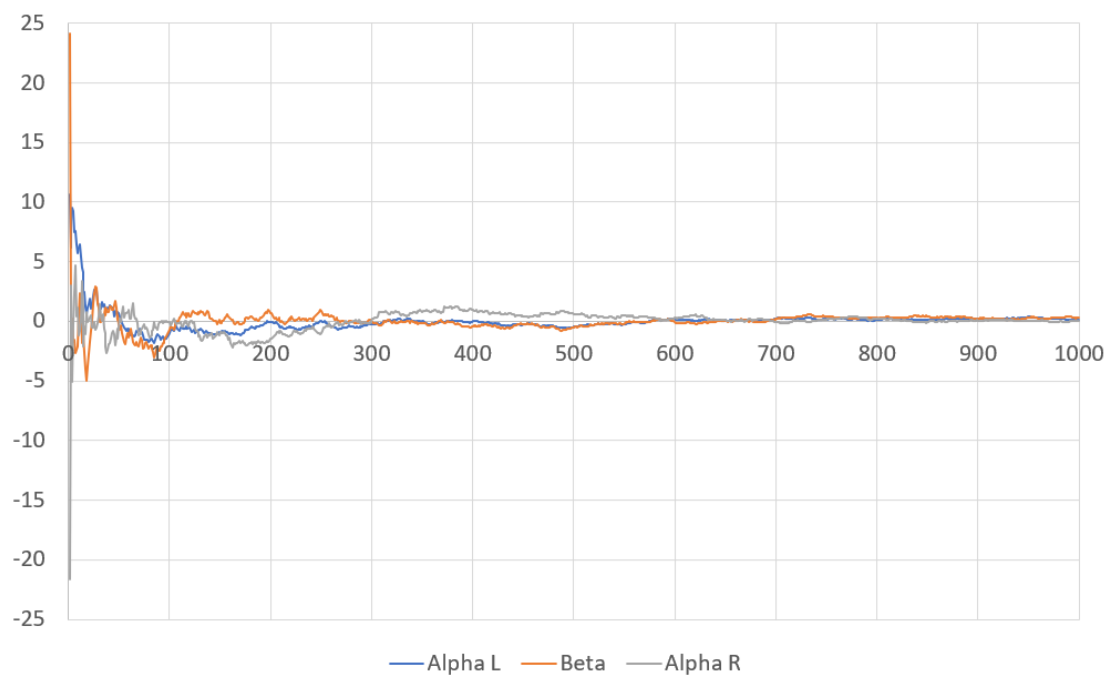
Figure 5.2: CMA of the calibration coefficients

5.7 Measuring the effects of debiasing

To analyse whether the modified reference class forecasting technique reduced bias - and if so, to what extend - some pre-defined ways to measure it is necessary. All the results will be presented either in terms of the average across 1000 iterations or in terms of a single iteration. The distinction will be clarified to avoid ambiguity.

The first method to measure improvements is to take the average absolute deviation between the assessments and the perfectly calibrated judge. This will give an indication of the average amount of bias in the original assessments and how effective the calibration technique is.

The second method is to plot the cumulative probabilities against the plotting positions. Both the calibrated- and uncalibrated cumulative probabilities should

be included in this plot, together with the plotting probabilities. Figure 5.3 illustrates how this would look for a single iteration. It is evident that there are many instances where the cumulative probabilities are zero or close to zero. This applies for 6 out of the 10 plotting positions in this example. As will be shown later, this is not an uncommon characteristic of the uncalibrated data.

It is also possible to calculate the average cumulative probability for each of the plotting positions for all the iterations. This provides an overview of the differences between the calibrated- and uncalibrated assessments, and the plotting probabilities.
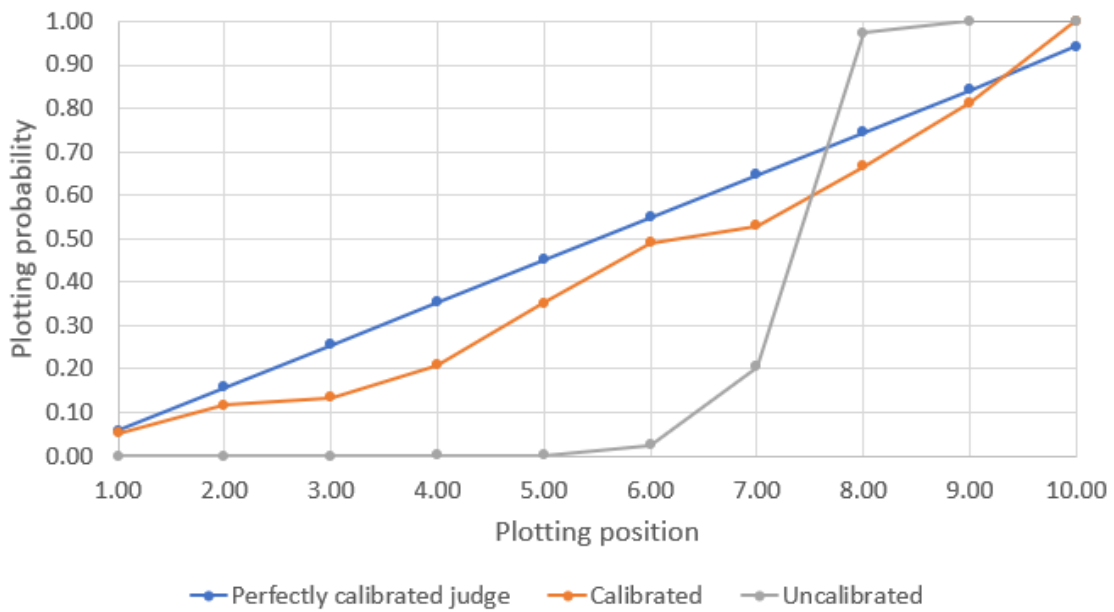
Figure 5.3: Plotting probabilities - cumulative probabilities versus plotting positions

Another way of presenting the data in Figure 5.3 is to take the squared difference between all the cumulative probabilities and the plotting probabilities. This yields the root mean square error (RMSE), which is commonly used when analysing forecasts. The formula for RMSE is presented in Equation 5.10.

$$RMSE = \sqrt{\frac{\sum_{j=1}^{J}(C_j - p_j)^2}{J}} \qquad (5.10)$$

where J is the total number of true values, C_j is the cumulative probability and p_j is the plotting probability for the jth plotting position. The probabilities must be plotted in ascending order.

Table 5.10: Absolute difference of true values above/below percentiles between the perfectly calibrated judge and the assessments, Year 0

Percentile	Uncalibrated	Calibrated	Improvement
Below P10	5.276	0.893	83%
Above P90	1.289	0.922	28%
Above/Below P50	1.942	1.588	18%

Table 5.11: Absolute difference of true values above/below percentiles between the perfectly calibrated judge and the assessments, Year 1

Percentile	Uncalibrated	Calibrated	Improvement
Below P10	3.683	0.827	78%
Above P90	1.094	0.808	26%
Above/Below P50	2.427	1.317	46%

Table 5.12: Absolute difference of true values above/below percentiles between the perfectly calibrated judge and the assessments, Year 2

Percentile	Uncalibrated	Calibrated	Improvement
Below P10	2.413	0.893	63%
Above P90	1.084	0.862	20%
Above/Below P50	1.436	1.348	6%

Table 5.13: Absolute difference of true values above/below percentiles between the perfectly calibrated judge and the assessments, Year 3

Percentile	Uncalibrated	Calibrated	Improvement
Below P10	2.205	0.845	62%
Above P90	0.736	0.834	-13%
Above/Below P50	1.432	1.403	2%

5.8 Results

The absolute difference between the assessments and the perfectly calibrated judge is illustrated in Table 5.10, 5.11, 5.12 and 5.13. The improvement from the uncalibrated to the calibrated is seen in the rightmost column in the tables. The tables indicates good results in terms of reducing the optimism and overconfidence. For the first year of production, the P10 estimates were 83 percent more accurate (according to this specific metric). For the most part, the calibration also aids in reducing bias in the other percentiles, although to a lesser degree than for the P10.
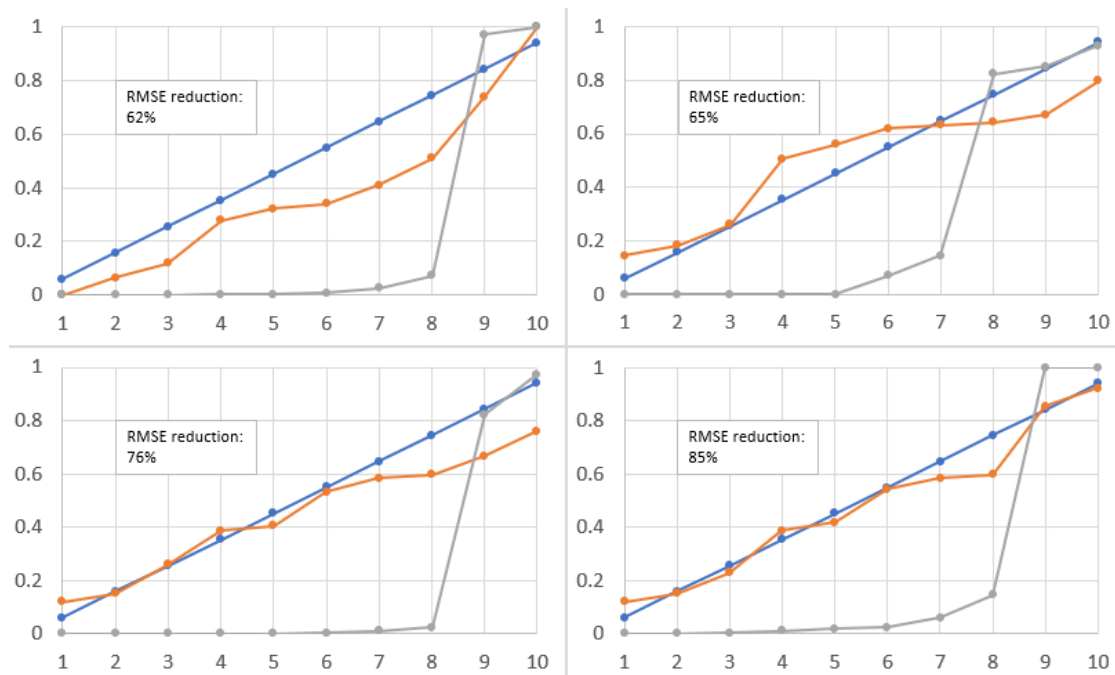
Figure 5.4: Plotting probabilities - Year 0, four random iterations
Blue = Perfectly calibrated judge, Orange = Calibrated, Gray = Uncalibrated

Figure 5.4, which is a combination of four subfigures, illustrates the efficiency of the calibration technique on an iteration-by-iteration basis for Year 0. RMSE reductions of more than 60 percent is seen for all these four iterations. The good results illustrated in this figure are also evident when averaging across 1000 iterations, seen in Figure 5.5.

The 1000 iteration-averages of Year 1, Year 2 and Year 3 is compiled in Figure 5.6. All the plots indicate that the calibration technique is useful in enabling the inclusion of more true values in the assessed 80 percent confidence intervals - mostly achieved through expanding the lower side of the interval.

Table 5.14 shows the RMSE values for both the calibrated- and uncalibrated data, as well as the improvements brought on by the calibration technique. The table shows the same trend as seen in the previous tables, with improvements in the first couple of years and reduced improvements in the following years. However, the RMSE improvements are significant for all the studied years.
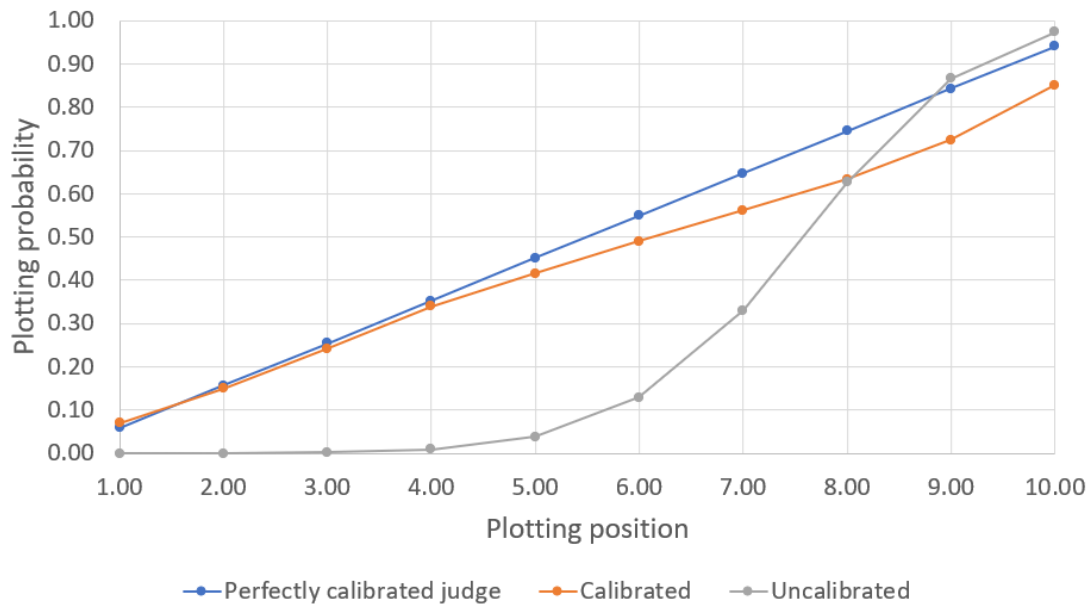
Figure 5.5: Plotting probabilities - Year 0, 1000 iterations
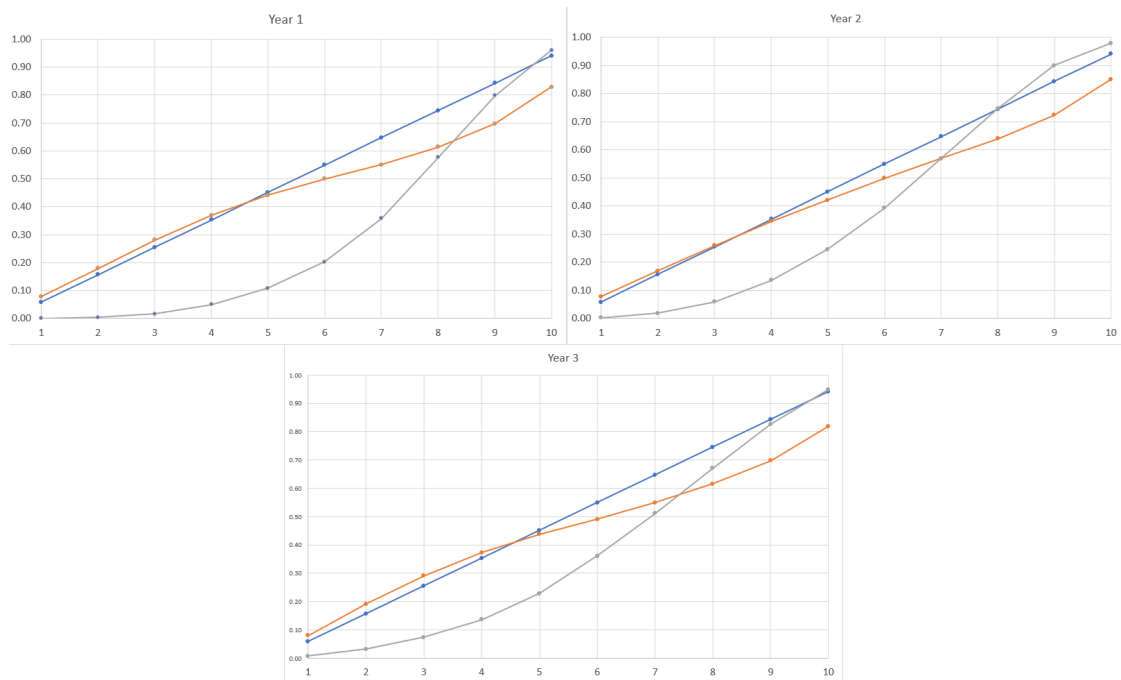
Figure 5.6: Plotting probabilities - Year 1, 2 and 3 - 1000 iterations

Table 5.14: Average RMSE across 1000 iterations, Year 0 through 3

Year	Uncalibrated	Calibrated	Improvement
0	0.32	0.14	55%
1	0.28	0.13	52%
2	0.20	0.13	34%
3	0.19	0.14	28%

A recurring theme in the graphs and tables in this section is that the calibration increases the accuracy of the assessments and reduces the bias, particularly for the first years. Figure 5.7 highlights the magnitude of the calibration coefficients for each year. It is evident that they are approaching 1. This indicates that the PDO assessments in this data set are less biased for later years, which is also evident from the "uncalibrated" column in, e.g., Table 5.14.
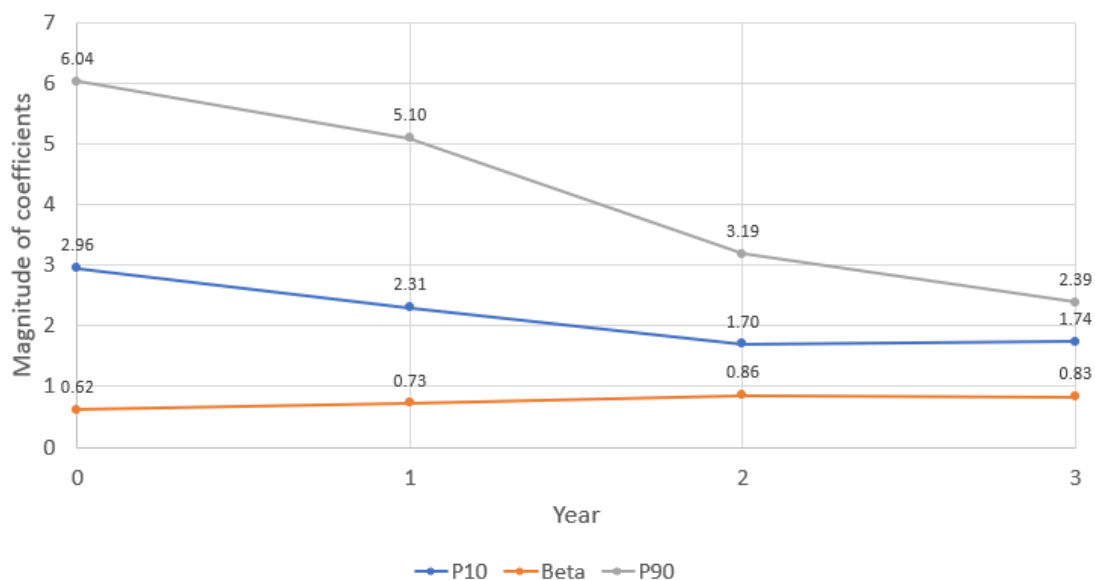
Figure 5.7: Calibration coefficients trend

5.9 Classical reference class forecasting

This section will present the results from applying the CRCF method to the same data set. Section 6.4 will compare the MRCF method to the CRCF method.

5.9.1 The method

As with MRCF, CRCF uses two sets of data - a reference class and an application class for each year, where the entire set of assessments is divided into one of these categories at random. All the true values are then normalized with the mean estimates and ordered in ascending order. Based on this list of ascending values, a metalog distribution [21] is fitted. The calibration coefficients are the output

of the metalog distribution. These coefficients are then applied to the application class, similar to the MRCF method.

5.9.2 Results

Table 5.15: Absolute difference of true values within percentiles between the perfectly calibrated judge and the assessments, Year 0

Percentile	Uncalibrated	Calibrated	Improvement
Below P10	5.249	0.724	86%
Above P90	1.283	0.785	39%
Above/Below P50	1.908	1.534	20%

Table 5.16: Absolute difference of true values within percentiles between the perfectly calibrated judge and the assessments, Year 1

Percentile	Uncalibrated	Calibrated	Improvement
Below P10	3.598	0.690	81%
Above P90	1.172	0.907	23%
Above/Below P50	2.311	1.405	39%

Table 5.17: Absolute difference of true values within percentiles between the perfectly calibrated judge and the assessments, Year 2

Percentile	Uncalibrated	Calibrated	Improvement
Below P10	2.420	0.617	75%
Above P90	1.103	0.896	19%
Above/Below P50	1.477	1.530	-4%

Table 5.18: Absolute difference of true values within percentiles between the perfectly calibrated judge and the assessments, Year 3

Percentile	Uncalibrated	Calibrated	Improvement
Below P10	2.183	0.774	65%
Above P90	0.758	0.822	-8%
Above/Below P50	1.416	1.317	7%

The absolute difference between the assessments and the perfectly calibrated judge after the CRCF method has been applied is seen in Table 5.15, 5.16, 5.17 and 5.18. The tables shows improvements with the calibration technique for most of the reviewed percentiles. The improvements are similar to that of the MRCF method. Table 5.19 investigates the differences between the two methods. The

values in the table express the percent change when going from the CRCF method to the MRCF method. Positive numbers indicate that the MRCF method performs better than the CRCF method and vice versa for negative numbers. By using this measurement, the modified method does not bring any improvement over the classical method and, in fact, it results in significantly worse results for the P10 calibrated assessments.

Table 5.19: Improvements seen when choosing MRCF instead of CRCF

	Year 0	Year 1	Year 2	Year 3
Below P10	-23%	-20%	-43%	-9%
Above P90	-17%	11%	4%	-1%
Above/Below P50	-4%	6%	12%	-7%
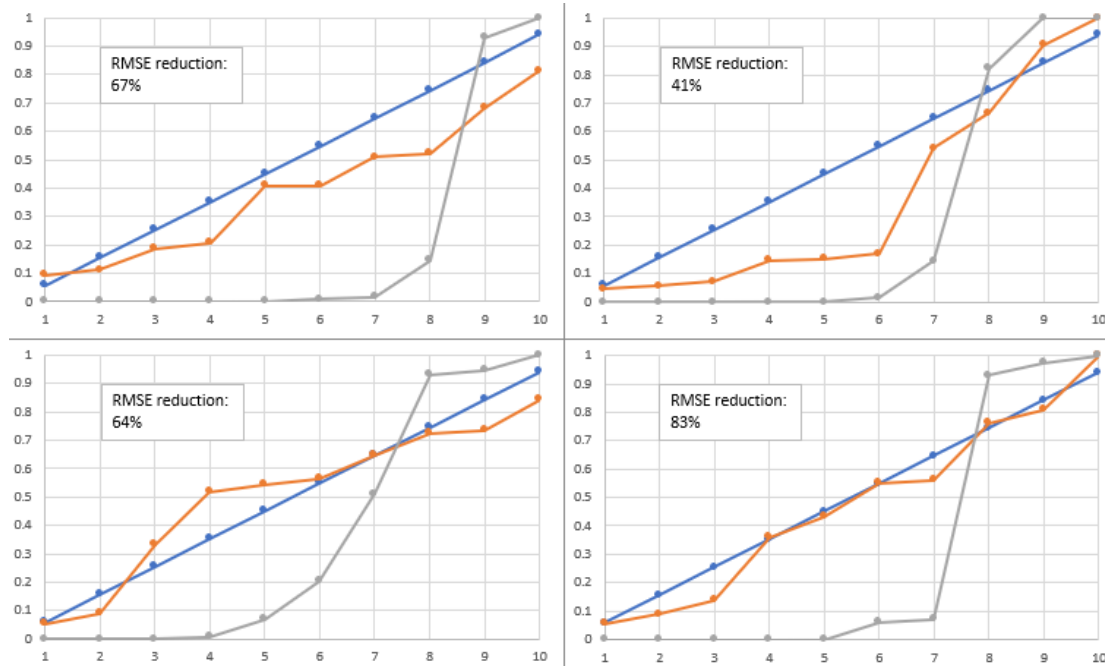
Figure 5.8: Plotting probabilities - Year 0, four random iterations
Blue = Perfectly calibrated judge, Orange = Calibrated, Gray = Uncalibrated

Figure 5.8 illustrates how the cumulative probabilities of the calibrated- and uncalibrated assessments match up to the perfectly calibrated judge in four random iterations of Year 0. There are some variations in the effectiveness of the calibration procedure, but it is always outperforming the uncalibrated data. It does not seem to be noticeably different from the MRCF method, shown in Figure 5.4.

Figure 5.9 illustrates the average cumulative probabilities across 1000 iterations for Year 0. This appears almost identical to that of the MRCF method, shown in Figure 5.5. Moreover, Figure 5.10 illustrates the average cumulative probabilities for Year 1, 2 and 3 as well. Compared to the MRCF method in Figure 5.6, the
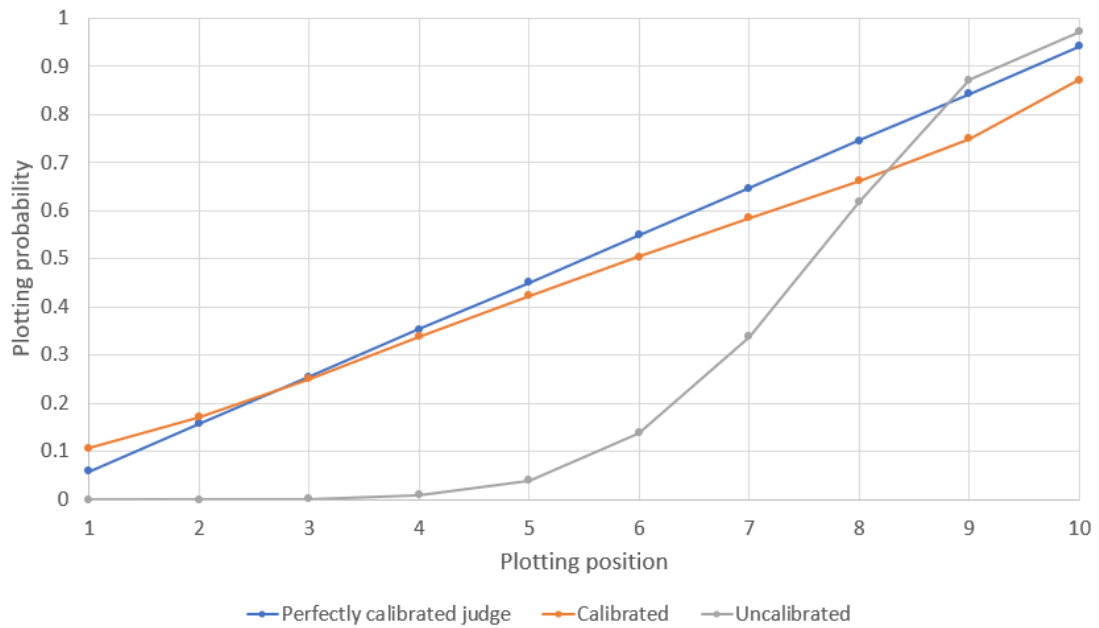
Figure 5.9: Plotting probabilities - Year 0, 1000 iterations

CRCF method is more accurately calibrating the P90 values. From the uncalibrated data, it is evident that the assessors are fairly good at estimating the P90 values correctly, and the MRCF method subsequently fails by overextending this measurement.
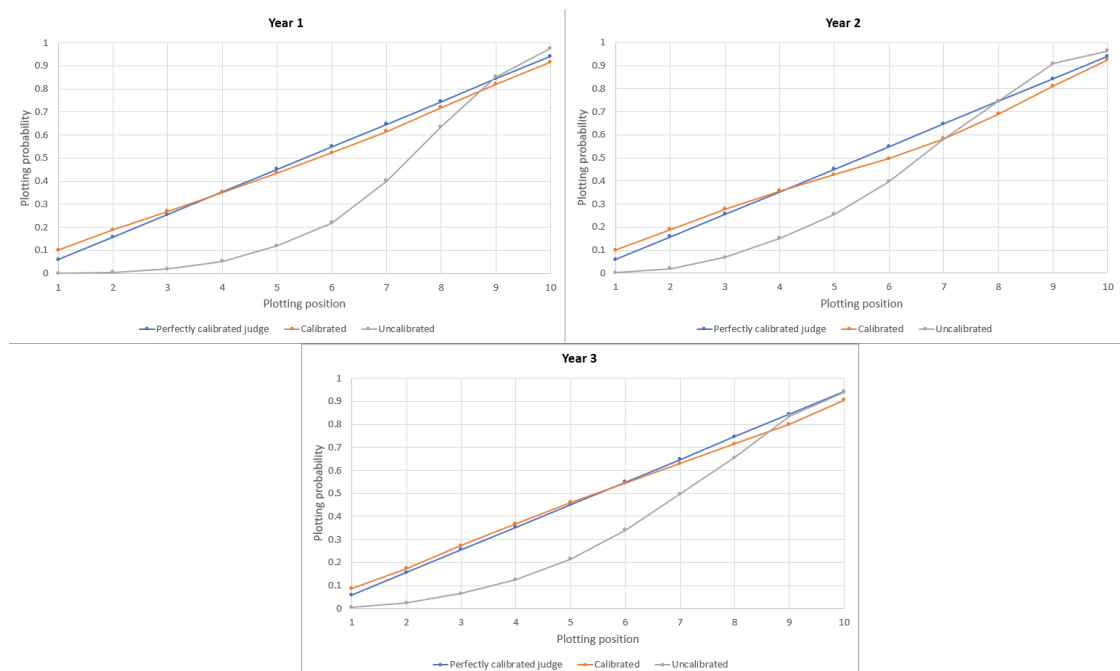
Figure 5.10: Plotting probabilities - Year 1, 2 and 3 - 1000 iterations

The average RMSE values and improvements is seen in Table 5.20. Table 5.21

compares the RMSE values from the MRCF and the CRCF method. Evidently, there are not much difference between the two. In fact, the CRCF method seems to be slightly more efficient than the MRCF method, in particular for Year 3.

Table 5.20: Average RMSE across 1000 iterations, Year 0 through 3

Year	Uncalibrated	Calibrated	Improvement
0	0.32	0.14	57%
1	0.27	0.12	54%
2	0.20	0.13	36%
3	0.19	0.12	35%

Table 5.21: Comparison of the RMSE improvements between MRCF and CRCF

Year	MRCF	CRCF
0	55%	57%
1	52%	54%
2	34%	36%
3	28%	35%

Chapter 6

Discussion

In the previous chapters, evidence of overprecision and optimism in production estimates have been discussed. Moreover, debiasing techniques have been applied to the data in order to evaluate whether it could improve the estimates. The following chapter will include a discussion about biases and the reasons why they are so evident in the data. There will also be a discussion about the debiasing techniques and the differences between them.

6.1 Explaining forecasting errors

6.1.1 Overruns in other industries

First, it is important to note that biased forecasts, similar to that which have been shown in this thesis, are not limited to forecasts on the NCS. Neither is it limited to forecasts in the petroleum industry. Several papers have investigated overruns in a multitude of industries, e.g. Flybjerg (2002) [16] and Flyvbjerg (2007) [14]. They conclude that, as a general rule of thumb, projects will arrive late, over budget and fail to meet expectations [13]. Overruns of more than 50 percent is common. Overruns of more than 100 percent is not uncommon. Although historical data and research suggests that overruns are common across a multitude of industries it should not be used as an excuse to continue to provide biased forecasts in the oil- and gas industry. In order to discontinue the trend and improve forecasts in the future, a better understanding of the challenges we face is needed.

6.1.2 Root causes

There are many reasons why projects underperform compared to the initial estimates. Scope changes, technological uncertainty and internal disagreements might all have a negative impact on projects, which could ultimately cause overruns, delays and reduced benefits [15]. These are typical reasons as to why projects do not

meet expectations and can be categorized as "causes". However, Flyvbjerg et al. [13] argues that these explanations are not the *root* causes. Rather, they focused on three causes they assumed to be the true, underlying causes for the lack of performance:

1. Bad luck
2. Delusion
3. Deception

6.1.3 Bad luck

In situations where project expectations misaligns with the outcome - either in terms of schedule, cost or quality - it is easy to blame bad luck. It is not uncommon that managers rationalize in this way. Why? Because it pushes the responsibility away from managers over to an intangible scapegoat.

The theory around black swans was introduced by Nassim Taleb in his 2007 publication "The Black Swan" [40]. A black swan is an event that is highly improbable - either assessed with a minuscule probability or not considered whatsoever. The financial crisis of 2007-2008 can be considered a black swan event because few, if any, knew that such an event would occur.

It may be argued that people who lost their lifetime savings during this period experienced a case of bad luck. However, bad luck should not be used to describe a situation where a gambler loses all of his savings in a casino. In this scenario the odds are stacked against him, and statistics will ensure that he will always lose in the long run. The gambler would likely disagree, however, insisting on his bad luck.

The aforementioned analogy was presented to showcase that bad luck could be a valid explanation in, e.g., a black swan event, but should not be used to describe all events ending with a negative outcome.

The problem with using bad luck as an explanation for all projects with poor outcome is that it is not statistically coherent. The bad luck-term, i.e., the error term, should have a mean in the vicinity of zero. If this assumption does not hold true, the model is biased. Based on the production estimates investigated in this thesis, it is clear that the data is biased and that bad luck or error cannot be the only explanation.

Furthermore, if bad luck could explain the poor data, it should be expected that with time, the estimates would improve and that the misalignment with reality would become smaller [13]. This should be expected as the potential benefit of improving predictions is usually vast. Improved models, more education and knowledge, and an ever-increasing amount of historical data should all help with

the reduction of forecasting errors. This can clearly be seen with the improvements in hurricane forecasting, illustrated in Figure 6.1. The forecasting error in, e.g., the 72 hours prediction has been improved from about 450 nautical miles to below 100 - an improvement of around 80 percent. From the data in this thesis, it was evident that there have been no signs of improvement across the sample period (see Section 4.3). More importantly, the data shows consistently optimistic forecasts, which disputes the argument that the poor data is a result of bad luck.

Advances in hurricane prediction

Data from the NOAA National Hurricane Center (NHC) (*13*) show that forecast errors for tropical storms and hurricanes in the Atlantic basin have fallen rapidly in recent decades. The graph shows the forecast error in nautical miles (1 n mi = 1.852 km) for a range of time intervals.
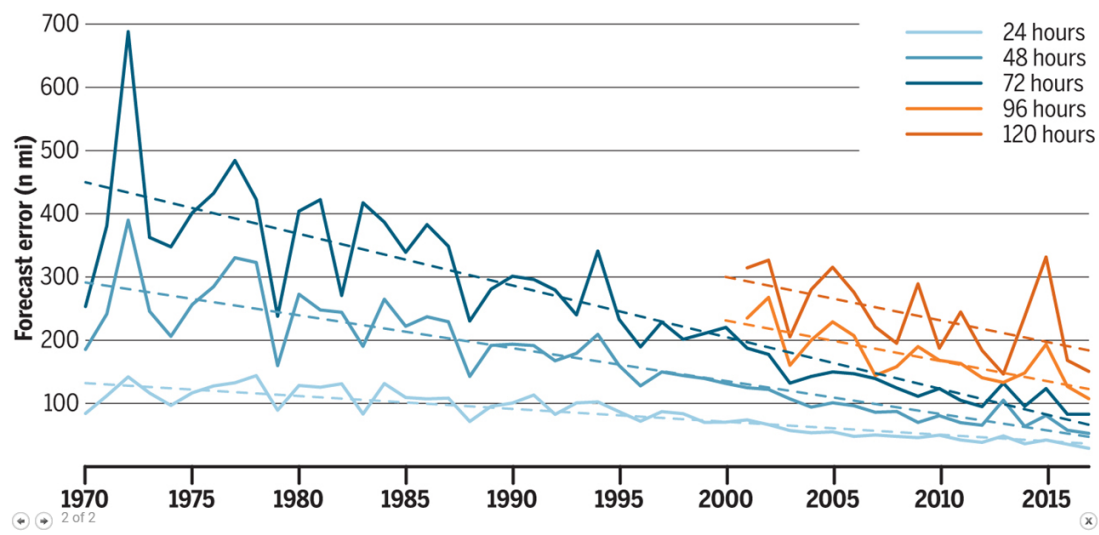
Figure 6.1: Improvements in hurricane forecasting [1]

6.1.4 Delusion

Bad luck is clearly a poor argument when trying to find the reasons why failing to meet expectations is so common. Flyvbjerg et al. [13] therefore propose another explanation: delusion. Too often, decision makers make decisions based on delusional optimism [18] rather than carefully and rationally weighting risks, probabilities, advantages and disadvantages. Benefits tend to be exaggerated, whilst underestimating costs. This also applies when people have access to historical data and are aware of the phenomenon. This behaviour is called *the planning fallacy* [20]. Most people are optimistic most of the time. People are specifically prone to overestimate their own abilities and their own ability to influence the outcome of a project [18].

Figure 6.2 shows the extent of the planning fallacy. The data is gathered from a study where forecasting teams were asked to assess the quality of their forecasts. The study shows no difference in accuracy between the teams who thought their forecast were excellent/good and the teams who assessed their forecast as

fair/poor.

People's optimistic nature could be part of the explanation why we see so poor assessments for the P10 values in particular. It could also help explaining why there are generally more true production outcomes below the P50 assessment than above.
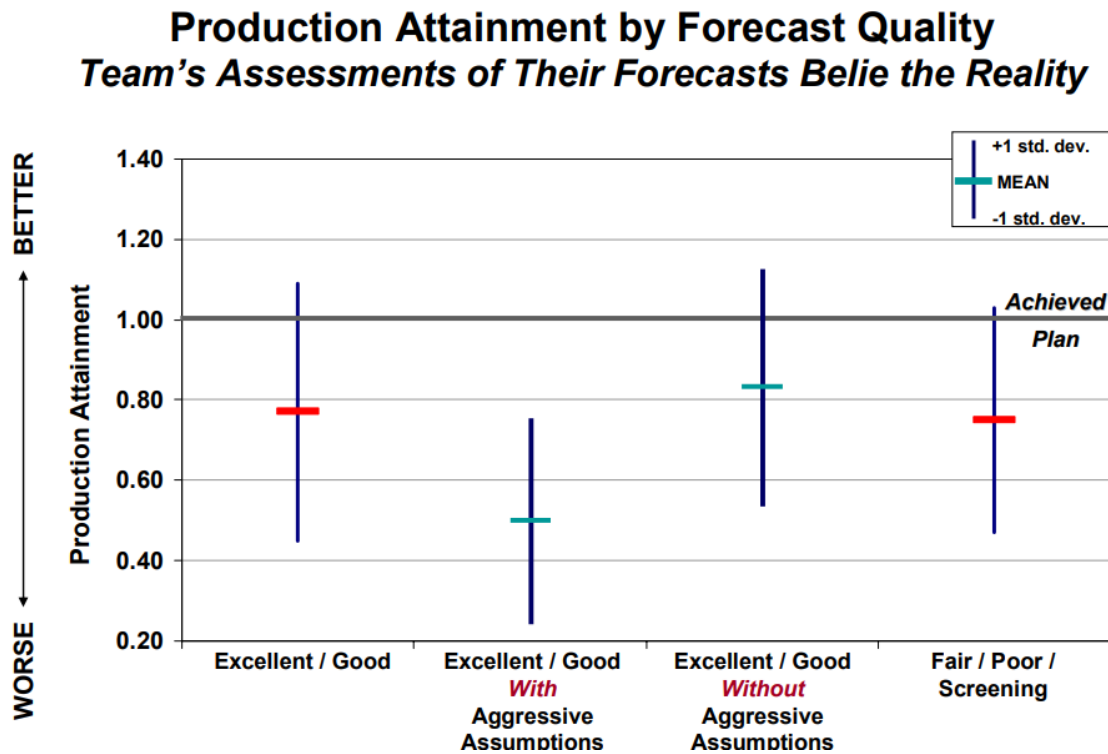
Figure 6.2: Assessment of the quality of forecasts [33]

Kahneman and Tversky wrote a great deal about *heuristics* in their research on cognitive biases [19]. A heuristic is a methodology applied by people in order to make sense of questions or problems that are otherwise difficult to solve. Heuristics are based on intuition and does not necessarily give an optimal, or even satisfactory solution to the problem. An example of a heuristic is the rule of thumb, where a *generalized* answer is applied to a *specific* question.

Frequently, people use the ease of which a specific event comes to mind in order to decide on the probability of such an occurrence [19]. This is called the *availability heuristic*. A good example is that the assessed probability of success for a given project would likely be exaggerated if recent successes came easily to mind for the assessors.

Another important heuristic is the tendency of *anchoring and adjusting*. This is also a theory that was constructed by Kahneman and Tversky [20]. It explains the human tendency of anchoring in on a base case-estimate - a value which is considered the most likely - and then adjusting the lower- and upper bounds ac-

cording to this value. The problem is that people generally put too much faith in the original estimate such that the lower- and upper bounds become biased. The following example will illustrate this phenomenon [19]:

Two groups of high school students were given the task to estimate, within a timeframe of 5 seconds, the answer to an arithmetic expression. The groups were given the same expression, mathematically speaking. However, the first group was given the expression 1, and the second, expression 2:

$$8 * 7 * 6 * 5 * 4 * 3 * 2 * 1 \tag{1}$$

$$1 * 2 * 3 * 4 * 5 * 6 * 7 * 8 \tag{2}$$

Because of the limited timeframe that was given, the students were only able to perform a few steps of the calculation and had to adjust from this value - the anchor - to achieve their answer. The evidence were conclusive. For the first group, the median of the students' estimates was 2250, while for the second group this value was 512.

Anchoring and adjusting is a particularly probable cause for some of the overprecision that is seen in the data in this thesis. It is not inconceivable that many of the assessors put a great deal of effort into creating an anchoring base case estimate - the "mean" - and then adjusting slightly upwards and downwards of this to create the P10 and P90.

Delusion stems from decision makers adopting an *inside view* [20]. The inside view is the term used when too much emphasis is put on particular details of the problem or project at hand. By ignoring the broader picture and falling into the pitfalls of delusion described above, bias is introduced to forecasts. Therefore, it is advantageous to adapt an *outside view*. By taking a step back and seeing the similarities between the current project and the projects of the past, historical data can be used to effectively reduce or eliminate a lot of the bias introduced as a consequence of delusion. This is the theory behind RCF.

6.1.5 Deception

The final root cause for underperformance, Flyvbjerg et al. [13] argues, is deception. Organizational deception is mainly caused by *the principal agent problem.*

The principal-agent problem occurs when a principal engages an agent to act on its behalf [13]. In general, such relationships exist between every two levels of an organizational hierarchy, e.g. a manager who engages an employee to act on his or

her behalf. The challenge that in every principal-agent relationship, the two parties usually have misaligning interests. If agents lack proper incentives to do the "right thing" by the principal, or is not penalized for promoting their own self-interests above the interests of the principal, it is likely that agents will misbehave. Typical issues include different time horizons, different risk preferences, uneven distribution of information between the principal and agent and diffuse accountability. These are all sources of strategic deception [13].

In large projects, such as offshore petroleum development projects, there are many principal-agent (P-A) relationships. This is known as a multi-tier P-A problem. An illustration is provided in Figure 6.3. This example is a generalization of a multi-tier (P-A) system and shows how principals and agents can be mapped. There are a number of other such relationships in a large scale project.
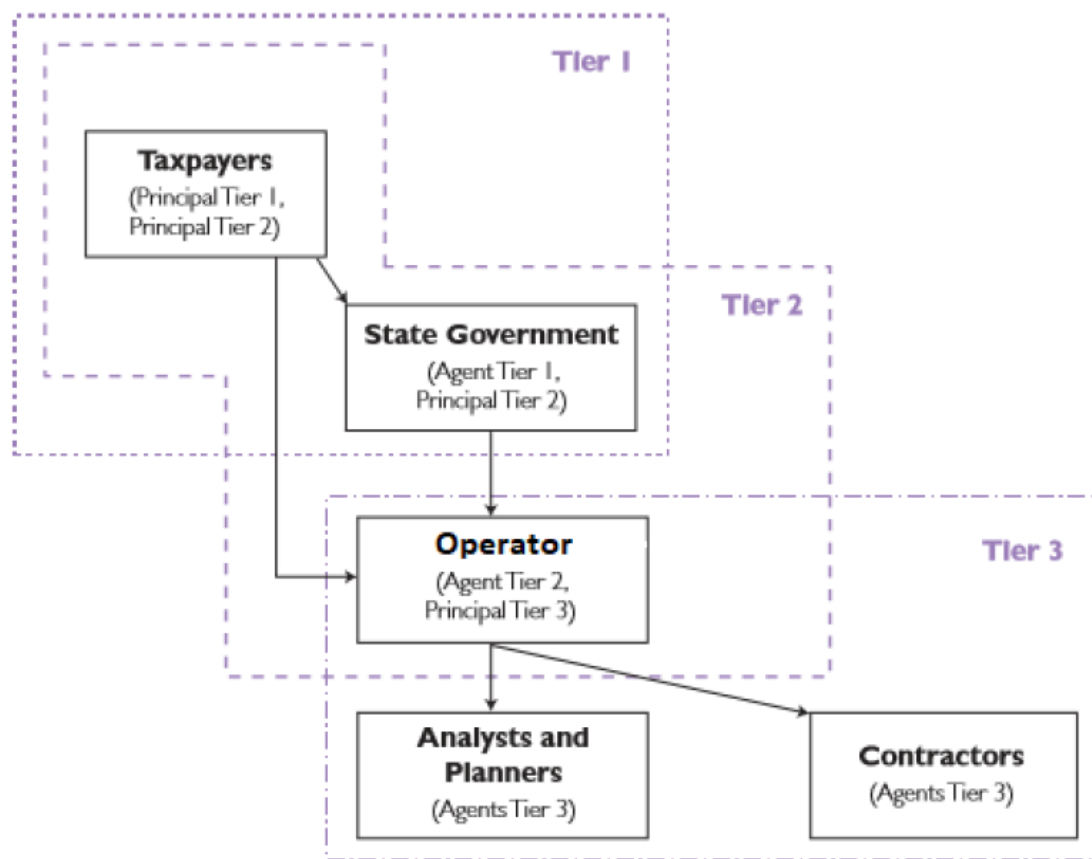
Figure 6.3: Example of multi-tier P-A relationships [32]

Tier 1 represents the P-A relationship between the taxpayers (principal) and the state government (agent). The Norwegian state government is elected by the taxpayers in order to act on their behalf. As discussed in Section 2.4, 78 percent of the revenue generated from petroleum activities accrues to the state government [27], which ultimately benefits the taxpayers. However, the taxpayers pay for 89 percent of the development costs [41]. Hence, taxpayers expect the state government to approve the PDOs that maximize the benefits. In this Tier 1 P-A relationship,

misalignments include different time horizons and different risk preferences. The Norwegian state government consists of members that are elected for four years at a time. Having to rely on a past accomplishments to remain relevant, politicians are incentivized to take more risk than their principals (the taxpayers) would like to. Moreover, as there are elections every fourth year, the state government has a shorter time horizon than the taxpayers, whose time horizon is generally long-term. These misalignments might not incentivize the state government to run proper quality controls of the assessments in the PDOs.

Tier 3 involves analysts and planners who act on behalf of an operator. In this P-A relationship, misalignments include different risk preferences, uneven distribution of information and diffuse accountability. The operator's primary objective is to maximize value from their assets. Analysts and planner, however, does not necessarily share this objective. They are the ones that sit on the majority of the information, and the ones assigned to develop forecasts to be included in the PDO. They are usually less concerned about risks due to diffuse accountability if forecasts end up being wrong. They are also incentivized to present forecasts that gives indications of a good field because it could ensure that they keep having a job to go to.

6.1.6 Delusion and deception

Flyvbjerg et al. [13] further argues that while both delusion and deception are the root causes of large-scale projects' failure, they are two inseparable, complementary concepts. However, in some occasions, the failure is easier explained by one, rather than the other. Their paper gives a few examples of certain projects that are more susceptible to either delusions or deceptions, shown in Figure 6.4.

It is evident that when learning is good (i.e., models are being continually updated with new information based on historical data as well as a keen focus on continuous education and development) and well aligned incentives and penalties are in place, forecasts tend to be unbiased [13]. Flyvbjerg et al. uses weather forecasts as an example where this applies. As shown earlier in Figure 6.1, it is a clear downward trend in the errors of hurricane forecasts, which supports this. Flyvbjerg et al. further points out the difference between rail- and road projects, which was studied by Flyvbjerg, Holm and Buhl [16].

Figure 6.5 and 6.6 illustrates the assessors' ability of forecasting costs related to rail- and road projects. Undoubtedly, the overruns for rail projects are significantly larger than for road projects. The reason for this is likely that rail projects occur a lot less frequent than road projects, limiting the amount of experience and improvement with such projects. Rail projects are also typically much larger than road projects, increasing the likelihood of principal-agent problems.

When it comes to forecasting of future production, both delusion and deception
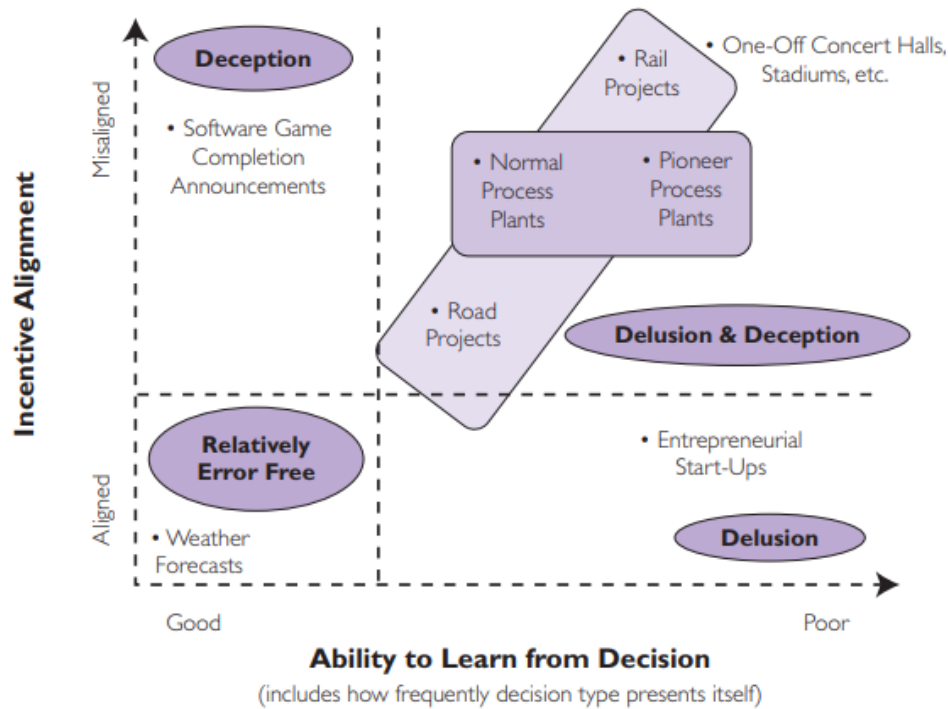
Figure 6.4: Delusion and deception in projects [13]

are causing biases. The bias observed in Chapter 4 could have been detrimental had it not been for the historically favorable prices for petroleum products. The following section will investigate the loss of value caused by bias on the NCS.

6.2   Loss of value due to biases

Mohus [32] analysed cost- and time overruns, as well as production shortfalls in the petroleum industry, with a specific emphasis on oil production fields. The paper includes a segment about the loss of value due to biased production estimates in the PDO. The present value (PV) is calculated using oil prices and exchange rates at the time of production. A 10 percent WACC and an inflation adjustment set to 1.24 is also factored in. The formula used was:

$$PV_i = \frac{Production_i * conversionrate * exhangerate_i * oilprice_i}{(1 + wacc_i)^t} \tag{6.1}$$

Mohus also provided an example of the calculation method for an arbitrary field on the NCS. The method is shown in Table 6.1.

Figure 6.5: Cost overruns in rail projects [16]
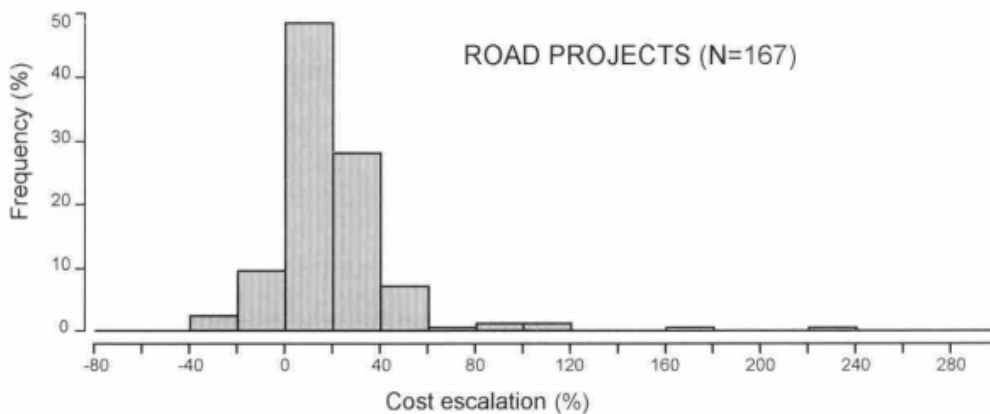
Figure 6.6: Cost overruns in road projects [16]

Table 6.1: Lost revenue for an arbitrary field on the NCS [32]

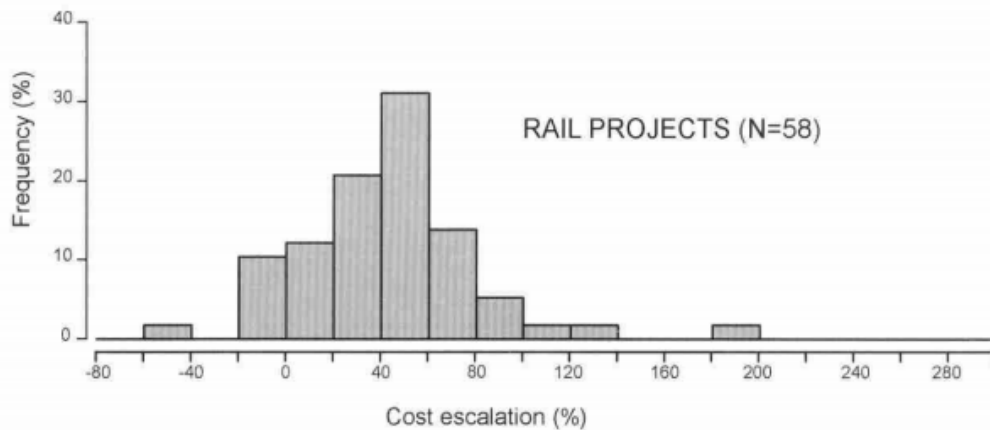| Field	XXX										
PDO approved	2007										
Discounting factor	10 %										
Years after PDO		0	1	2	3	4	5	6	7	8	9	10
Mean estimated production		0	0	0,26	0,21	0,14	0,09	0,07	0,05	0,04	0,03	0,03
Actual production		0	0	0,26	0,24	0,32	0,24	0,14	0,18	0,15	0,15	0,18
Oil price [USD/bbl]		73	97	62	79	111	112	109	99	52	44	54
Exhange rate [NOK/USD]		5,9	5,6	6,3	6,0	5,6	5,8	5,9	6,3	8,1	8,4	8,3
Conversion factor [bbl/Sm3]		6,3	6,3	6,3	6,3	6,3	6,3	6,3	6,3	6,3	6,3	6,3
Annual economic value of underproduction [mill NOK]			-4	-96	-714	-609	-271	-525	-300	-266	-425
Discounted value [mill 2007-NOK]		0	0	-3	-72	-488	-378	-153	-270	-140	-113	-164
Total revenue lost [mill 2007-NOK]	**-1 781**										
Inflation adjustment	1,24										
Total revenue lost [mill 2017-NOK]	**-2 215**										

When neglecting lost value due to delayed production start, the total accrued lost value due to underproduction was estimated to almost 19 billion NOK, expressed in 2017 PV. However, this calculation did not consider that many of the fields saw further investments after production start; investments that was not factored in at the time of the PDO. When correcting for these investments, Mohus ended up with a total loss in PV of 200 billion NOK due to underproduction.

The expected NPV loss caused by overconfidence in the petroleum industry is also documented in Welsh et al. [45], which is summarized in Figure 6.7. Given no overconfidence, the expected NPV was USD 246 million. Given 5 percent overconfidence, this number would have been reduced to USD 224 million - a loss of nearly 10 percent. Furthermore, the decrease in NPV is accelerating for every incremental increase in the overconfidence level. At 30 percent overconfidence, the expected NPV would have decreased to USD -10 million.
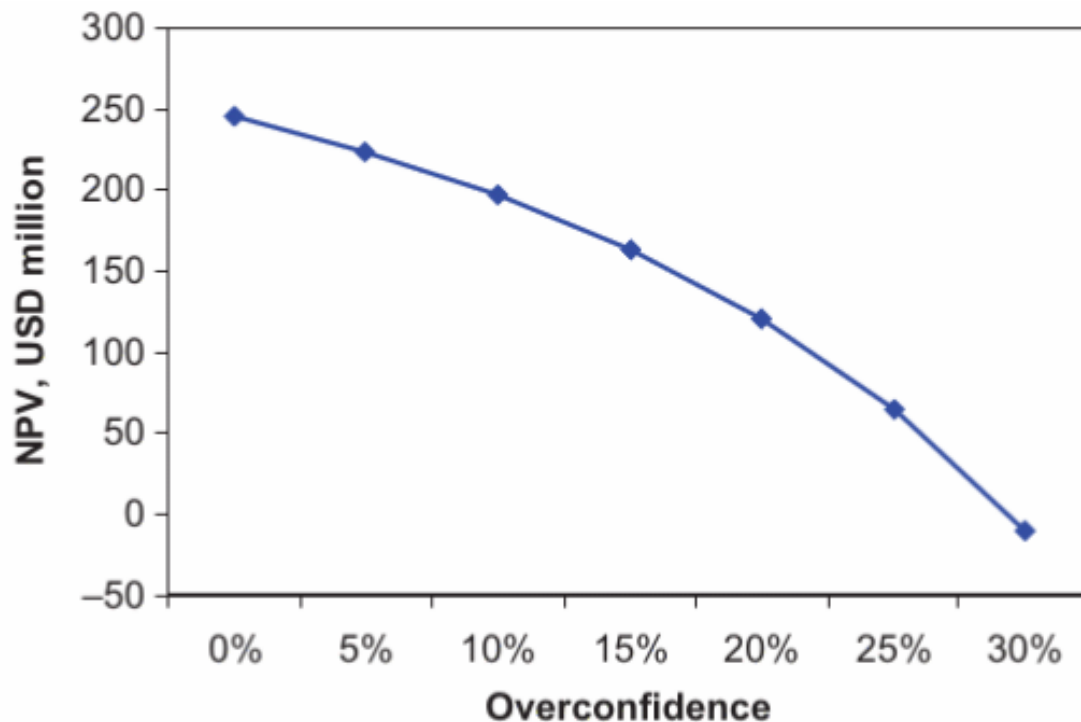
Figure 6.7: NPV loss due to overconfidence [2] [45]

6.3   Does the industry want to improve?

The evidence provided in Nandurdikar and Wallace [33] indicate the petroleum industry's unwillingness to improve future forecasts. Their research showed no improvements in forecasting between 1995 and 2011. This was further analysed and supported in this thesis. This is alarming given all the technological advances, as well as the increase in the number of publications (illustrated Figure 6.8) in the same time span.

Figure 6.8: Growth in the number of papers on probabilistic production forecasting [3]

Nandurdikar and Wallace point out that in only 30 percent of projects were there conducted an analysis to figure out the reasons behind the errors in their forecasts. This is startling given the improvements such analyses could induce.

Given how grave the economic consequenses of biased estimates can be, it is peculiar that the industry is not more eager to improve their forecasts and red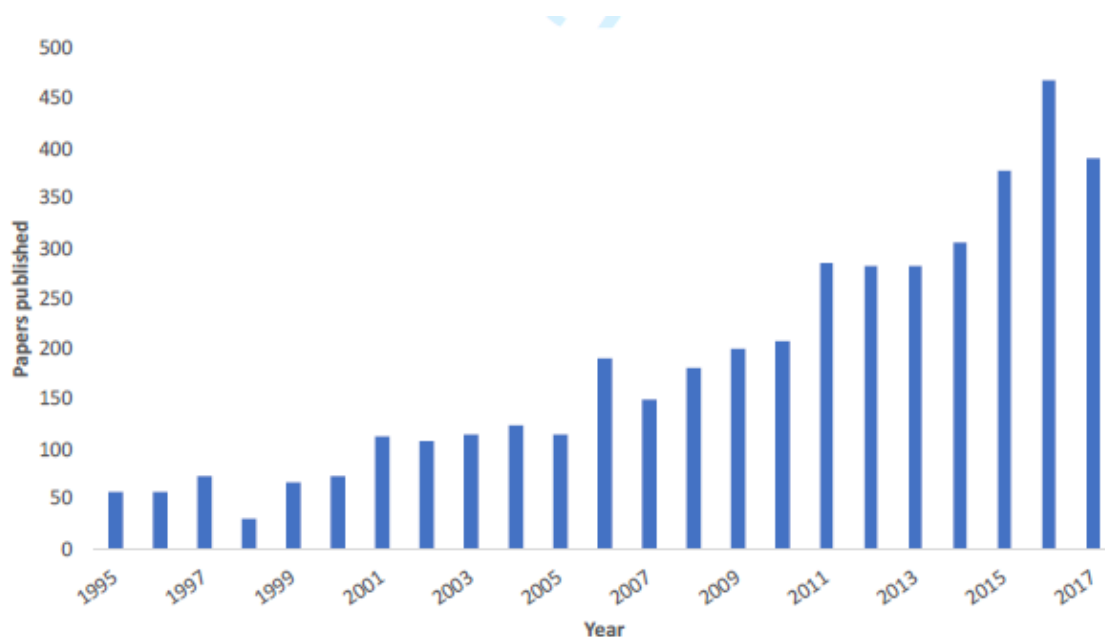uce the bias in their predictions. Nandurdikar and Wallace argues that the lack of accountability makes it difficult to achieve any progress. Since no one really bears the responsibility for the poor assessments, no one will be told that they will have to sit down and evaluate their own mistakes in order to improve in the future.

Lastly, it is reasonable to ask whether the industry is *really* aware of the magnitude of bias in their predictions and the impact it has on companies' bottom line. Bratvold and Begg [2] explain how companies in the oil- and gas industry rarely distinguish between good *decisions* and good *outcomes*. Generally, if a good outcome follows a decision, the decision is considered good and vice versa. As explained in Section 2.5.3, this strictly contradicts the theory. A plausible explanation for the inability to distinguish between outcomes and decisions is that the outcomes in this industry are generally not *bad*. The companies make a great deal of money because of the high oil prices, and are potentially not fully aware of how much value that is vasted. However, the oil- and gas industry is not entitled for future life, and the increased demand for more renewable energy will put the petroleum industry to the test and might induce a reality check for many of the companies in the market.

6.4 Discussion on RCF

There seem to be no improvements from using the modified reference class forecasting rather than the classical. On the contrary, the classical method seems to be better at calibrating the data in this thesis. This is a surprising outcome because the modified reference class forecasting uses more of the the information provided by the assessors. What this implies is that it is beneficial to abandon the forecasters' P10- and P90 estimates entirely and rather just create new ones using the reference class forecasting model. This could be explained if the amount of bias in the P10- and P90 estimates was not homogeneous, which would violate one of the general assumption for the procedure. Another explanation is that the modified reference class forecasting method needs refinement in order to better calibrate the estimates. Both of these suggestions will be discussed in the following.

6.4.1 Should the P10- and P90 estimates be abandoned?

There are several reasons why it might be better to abandon the P10- and P90 estimates altogether:

1. It was argued in Section 6.1.4 that most people are optimistic most of the time. This implies that the assessors are likely to make wrongful forecasts by assessing too high production volumes, which was also evident from the data. The data set contained forecasts from several different companies, and presumably from an even greater number of assessors. It must further be assumed that these assessors have a varying degree of inherent optimism. It is therefore conceivable that the MRCF method cannot appropriately calibrate the data to account for this. Since the CRCF method discards the P10- and P90 estimates entirely, this inconsistency is not an issue.

2. It was also argued that the anchoring and adjusting heuristic could impact the quality of assessments. If it is assumed that it is common for companies to develop a base case and then adjust the P10- and P90 values accordingly, it would be difficult for the MRCF method to correct for this. The P10 and P90 would neither be random nor based on probabilistic forecasting. Since every assessor in every company might adjust the P10s and P90s in a varying degree, the general assumption that the bias in the reference class and application class is equal might be violated.

3. Variation in optimism might also be seen even when the assessors remain the same. For example, the availability heuristic makes it more likely for assessors to be biased towards optimism and overconfidence if previous assessments have been good. This further complicates the matter.

6.4.2 Does the modified RCF method need refinement?

It became evident from the results acquired with the MRCF method that, on average, it was highly efficient at calibrating the P10 estimates, but not as efficient at calibrating the P90 estimates. The question becomes: does the method need refinement in order to improve the calibration? There are many ways that the method may be improved. Mostly, it comes down to the amount and quality of the information used as input.

For the purpose of this thesis, the mean estimate has been assumed equal to the P50. This is a simplification which generally does not hold true. In statistical analysis, P50 is a much more useful measurement than the mean. A reccomendation is therefore that the NPD in the future requests companies to report the middle value in terms of P50, rather than the mean. With the MRCF method, the middle value also has an impact on the P10- and P90 values during the calibration procedure and, thus, a more precise middle value could prove valuable.

Since the P10- and P90 values are reliant on the middle value, the MRCF method often exaggerate the adjustments to the P90 values - which were generally well estimated by the assessors (see Figure 5.5 and 5.6 and Table 5.10 through 5.13). Thus, it is possible that some alterations to the model might improve the P90 estimates. Suggestions include:

1. Using the metalog distribution in the MRCF method
2. Create a hybrid model between the CRCF method and the MRCF method such that the P90 calibration can be improved
3. Only partly calibrate the P90 values, ending up with a value that is somewhere between the original estimate and the result of the calibration procedure

Nonetheless, the fact that the classical reference class forecasting model outperforms the modified is a surprising outcome. Presumably, companies *do* put time and money into forecasting P10 and P90 and, thus, it should have been improving the accuracy of the calibration.

6.5 Further research

In this thesis, the modified reference class forecasting method has been implemented to debias production forecasts of oil on the NCS. The data on which the methodology is applied all stem from the NCS, which implies limited geological diversity and a assessments from a limited number of companies. Clearly, no general conclusions can be drawn based on this NCS study. However, there is no reason to believe that PDO based forecasts provided by operators on the NCS differ significantly from forecasts provided in other regions for conventional field development are any better. However, we hope that this study inspires similar studies outside

the NCS.

Some examples for further research include:

1. Base the reference class on a single company or a few selected companies. This could be helpful to more specifically address the bias within a company and might yield even better results

2. Base the reference class on specific geological characteristics or other specific characteristics which could have an impact on the degree of bias in the original model

3. In this thesis, RCF has been applied on the finalized forecast in the PDO that the company provided the NPD with. It would be interesting to see whether RCF could have a larger positive effect if applied earlier in the forecasting process - either to debias the inputs that go into the models that the companies use, or to debias the direct output of the models. A possibility would also be to apply RCF in multiple steps

As discussed, adjustments might be made to the modified reference class forecasting method in order to improve its efficiency.

Chapter 7

Conclusion

In this thesis, the modified reference class forecasting approach has been applied to production data from the NCS. The goal was to reduce forecasting biases. In previous research, the classical reference class model has been applied on similar data with great success. The goal of this thesis was therefore to see whether the modified model could improve on the already efficient classical model.

Production data from 1997-2019 were used for forecast validation and calibration. Due to inconsistent probabilistic assessments, some of the data points could not be used, leaving 51 fields and 1320 consistent forecasts. The data was evaluated on a year-by-year basis. On average, the 80 percent interval that the assessors had forecasted captured about 40 percent of the true values reported. Moreover, 41 percent of the true values fell below the assessed P10 estimates. These results confirms overprecision and optimism in the forecasts.

The reason behind the bias is multifaceted. Although bad luck is often used as an explanation for suboptimal forecasting performance, two alternative causes, delusion and deception, have been introduced and discussed and deemed more likely. A number of cognitive biases, including optimism, poor self-knowledge, and the anchoring and adjusting heuristic all contribute to delusional forecasts. The principal-agent problem causes deceptional forecasts. Previous studies have suggested that better learning environments, and increased accountability and transparency could help reducing these biases.

In this thesis, the fundamental research by Kahneman and Tversky on the outside view and reference class forecasting was adapted in an attempt to debias the production forecasts used for the development and operation decision (PDO). The modified reference class forecasting approach resulted in RMSE reductions of 55-, 52-, 34- and 28 percent for Year 0, 1, 2 and 3, respectively. These are significant improvements. However, the classical reference class forecasting approach, where

the forecasted P10 and P90 are ignored, perform marginally better, with RMSE improvements of 57-, 54-, 36- and 35 percent.

The fact that the CRCF approach perform better than the MRCF approach implies that the P10- and P90 estimates should be ignored in favour of simply using the forecasted middle value and create a complete forecasting distribution from this single value. It is not unreasonable to assume that the companies put a great deal of time, effort and money into forecasting P10 and P90 and, thus, the fact that it does not improve the calibration procedure whatsoever should warrant serious evaluation of how the operators generate their probabilistic forecasts.

The research provided in this thesis leaves room for further research. Rather than request mean forecasts, the NPD should ask operators to provide median values. This would significantly ease and improve both forecast verification and calibration. With this information, the MRCF approach could be re-evaluated. Moreover, it would be useful to attempt to apply the presented methodology in other countries and environments to study the effect it might have on an expanded set of data - or a different data set entirely.

Bibliography

[1]   Richard B. Alley, Kerry A. Emmanuel, and Fuqing Zhang. "Advances in weather prediction". In: (2019): *Science*. DOI: `doi.org/10.1126/science.aav7274`.

[2]   Reidar B. Bratvold and Steve Begg. *Making Good Decisions*. Society of Petroleum Engineers, 2010.

[3]   Reidar B. Bratvold et al. "Production Forecasting: Optimistic and Overconfident - Over and Over Again". In: (2020).

[4]   Eric Brun. *What is "Fuzziness" - or "the Unknown" -at the Front End of New Product Development Projects?* Working Paper. University of Stavanger, 2011.

[5]   Robert T. Clemen et al. "Modeling and correction of over-precision bias in subjective confidence intervals: The Bi-Normal and the Multi Parameter Bi-Normal approaches". In: *Management Science* (2020).

[6]   C. Cunnane. "Unbiased plotting positions - A review". In: *Journal of Hydrology* 37(3) (1978). DOI: `https://doi.org/10.1016/0022-1694(78)90017-3`.

[7]   Richard L. Daft and Robert H. Lengel. "Organizational Information Requirements, Media Richness and Structural Design". In: *Management Science* (1986). DOI: `https://doi.org/10.1287/mnsc.32.5.554`.

[8]   Norwegian Petroleum Directorate. *Act 29 November 1996 No. 72 relating to petroleum activities*. 2019. URL: `https://www.npd.no/en/regulations/acts/act-29-november-1996-no2.-72-relating-to-petroleum-activities/`. (Accessed: March 2, 2020).

[9]   Norwegian Petroleum Directorate. *Faktasider*. (Accessed: July 14, 2020).

[10]   Norwegian Petroleum Directorate. *Fields*. URL: `https://www.norskpetroleum.no/en/facts/field/`. (Accessed: April 6, 2020).

[11]   Bent Flybjerg. "Curbing Optimism Bias and Strategic Misrepresentation in Planning: Reference Class Forecasting in Practice". In: 16 (2007): *European Planning Studies*.

[12]   Bent Flybjerg and COWI. "Procedures for Dealing with Optimism Bias in Transport Planning". In: (2004).

[13]   Bent Flybjerg, Massimo Garbuio, and Dan Lovallo. "Delusion and Deception in Large Infrastructure Projects: Two Models for Explaining and Preventing Executive Disaster". In: 51 (2009): *California Management Review.*

[14]   Bent Flyvbjerg. "Cost Overruns and Demand Shortfalls in Urban Rail and Other Infrastructure". In: 30 (2007): *Transportation Planning and Technology.* DOI: `doi.org/10.1080/03081060701207938`.

[15]   Bent Flyvbjerg. "Over Budget, Over Time, Over and Over Again: Managing Major Projects". In: (2011): *The Oxford Handbook of Project Management.*

[16]   Bent Flyvbjerg, Mette Skamris Holm, and Søren Buhl. In: 68 (2002): *Journal of the American Planning Association.* DOI: `doi.org/10.1080/01944360208976273`.

[17]   John Kenneth Galbraith. *The Age of Uncertainty.* 1977.

[18]   Daniel Kahneman and Dan Lovallo. "Delusions of Success: How Optimism Undermines Executives' Decisions". In: (2003): *Harward Business Review.*

[19]   Daniel Kahneman and Amos Tversky. *Judgment under uncertainty: Heuristics and biases.* 1974. Accessed through the 1982 book publication under the same name.

[20]   Daniel Kahneman and Amos Tversky. *Intuitive Predictions: Biases and Corrective Procedures.* 1977.

[21]   Tom W. Keelin. "The Metalog Distributions". In: *Decision Analysis* 13 (2016). DOI: `https://doi.org/10.1287/deca.2016.0338`.

[22]   MathIsFun. *Normal Distribution.* 2019. URL: `https://www.mathsisfun.com/data/standard-normal-distribution.html`. (Accessed: June 5, 2020.

[23]   Ed Merrow. *Industrial Megaprojects: Concepts, Strategies and Practices for Success.* 2019.

[24]   Ministry of Petroleum and Energy and Norwegian Petroleum Directorate. *Activity per sea area.* 2020. URL: `https://www.norskpetroleum.no/en/developments-and-operations/activity-per-sea-area/`. (Accessed: February 28, 2020).

[25]   Ministry of Petroleum and Energy and Norwegian Petroleum Directorate. *Licenses.* 2020. URL: `https://www.norskpetroleum.no/en/facts/licences/`. (Accessed: February 28, 2020).

[26]   Ministry of Petroleum and Energy and Norwegian Petroleum Directorate. *The Government's Revenues.* 2020. URL: `https://www.norskpetroleum.no/en/economy/governments-revenues/`. (Accessed: July 11, 2020).

[27]   Ministry of Petroleum and Energy and Norwegian Petroleum Directorate. *The Government's Revenues.* 2020. URL: `https://www.norskpetroleum.no/en/economy/petroleum-tax/`. (Accessed: July 11, 2020).

[28]   Ministry of Petroleum and Energy and Norwegian Petroleum Directorate. *Classification of petroleum resources.* URL: `https://www.norskpetroleum.no/en/petroleum-resources/resource-classification/`. (Updated: March 21, 2020).

[29] Ministry of Petroleum and Energy and Norwegian Petroleum Directorate. *Original reserves*. URL: `https://www.norskpetroleum.no/en/facts/original-reserves/`. Updated: January 29, 2020.

[30] Ministry of Petroleum and Energy and Norwegian Petroleum Directorate. *Remaining reserves*. URL: `https://www.norskpetroleum.no/en/facts/remaining-reserves/`. (Updated: February 21, 2020).

[31] Ministry of Petroleum and Energy and Ministry of Labour and Social Affairs. *Guidelines for plan for development and operation of a petroleum deposit (PDO) and plan for installation and operation of facilities for transport and utilisation of petroleum (PIO)*. 2018.

[32] Erlend Mohus. "Over Budget, Over Time, and Reduced Revenue, Over and Over Again – An Analysis of the Norwegian Petroleum Industry's Inability to Forecast Production". In: (2018). Ed. by Reidar B. Bratvold.

[33] N. S. Nandurdikar and L. Wallace. *Failure to Produce: An Investigation of Deficiencies in Production Attainment*. Society of Petroleum Engineers, 2011. DOI: `doi:10.2118/145437-MS`.

[34] Norwegian Petroleum Directorate. *Guidelines to plan for development and operation of a petroleum deposit (PDO)*. 2000. DOI: `https://cutt.ly/GpGXkCz`.

[35] Oil and Gas Portal. *Upstream — Petroleum Exploration*. URL: `http://www.oil-gasportal.com/upstream/petroleum-exploration/`. (Accessed: July 7, 2020.

[36] Oil and Gas Portal. *Upstream — Petroleum Production Phase*. URL: `http://www.oil-gasportal.com/upstream/petroleum-production-phase/`. (Accessed: March 10, 2020).

[37] Regjeringen.no. *Kontinentalsokkelen: Spørsmål og svar*. 2014. URL: `https://www.regjeringen.no/no/tema/utenrikssaker/folkerett/kontinentalsokkelen-sporsmal-og-svar/id448309/`. (Accessed: February 27, 2020).

[38] Kenneth C. Lichtendahl Jr. Robert T. Clemen. "Debiasing Expert Overconfidence: A Bayesian Calibration Model". In: *Probabilistic Safety Assessment and Management 6* (2002).

[39] Michael Schrage. "Daniel Kahneman: The Thought Leader Interview". In: (33 2003): *strategy+business*.

[40] Nassim N. Taleb. *The Black Swan: The Impact of the Highly Improbable*. 2007.

[41] Lars Taraldsen. *Norske oljeprosjekter har sprukket med over 200 milliarder på 14 år*. 2015. URL: `https://www.tu.no/artikler/norske-oljeprosjekter-har-sprukket-med-over-200-milliarder-pa-14-ar/275922`. (Accessed: July 11, 2020).

[42] Offshore Technology. *The history of the oil and gas industry from 347 AD to today*. URL: `https://www.offshore-technology.com/comment/history-oil-gas/`. (Accessed: April 7, 2020).

[43] The Ministry of Petroleum and Energy and the Ministry of Labour and Social Affairs. *Guidelines for plan for development and operation of a petroleum deposit (PDO) and plan for installation and operation of facilities for transport and utilisation of petroleum (PIO)*. 2018. URL: `https://www.npd.no/globalassets/1-npd/regelverk/forskrifter/en/pdo-and-pio.pdf`.

[44] Duke University. URL: `https://www.fuqua.duke.edu/faculty/robert-clemen`. (Accessed: April 16, 2020).

[45] Matthew B. Welsh, Reidar B. Bratvold, and Steve H. Begg. "Modeling the Economic Impact of Cognitive Biases on Oil and Gas Decisions". In: (2007).

Appendix A

Assumption: median = mean

Since the underlying distributions of the forecasts are unknown, their skewness cannot be evaluated directly. Therefore, to evaluate the decision to assume that the median is equal to the mean, the metalog distribution is introduced. This is a versatile distribution, proposed by Keelin [21], which can handle percentiles as input. Based on these inputs, a complete distribution is generated.

Figure A.1 illustrates how data is used as input to the distribution. Since the distribution uses percentiles as input, the mean estimate cannot be used as the middle value. However, the metalog distribution also outputs the mean of the generated distribution. The P50 value has therefore been found by changing the input value until the mean of the distribution is equal to the forecasted mean. In this example, the forecasted mean is 0.2572, whilst the calculated median is 0.2571. Thus, the distribution has close to zero skew. Although these results cannot be guaranteed for every assessed distribution, it is a general assumption for the MRCF approach presented in this thesis. If this assumption does not hold true, it could cause suboptimal results.
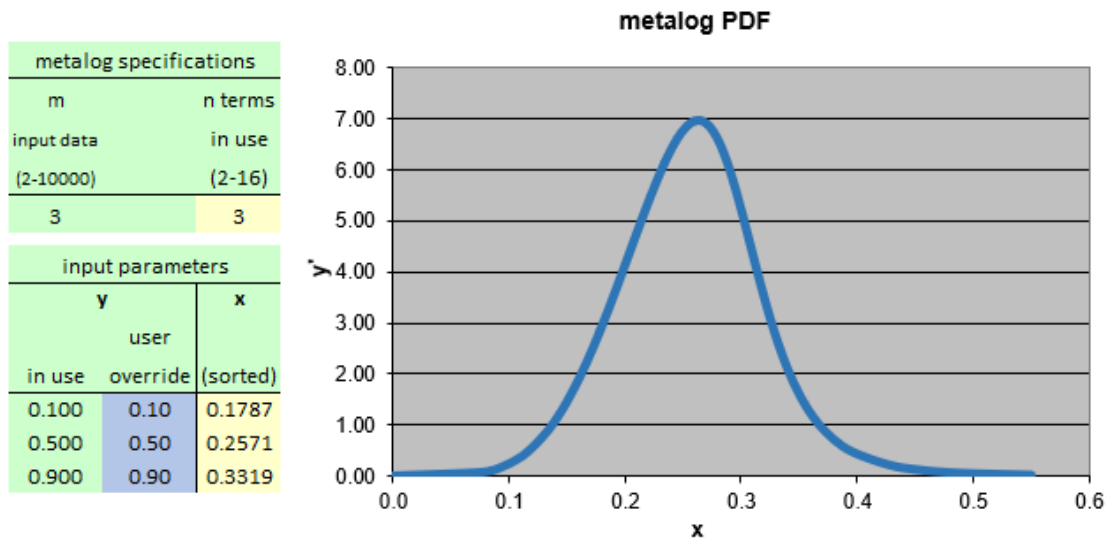
metalog specifications	
m	n terms
input data	in use
(2-10000)	(2-16)
3	3

input parameters		
y		x
	user	
in use	override	(sorted)
0.100	0.10	0.1787
0.500	0.50	0.2571
0.900	0.90	0.3319

metalog PDF

Figure A.1: Illustration of how the metalog distribution is used to fit the data
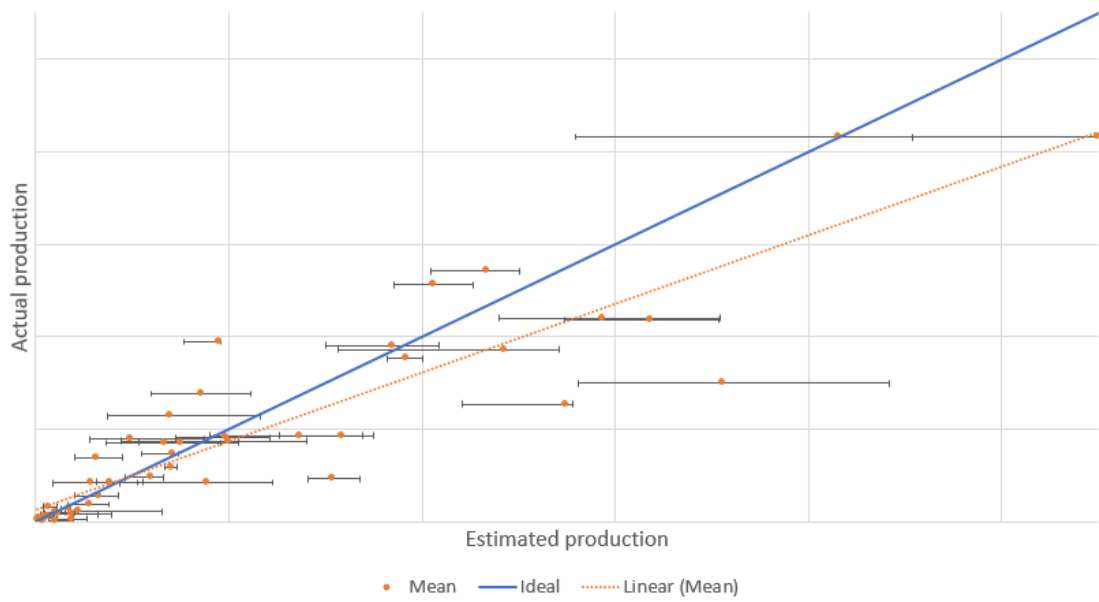
Appendix B

Figures

B.1 Confidence intervals

Figure B.1: Confidence intervals for estimates versus actual production, third year of production, all fields
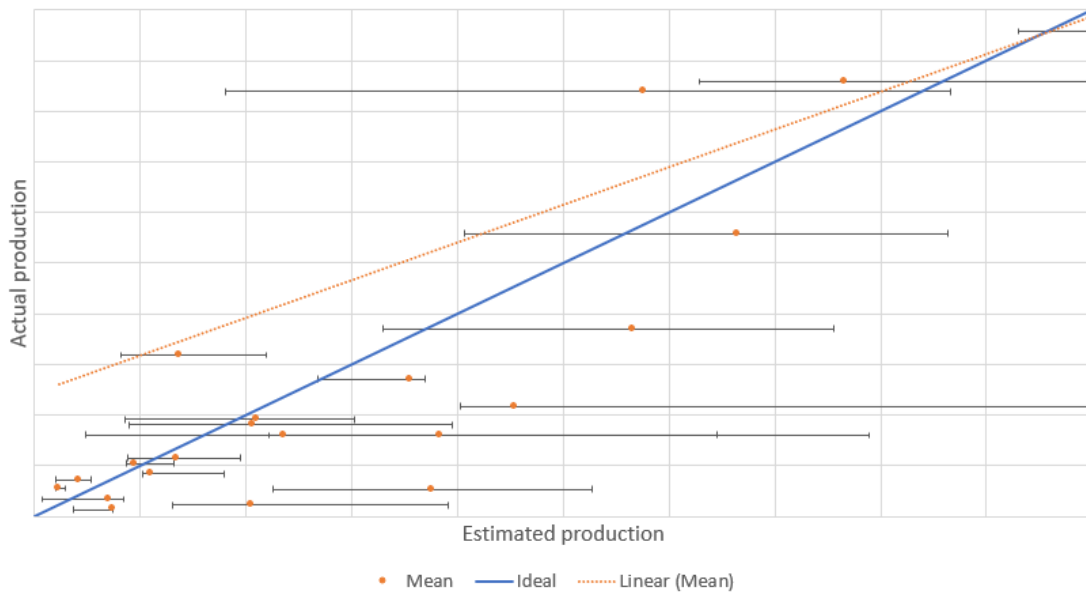
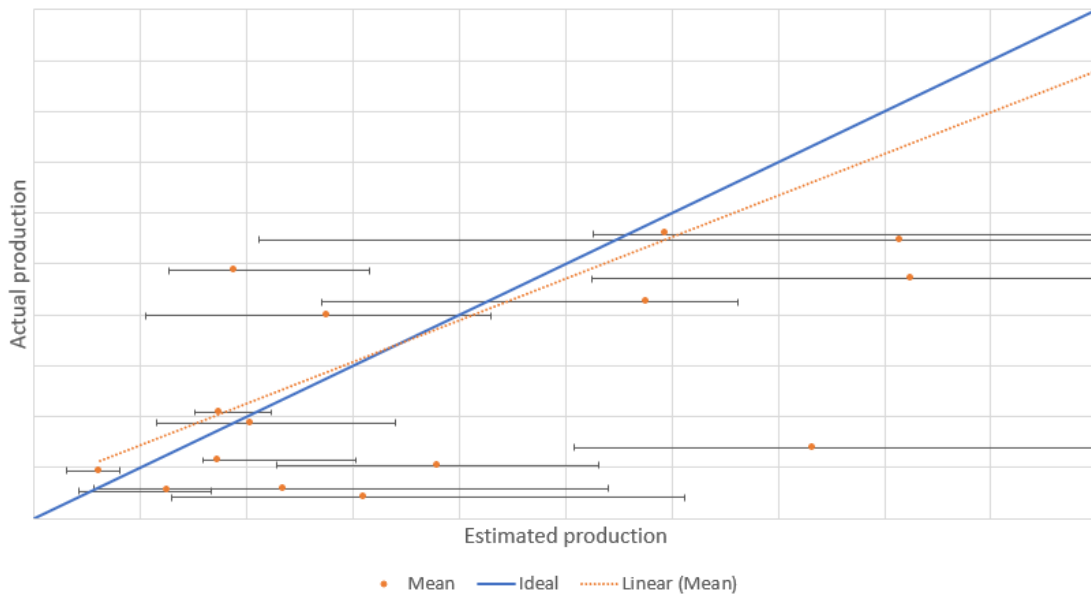**Figure B.2:** Confidence intervals for estimates versus actual production, third year of production, smaller fields
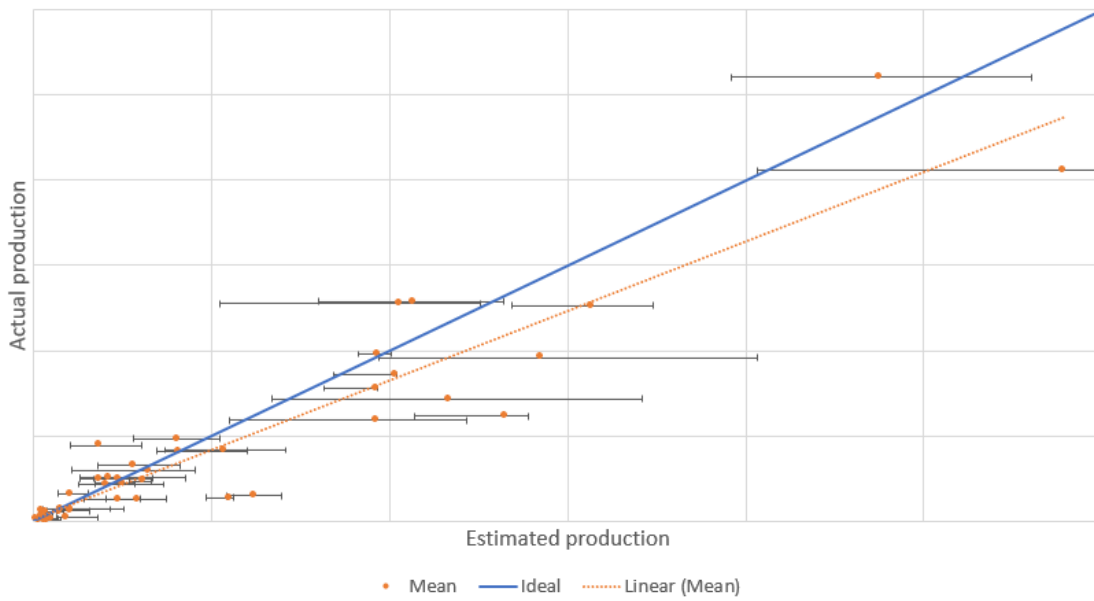
Figure B.3: Confidence intervals for estimates versus actual production, fourth year of production, all fields
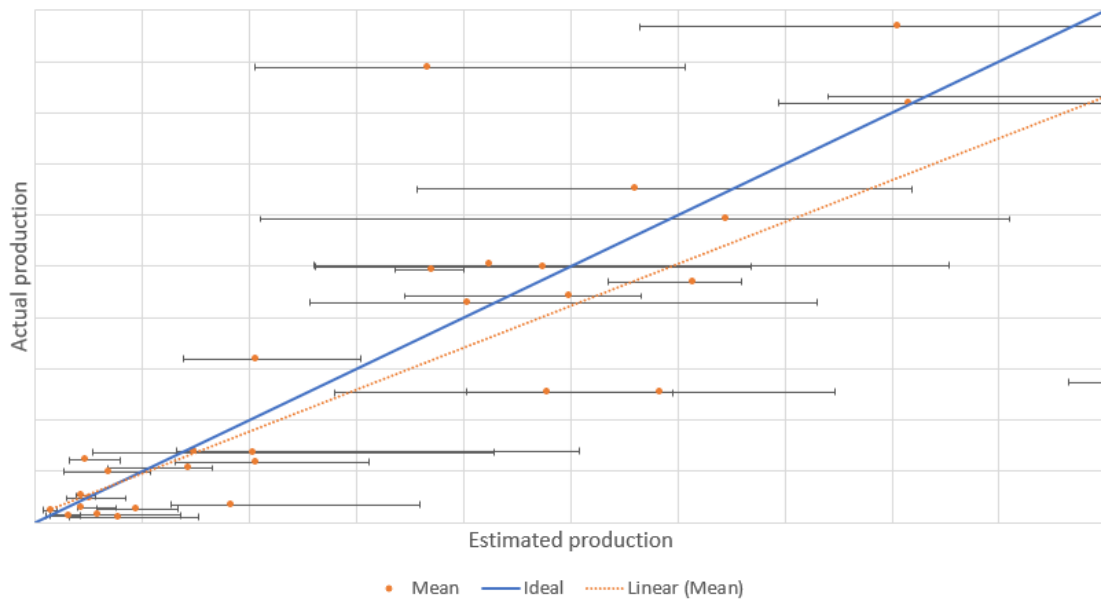
Figure B.4: Confidence intervals for estimates versus actual production, fourth year of production, smaller fields

Figure B.5: Confidence intervals for estimates versus actual production, fourth year of production, even smaller fields

B.2 Stabilization of the CMA of the calibration coefficients
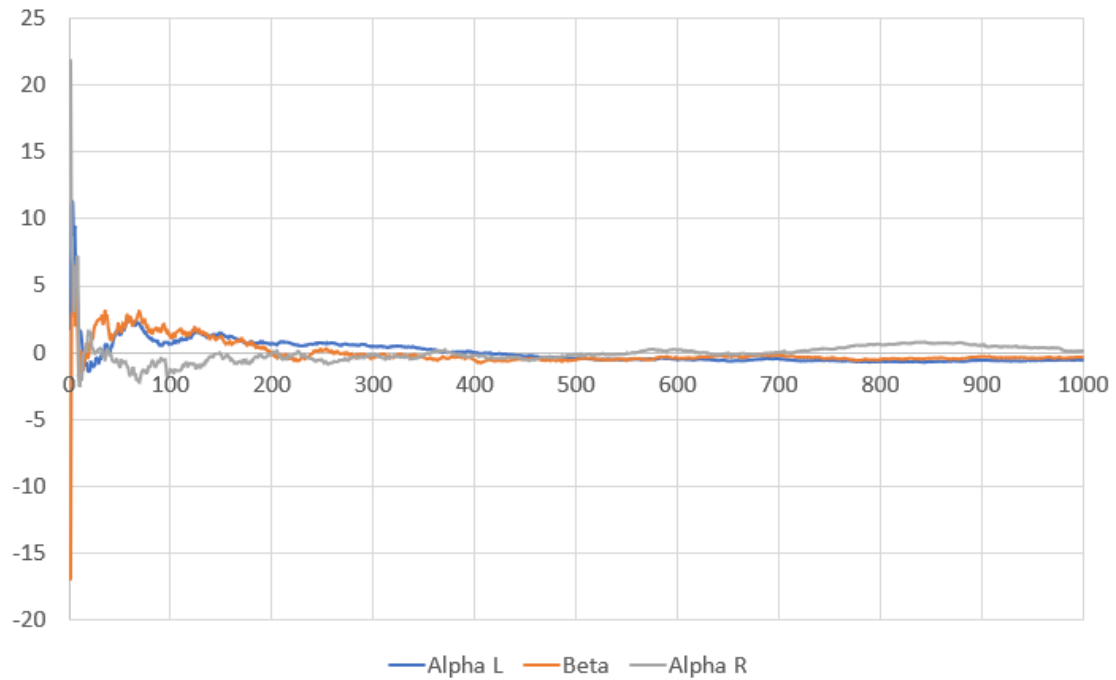
Figure B.6: Stabilization of the CMA of the calibration coefficients, Year 1
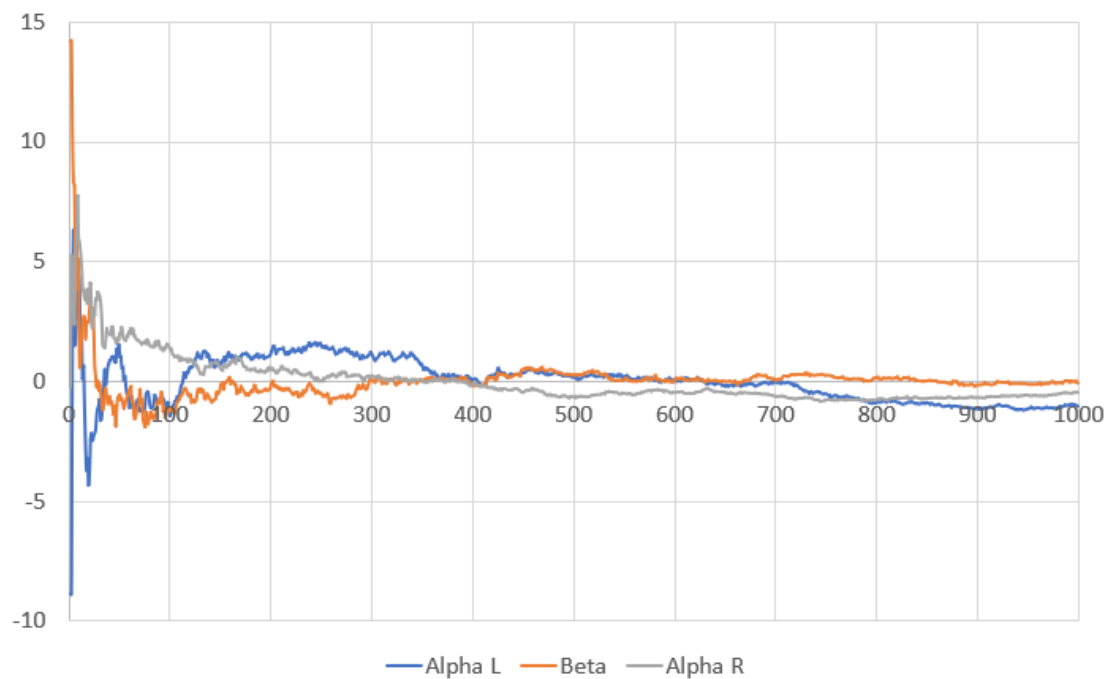

Figure B.7: Stabilization of the CMA of the calibration coefficients, Year 2
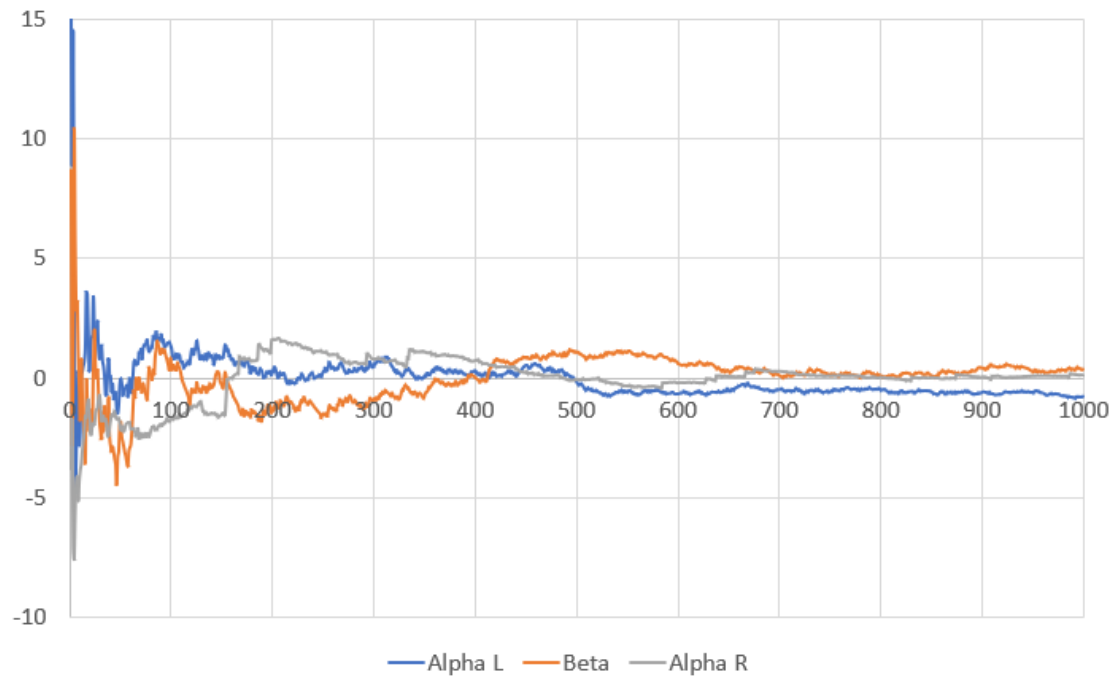
Figure B.8: Stabilization of the CMA of the calibration coefficients, Year 3

Appendix C

Excel worksheet

	A	B	C	D	E	F	G	H	I
1	Field	Year	P10	Mean	P90	Actual	Rand()		Total TRUE
2	Field 1	0	1	2	3	4	FALSE		22
3	Field 2	0	2	3	4	2	FALSE		
4	Field 3	0	3	4	5	3	FALSE		
5	Field 4	0	4	5	6	5	TRUE		Total FALSE
6	Field 5	0	5	6	7	4	TRUE		10
7	Field 6	0	6	7	8	2	TRUE		
8	Field 7	0	7	8	9	8	TRUE		
9	Field 8	0	8	9	10	5	TRUE		Total Count
10	Field 9	0	9	10	11	6	TRUE		32
11	Field 10	0	10	11	12	9	TRUE		
12	Field 11	0	11	12	13	8	FALSE		
13	Field 12	0	12	13	14	9	TRUE		
14	Field 13	0	13	14	15	15	FALSE		
15	Field 14	0	14	15	16	19	TRUE		
16	Field 15	0	15	16	17	23	FALSE		
17	Field 16	0	16	17	18	15	TRUE		
18	Field 17	0	17	18	19	16	TRUE		
19	Field 18	0	18	19	20	20	TRUE		
20	Field 19	0	19	20	21	14	FALSE		
21	Field 20	0	20	21	22	21	TRUE		
22	Field 21	0	21	22	23	19	TRUE		
23	Field 22	0	22	23	24	25	TRUE		
24	Field 23	0	23	24	25	20	FALSE		
25	Field 24	0	24	25	26	18	TRUE		
26	Field 25	0	25	26	27	28	TRUE		
27	Field 26	0	26	27	28	24	TRUE		
28	Field 27	0	27	28	29	27	TRUE		
29	Field 28	0	28	29	30	25	TRUE		
30	Field 29	0	29	30	31	30	FALSE		
31	Field 30	0	30	31	32	34	FALSE		
32	Field 31	0	31	32	33	26	TRUE		
33	Field 32	0	32	33	34	35	TRUE		

Figure C.1: Excel worksheet - Input
The data points are random, assigned only for the purpose of this example

	K	L	M	N
	Variable #	**Plotting probability**		**From Plotting Probability**
	1	0.027027027		
	2	0.072072072		**Lower tail**
	3	0.117117117		2
	4	0.162162162		
	5	0.207207207		
	6	0.252252252		**Upper tail**
	7	0.297297297		2
	8	0.342342342		
	9	0.387387387		
	10	0.432432432		**Above P50**
	11	0.477477477		11
	12	0.522522523		
	13	0.567567568		
	14	0.612612613		**Below P50**
	15	0.657657658		11
	16	0.702702703		
	17	0.747747748		
	18	0.792792793		
	19	0.837837838		
	20	0.882882883		
	21	0.927927928		
	22	0.972972973		

Figure C.2: Excel worksheet - Perfectly calibrated judge, in sample

Figure C.3: Excel worksheet - in sample data calibration

Figure C.4: Excel worksheet - out of sample data calibration

BS	BT	BU	BV
Variable #	**Plotting probability**		**From Plotting Probability**
1	0.059		
2	0.157		**Lower tail**
3	0.255		1
4	0.353		
5	0.451		
6	0.549		**Upper tail**
7	0.647		1
8	0.745		
9	0.843		
10	0.941		**Above P50**
			5
			Below P50
			5

Figure C.5: Excel worksheet - perfectly calibrated judge, out of sample

Appendix D

Excel VBA computer code

```vba
Sub out_of_sample()

' Timer
Dim StartTime As Double
Dim MinutesElapsed As String

'Remember time when macro starts
    StartTime = Timer




' Clear Contents in the necessary places
Range("K2:L500").ClearContents
Range("Q2:Y500").ClearContents
Range("AD2:AL500").ClearContents
Range("AS2:BA500").ClearContents
Range("BF2:BN500").ClearContents
Range("BS2:BT500").ClearContents
Range("CF2:CI2501").ClearContents
Range("G2:G500").ClearContents
Range("CL2:CQ2501").ClearContents
Range("CD3:CE2501").ClearContents
Range("CV2:AQQ100").ClearContents



Application.ScreenUpdating = False



Dim i As Integer
Dim j As Integer
Dim NoTrue As Integer
Dim randfun As Variant
Dim lastRow As Variant

' Define the last row of data input; this is the amount of datapoints + 1!
lastRow = Range("A" & Rows.Count).End(xlUp).Row


Dim NoValues As Integer
Dim NoHalfValue As Integer
Dim nthSmall As Double

' Count number of datapoints
Range("I10").Value = "=Count(R2C6:R" & lastRow & "C6)"
```

```vba
NoValues = Range("I10").Value

' This is half of the values (J/2 if even, (J-1)/2 if odd). This is the "False"
    values
NoHalfValues = NoValues / 2

Dim sizeoos As Integer

sizeoos = NoValues * 0.2


' Dim flags
Dim flg As Boolean
Dim flgl As Boolean
Dim flgr As Boolean


' Final calculation assignment
Dim uLT As Double
Dim uUT As Double
Dim uAP50 As Double
Dim uBP50 As Double

Dim CLT As Double
Dim cUT As Double
Dim cAP50 As Double
Dim cBP50 As Double

Dim ppLT As Double
Dim ppUT As Double
Dim ppAP50 As Double
Dim ppBP50 As Double


Dim tempuLT As Double
Dim tempuUT As Double
Dim tempuAP50 As Double
Dim tempuBP50 As Double

Dim tempcLT As Double
Dim tempcUT As Double
Dim tempcAP50 As Double
Dim tempcBP50 As Double

Dim tempppLT As Double
Dim tempppUT As Double
Dim tempppAP50 As Double
Dim tempppBP50 As Double

Dim Count As Integer
Dim lastRowFalse As Integer

Dim tempbeta As Double
Dim tempalphaL As Double
Dim tempalphaR As Double

Dim beta As Double
Dim alphaL As Double
Dim alphaR As Double

Dim tempcLTabs As Double
Dim cLTabs As Double
Dim tempcUTabs As Double
Dim cUTabs As Double
Dim tempaP50abs As Double
Dim aP50abs As Double
Dim tempbP50abs As Double
Dim bP50abs As Double

Dim tempuLTabs As Double
```

```vba
Dim uLTabs As Double
Dim tempuUTabs As Double
Dim uUTabs As Double
Dim tempuaP50abs As Double
Dim uaP50abs As Double
Dim tempubP50abs As Double
Dim ubP50abs As Double


,,,,,
' SLETT?!

' Dim rmseu As Double
' Dim temprmseu As Double
' Dim rmsec As Double
' Dim temprmsec As Double

' Dim rmseu1 As Double
' Dim temprmseu1 As Double
' Dim rmsec1 As Double
' Dim temprmsec1 As Double
,,,,,


' Start iterations (100? 1000?)
Dim looping As Integer

For looping = 1 To 100

' Reset calibration parameters
Range("AN4").Value = 1
Range("AN8").Value = 1
Range("AN12").Value = 1

' Clear Contents in the necessary places
Range("K2:L500").ClearContents
Range("Q2:Y500").ClearContents
Range("AD2:AL500").ClearContents
Range("AS2:BA500").ClearContents
Range("BF2:BN500").ClearContents
Range("BS2:BT500").ClearContents

' Use rand() function to generate a random number for each field
Range("G2").Formula = "=rand()"
Range("G2:G" & lastRow).FillDown
Range("G2:G" & lastRow).Value = Range("G2:G" & lastRow).Value


' The n-th smallest value in the range. The 2nd argument decides how many
    datapoints will generate the calibration parameters
nthSmall = Application.WorksheetFunction.Small(Range("G2:G" & lastRow & ""), 6)

i = 2

' Set "True" or "False" for each of the random numbers
Do While Cells(i, 7) <> ""

    If Cells(i, 7) <= nthSmall Then

        Cells(i, 7).Value = "False"

    Else

        Cells(i, 7).Value = "True"

    End If

    i = i + 1

Loop
```

```vba
' Count TRUE
Range("I2").Formula = "=CountIf(R2C7:R" & lastRow & "C7, TRUE)"
NoTrue = Range("I2").Value

' Count FALSE
Range("I6").Formula = "=CountIf(R2C7:R" & lastRow & "C7, FALSE)"
NoFalse = Range("I6").Value



i = 2
j = 1

' Variable # (K:K)
Do While Cells(i, 7) <> ""

    If Cells(i, 7) = "True" Then

        Cells(j + 1, 11).Value = j

        j = j + 1

    End If

    i = i + 1

Loop



' Reset i
i = 2

' Plotting Probability (L:L)
Do Until i = (j + 1)

    Cells(i, 12).Value = (Cells(i, 11) - 0.4) / (1 + (j - 1) - 2 * 0.4)

    i = i + 1

Loop


' Data in tails (plotting probability) (N:N)

' Count lower tail
Range("N4").Value = Application.CountIf(Range("L2:L" & (NoTrue + 1) & ""), "<0.1
    ")

' Count upper tail
Range("N8").Value = Application.CountIf(Range("L2:L" & (NoTrue + 1) & ""), ">0.9
    ")

' Count true values above P50
Range("N12").Value = Application.CountIf(Range("L2:L" & (NoTrue + 1) & ""), "
    >0.5")

' Count true values below P50
Range("N16").Value = Application.CountIf(Range("L2:L" & (NoTrue + 1) & ""), "
    <0.5")



' Reset i, j
i = 2
```

```vba
j = 2

' Input the datasets of the fields where the rand() function is TRUE in another
    table (Q:V)
Do While Cells(i, 7) <> ""

    If Cells(i, 7) = "True" Then

        Cells(j, 17).Value = Cells(i, 1).Value
        Cells(j, 18).Value = Cells(i, 2).Value
        Cells(j, 19).Value = Cells(i, 3).Value
        Cells(j, 20).Value = Cells(i, 4).Value
        Cells(j, 21).Value = Cells(i, 5).Value
        Cells(j, 22).Value = Cells(i, 6).Value

        j = j + 1

    End If

    i = i + 1

Loop



' Reset i, j
i = 2
j = 2

' Calculate standard deviation and the cumulative probability (W:Y)
Do While Cells(i, 17) <> ""

    Cells(i, 23).Value = (Cells(i, 20).Value - Cells(i, 19).Value) / 1.28
    Cells(i, 24).Value = (Cells(i, 21).Value - Cells(i, 20).Value) / 1.28

    If Cells(i, 22).Value < Cells(i, 20) Then

        Cells(i, 25).Value = WorksheetFunction.Norm_Dist(Cells(i, 22).Value,
            Cells(i, 20).Value, Cells(i, 23).Value, True)

    Else

        Cells(i, 25).Value = WorksheetFunction.Norm_Dist(Cells(i, 20).Value,
            Cells(i, 20).Value, Cells(i, 23).Value, True) + _
                        WorksheetFunction.Norm_Dist(Cells(i, 22).Value,
                            Cells(i, 20).Value, Cells(i, 24).Value, True) - _
                        WorksheetFunction.Norm_Dist(Cells(i, 20).Value,
                            Cells(i, 20).Value, Cells(i, 24).Value, True)

    End If

    i = i + 1

Loop



' Data in tails (uncalibrated) (AA:AA)

' Count lower tail
Range("AA4").Formula = "=sumproduct(--(R2C22:R" & (NoTrue + 1) & "C22<R2C19:R" &
    (NoTrue + 1) & "C19))"

' Count upper tail
Range("AA8").Formula = "=sumproduct(--(R2C22:R" & (NoTrue + 1) & "C22>R2C21:R" &
    (NoTrue + 1) & "C21))"
```

```vba
' Count true values above P50
Range("AA12").Formula = "=sumproduct(--(R2C22:R" & (NoTrue + 1) & "C22>R2C20:R"
    & (NoTrue + 1) & "C20))"

' Count true values below P50
Range("AA16").Formula = "=sumproduct(--(R2C22:R" & (NoTrue + 1) & "C22<R2C20:R"
    & (NoTrue + 1) & "C20))"




' Reset i, j
i = 2
j = 2

' Input the datasets of the fields where the rand() function is TRUE in another
    table (AD:AL)
Do While Cells(i, 17) <> ""

    Cells(i, 30).Value = Cells(i, 17).Value
    Cells(i, 31).Value = Cells(i, 18).Value
    Cells(i, 33).Formula = "=R" & i & "C20 * R4C40"
    Cells(i, 32).Formula = "=IF(R[0]C[1] - R8C40 * (R" & i & "C20 - R" & i & "
        C19)<0,0,R[0]C[1] - R8C40 * (R" & i & "C20 - R" & i & "C19))"
    Cells(i, 34).Formula = "=R[0]C[-1] - R12C40 * (R" & i & "C20 - R" & i & "C21
        )"
    Cells(i, 35).Value = Cells(i, 22).Value

    Cells(i, 36).Formula = "=(R" & i & "C33 - R" & i & "C32) / 1.28"
    Cells(i, 37).Formula = "=(R" & i & "C34 - R" & i & "C33) / 1.28"

    Cells(i, 38).Formula = "=IF(R[0]C[-3] < R[0]C[-5], Norm.Dist(R" & i & "C35,
        R" & i & "C33, R" & i & "C36, True), Norm.Dist(R" & i & "C33, R" & i & "
        C33, R" & i & "C36, True) + Norm.Dist(R" & i & "C35, R" & i & "C33, R" &
        i & "C37, True) - Norm.Dist(R" & i & "C33, R" & i & "C33, R" & i & "C37
        , True)) "

    i = i + 1

Loop




' Data in tails (calibrated) (AP:AP)

' Minimize Median(Beta*Mean - Actual)
Range("AP20").FormulaArray = "=Median(R2C33:R" & (NoTrue + 1) & "C33-R2C35:R" &
    (NoTrue + 1) & "C35)"

' Count lower tail
Range("AP4").Formula = "=sumproduct(--(R2C35:R" & (NoTrue + 1) & "C35<R2C32:R" &
    (NoTrue + 1) & "C32))"

' Count upper tail
Range("AP8").Formula = "=sumproduct(--(R2C35:R" & (NoTrue + 1) & "C35>R2C34:R" &
    (NoTrue + 1) & "C34))"

' Count true values above P50
Range("AP12").Formula = "=sumproduct(--(R2C35:R" & (NoTrue + 1) & "C35>R2C33:R"
    & (NoTrue + 1) & "C33))"

' Count true values below P50
Range("AP16").Formula = "=sumproduct(--(R2C35:R" & (NoTrue + 1) & "C35<R2C33:R"
    & (NoTrue + 1) & "C33))"
```

```vba
'Calibration Parameters Calculation

' Beta
flg = False

If Range("AP12").Value > Range("N12").Value Then

    Do While Range("AP20").Value > 0

        Range("AN4").Value = Range("AN4").Value + 0.1

        If Range("AP20").Value < 0 Then

            Do While Range("AP20").Value < 0

                Range("AN4").Value = Range("AN4").Value - 0.01

                If Range("AP20").Value > 0 Then

                    Do While Range("AP20").Value > 0

                        Range("AN4").Value = Range("AN4").Value + 0.001

                        If Range("AP20").Value < 0 Then

                            Do While Range("AP20").Value < 0

                                Range("AN4").Value = Range("AN4").Value - 0.0001

                                If Range("AP20").Value > 0 Then

                                    flg = True

                                    Exit Do

                                End If

                            Loop

                        End If

                        If flg = True Then

                            Exit Do

                        End If

                    Loop

                End If

                If flg = True Then

                    Exit Do

                End If

            Loop

        End If

        If flg = True Then

            Range("AN4").Value = Range("AN4").Value + 0.0001

            Exit Do

        End If

    Loop
```

```vba
ElseIf Range("AP12").Value < Range("N12").Value Then
    Do While Range("AP20").Value > 0
        Range("AN4").Value = Range("AN4").Value - 0.1
        If Range("AP20").Value < 0 Then
            Do While Range("AP20").Value < 0
                Range("AN4").Value = Range("AN4").Value + 0.01
                If Range("AP20").Value > 0 Then
                    Do While Range("AP20").Value > 0
                        Range("AN4").Value = Range("AN4").Value - 0.001
                        If Range("AP20").Value < 0 Then
                            Do While Range("AP20").Value < 0
                                Range("AN4").Value = Range("AN4").Value + 0.0001
                                If Range("AP20").Value > 0 Then
                                    flg = True
                                    Exit Do
                                End If
                            Loop
                        End If
                        If flg = True Then
                            Exit Do
                        End If
                    Loop
                End If
                If flg = True Then
                    Exit Do
                End If
            Loop
        End If
        If flg = True Then
            Exit Do
        End If
    Loop
End If
```

```vba
' Alpha L
flgl = False

If Range("AP4").Value > Range("N4").Value Then

    Do Until Range("AP4").Value = Range("N4").Value

        Range("AN8").Value = Range("AN8").Value + 0.1

        If Range("AP4").Value <= Range("N4").Value Then

            Do While Range("AN4").Value <= Range("N4").Value

                Range("AN8").Value = Range("AN8").Value - 0.01

                If Range("AP4").Value > Range("N4").Value Then

                    Do Until Range("AP4").Value <= Range("N4").Value

                        Range("AN8").Value = Range("AN8").Value + 0.001

                        If Range("AP4").Value <= Range("N4").Value Then

                            Do While Range("AN4").Value <= Range("N4").Value

                                Range("AN8").Value = Range("AN8").Value - 0.0001

                                If Range("AP4").Value > Range("N4").Value Then

                                    flgl = True

                                    Exit Do

                                End If

                            Loop

                        End If

                        If flgl = True Then

                            Exit Do

                        End If

                    Loop

                End If

                If flgl = True Then

                    Exit Do

                End If

            Loop

        End If

        If flgl = True Then

            Range("AN8").Value = Range("AN8").Value + 0.0001

            Exit Do

        End If

    Loop
```

```vba
ElseIf Range("AP4").Value < Range("N4").Value Then
    Do Until Range("AP4").Value = Range("N4").Value
        Range("AN8").Value = Range("AN8").Value - 0.1
        If Range("AP4").Value >= Range("N4").Value Then
            Do While Range("AP4").Value >= Range("N4").Value
                Range("AN8").Value = Range("AN8").Value + 0.01
                If Range("AP4").Value < Range("N4").Value Then
                    Do Until Range("AP4").Value >= Range("N4").Value
                        Range("AN8").Value = Range("AN8").Value - 0.001
                        If Range("AP4").Value >= Range("N4").Value Then
                            Do While Range("AP4").Value >= Range("N4").Value
                                Range("AN8").Value = Range("AN8").Value + 0.0001
                                If Range("AP4").Value < Range("N4").Value Then
                                    flgl = True
                                    Exit Do
                                End If
                            Loop
                        End If
                        If flgl = True Then
                            Exit Do
                        End If
                    Loop
                End If
                If flgl = True Then
                    Exit Do
                End If
            Loop
        End If
        If flgl = True Then
            Range("AN8").Value = Range("AN8").Value - 0.0001
            Exit Do
        End If
    Loop
End If
```

```vba
' Alpha R
flgr = False

If Range("AP8").Value > Range("N8").Value Then

    Do Until Range("AP8").Value = Range("N8").Value

        Range("AN12").Value = Range("AN12").Value + 0.1

        If Range("AP8").Value <= Range("N8").Value Then

            Do While Range("AP8").Value <= Range("N8").Value

                Range("AN12").Value = Range("AN12").Value - 0.01

                If Range("AP8").Value > Range("N8").Value Then

                    Do Until Range("AP8").Value <= Range("N8").Value

                        Range("AN12").Value = Range("AN12").Value + 0.001

                        If Range("AP8").Value <= Range("N8").Value Then

                            Do While Range("AP8").Value <= Range("N8").Value

                                Range("AN12").Value = Range("AN12").Value -
                                    0.0001

                                If Range("AP8").Value > Range("N8").Value Then

                                    flgr = True

                                    Exit Do

                                End If

                            Loop

                        End If

                        If flgr = True Then

                            Exit Do

                        End If

                    Loop

                End If

                If flgr = True Then

                    Exit Do

                End If

            Loop

        End If

        If flgr = True Then

            Range("AN12").Value = Range("AN12").Value + 0.0001

            Exit Do

        End If

    Loop
```

```vba
ElseIf Range("AP8").Value < Range("N8").Value Then
    Do Until Range("AP8").Value = Range("N8").Value
        Range("AN12").Value = Range("AN12").Value - 0.1
        If Range("AP8").Value >= Range("N8").Value Then
            Do While Range("AP8").Value >= Range("N8").Value
                Range("AN12").Value = Range("AN12").Value + 0.01
                If Range("AP8").Value < Range("N8").Value Then
                    Do Until Range("AP8").Value >= Range("N8").Value
                        Range("AN12").Value = Range("AN12").Value - 0.001
                        If Range("AP8").Value >= Range("N8").Value Then
                            Do While Range("AP8").Value >= Range("N8").Value
                                Range("AN12").Value = Range("AN12").Value +
                                    0.0001
                                If Range("AP8").Value < Range("N8").Value Then
                                    flgl = True
                                    Exit Do
                                End If
                            Loop
                        End If
                        If flgl = True Then
                            Exit Do
                        End If
                    Loop
                End If
                If flgl = True Then
                    Exit Do
                End If
            Loop
        End If
        If flgl = True Then
            Range("AN12").Value = Range("AN12").Value - 0.0001
            Exit Do
        End If
    Loop
End If
```

```vba
' Reset i, j
i = 2
j = 2


' Input the datasets of the fields where the rand() function is FALSE - without
    calibration - in another table (AS:BA)
Do While Cells(i, 7) <> ""

    If Cells(i, 7) = "False" Then

        Cells(j, 45).Value = Cells(i, 1).Value
        Cells(j, 46).Value = Cells(i, 2).Value
        Cells(j, 48).Value = Cells(i, 4).Value
        Cells(j, 47).Value = Cells(i, 3).Value
        Cells(j, 49).Value = Cells(i, 5).Value
        Cells(j, 50).Value = Cells(i, 6).Value

        Cells(j, 51).Value = (Cells(j, 48).Value - Cells(j, 47).Value) / 1.28
        Cells(j, 52).Value = (Cells(j, 49).Value - Cells(j, 48).Value) / 1.28

        Cells(j, 53).Formula = "=IF(R[0]C[-3] < R[0]C[-5], Norm.Dist(R" & j & "
            C50, R" & j & "C48, R" & j & "C51, True), Norm.Dist(R" & j & "C48, R
            " & j & "C48, R" & j & "C51, True) + Norm.Dist(R" & j & "C50, R" & j
             & "C48, R" & j & "C52, True) - Norm.Dist(R" & j & "C48, R" & j & "
            C48, R" & j & "C52, True)) "

        j = j + 1

    End If

    i = i + 1

Loop


' Data in tails (out-of-sample uncalibrated) (BC:BC)

' Count lower tail
Range("BC4").Formula = "=sumproduct(--(R2C50:R" & (NoTrue + 1) & "C50<R2C47:R" &
    (NoTrue + 1) & "C47))"

' Count upper tail
Range("BC8").Formula = "=sumproduct(--(R2C50:R" & (NoTrue + 1) & "C50>R2C49:R" &
    (NoTrue + 1) & "C49))"

' Count true values above P50
Range("BC12").Formula = "=sumproduct(--(R2C50:R" & (NoTrue + 1) & "C50>R2C48:R"
    & (NoTrue + 1) & "C48))"

' Count true values below P50
Range("BC16").Formula = "=sumproduct(--(R2C50:R" & (NoTrue + 1) & "C50<R2C48:R"
    & (NoTrue + 1) & "C48))"




' Reset i, j
i = 2
j = 2


' Input the datasets of the fields where the rand() function is FALSE - with
    calibration - in another table (BF:BN)
Do While Cells(i, 7) <> ""
```

```vba
    If Cells(i, 7) = "False" Then

        Cells(j, 58).Value = Cells(i, 1).Value
        Cells(j, 59).Value = Cells(i, 2).Value
        Cells(j, 61).Value = Cells(i, 4).Value * Range("AN4")
        Cells(j, 60).Value = Cells(j, 61).Value - Range("AN8") * (Cells(i, 4) -
            Cells(i, 3).Value)

        If Cells(j, 60).Value < 0 Then

            Cells(j, 60).Value = 0

        End If

        Cells(j, 62).Value = Cells(j, 61).Value - Range("AN12") * (Cells(i, 4) -
            Cells(i, 5).Value)
        Cells(j, 63).Value = Cells(i, 6).Value

        Cells(j, 64).Value = (Cells(j, 61).Value - Cells(j, 60).Value) / 1.28
        Cells(j, 65).Value = (Cells(j, 62).Value - Cells(j, 60).Value) / 1.28

        Cells(j, 66).Formula = "=IF(R[0]C[-3] < R[0]C[-5], Norm.Dist(R" & j & "
            C63, R" & j & "C61, R" & j & "C64, True), Norm.Dist(R" & j & "C61, R
            " & j & "C61, R" & j & "C64, True) + Norm.Dist(R" & j & "C63, R" & j
            & "C61, R" & j & "C65, True) - Norm.Dist(R" & j & "C61, R" & j & "
            C61, R" & j & "C65, True)) "

        j = j + 1

    End If

    i = i + 1

Loop


' Data in tails (out-of-sample calibrated) (BP:BP)

' Count lower tail
Range("BP4").Formula = "=sumproduct(--(R2C63:R" & (NoTrue + 1) & "C63<R2C60:R" &
    (NoTrue + 1) & "C60))"

' Count upper tail
Range("BP8").Formula = "=sumproduct(--(R2C63:R" & (NoTrue + 1) & "C63>R2C62:R" &
    (NoTrue + 1) & "C62))"

' Count true values above P50
Range("BP12").Formula = "=sumproduct(--(R2C63:R" & (NoTrue + 1) & "C63>R2C61:R"
    & (NoTrue + 1) & "C61))"

' Count true values below P50
Range("BP16").Formula = "=sumproduct(--(R2C63:R" & (NoTrue + 1) & "C63<R2C61:R"
    & (NoTrue + 1) & "C61))"




' Plotting Probability for datasets where rand() is FALSE

' Reset i, j
i = 2
j = 1

' Variable # (BS:BS)
Do While Cells(i, 7) <> ""

    If Cells(i, 7) = "False" Then

        Cells(j + 1, 71).Value = j
```

```vba
        j = j + 1

    End If

    i = i + 1

Loop



' Reset i
i = 2

' Plotting Probability (BT:BT)
Do Until i = (j + 1)

    Cells(i, 72).Value = (Cells(i, 71) - 0.4) / (1 + (j - 1) - 2 * 0.4)

    i = i + 1

Loop


' Data in tails (out-of-sample plotting probability) (BV:BV)

' Count lower tail
Range("BV4").Formula = "=CountIf(R2C72:R" & (NoFalse + 1) & "C72,""<0.1"")"

' Count upper tail
Range("BV8").Formula = "=COUNTIF(R2C72:R" & (NoFalse + 1) & "C72,"">0.9"")"

' Count true values above P50
Range("BV12").Formula = "=COUNTIF(R2C72:R" & (NoFalse + 1) & "C72,"">0.5"")"

' Count true values below P50
Range("BV16").Formula = "=COUNTIF(R2C72:R" & (NoFalse + 1) & "C72,""<0.5"")"




' Calculate sum of out-of-sample uncalibrated data
tempuLT = Range("BC4").Value
uLT = uLT + tempuLT

tempuUT = Range("BC8").Value
uUT = uUT + tempuUT

tempuAP50 = Range("BC12").Value
uAP50 = uAP50 + tempuAP50

tempuBP50 = Range("BC16").Value
uBP50 = uBP50 + tempuBP50


' Calculate sum of out-of-sample calibrated data
tempcLT = Range("BP4").Value
CLT = CLT + tempcLT

tempcUT = Range("BP8").Value
cUT = cUT + tempcUT

tempcAP50 = Range("BP12").Value
cAP50 = cAP50 + tempcAP50

tempcBP50 = Range("BP16").Value
cBP50 = cBP50 + tempcBP50
```

```vba
' Calculate sum of plotting probability applicable to the out-of-sample sample
    size
temppppLT = Range("BV4").Value
ppLT = ppLT + temppppLT

temppppUT = Range("BV8").Value
ppUT = ppUT + temppppUT

temppppAP50 = Range("BV12").Value
ppAP50 = ppAP50 + temppppAP50

temppppBP50 = Range("BV16").Value
ppBP50 = ppBP50 + temppppBP50


' Absolute difference from the PP (uncalibrated)

tempuLTabs = Abs(Range("BC4").Value - Range("BV4").Value)
uLTabs = tempuLTabs + uLTabs

tempuUTabs = Abs(Range("BC8").Value - Range("BV8").Value)
uUTabs = tempuUTabs + uUTabs

tempuaP50abs = Abs(Range("BC12").Value - Range("BV12").Value)
uaP50abs = tempuaP50abs + uaP50abs

tempubP50abs = Abs(Range("BC16").Value - Range("BV16").Value)
ubP50abs = tempubP50abs + ubP50abs



' Absolute difference from the PP (calibrated)

tempcLTabs = Abs(Range("BP4").Value - Range("BV4").Value)
cLTabs = tempcLTabs + cLTabs

tempcUTabs = Abs(Range("BP8").Value - Range("BV8").Value)
cUTabs = tempcUTabs + cUTabs

tempaP50abs = Abs(Range("BP12").Value - Range("BV12").Value)
aP50abs = tempaP50abs + aP50abs

tempbP50abs = Abs(Range("BP16").Value - Range("BV16").Value)
bP50abs = tempbP50abs + bP50abs




' Average of the calibration parameters for each iteration

tempbeta = Range("AN4").Value
tempalphaL = Range("AN8").Value
tempalphaR = Range("AN12").Value

beta = (tempbeta + beta)
alphaL = (tempalphaL + alphaL)
alphaR = (tempalphaR + alphaR)

Cells(looping + 1, 84).Value = looping
Cells(looping + 1, 85).Value = alphaL / looping
Cells(looping + 1, 86).Value = beta / looping
Cells(looping + 1, 87).Value = alphaR / looping

Cells(looping + 1, 90).Value = tempalphaL
Cells(looping + 1, 91).Value = tempbeta
Cells(looping + 1, 92).Value = tempalphaR
```

```vba
' Reset i
' i = 2

' Count false cells
lastRowFalse = Range("BT" & Rows.Count).End(xlUp).Row
CountFalse = WorksheetFunction.Count(Range("BT2:BT" & lastRowFalse & ""))


' Sum up RMSE values for ALL iterations
For i = 1 To CountFalse

    temprmseu = (Application.WorksheetFunction.Small(Range("BA2:BA" &
        lastRowFalse & ""), i) - Application.WorksheetFunction.Small(Range("BT2:
        BT" & lastRowFalse & ""), i)) ^ 2
    rmseu = rmseu + temprmseu

    temprmsec = (Application.WorksheetFunction.Small(Range("BN2:BN" &
        lastRowFalse & ""), i) - Application.WorksheetFunction.Small(Range("BT2:
        BT" & lastRowFalse & ""), i)) ^ 2
    rmsec = rmsec + temprmsec

    Cells(i + 1, looping + 99) = Application.WorksheetFunction.Small(Range("BN2:
        BN" & lastRowFalse & ""), i)

    Cells(i + 29, looping + 99) = Application.WorksheetFunction.Small(Range("BA2
        :BA" & lastRowFalse & ""), i)

Next

rmsec = Sqr(rmsec / CountFalse)
'totrmsec = totrmsec + rmsec

Cells(looping + 1, 95).Value = rmsec

'Cells(looping + 1, 86).Value = totrmsec / looping

rmsec = 0

'Cells(looping + 1, 85).Value = Sqr(rmseu / (looping * CountFalse))
'Cells(looping + 1, 86).Value = Sqr(rmsec / (looping * CountFalse))



' RMSE for each iteration




' This marks the end of the main framework (looping)
Next



' Calculate the average of out-of-sample uncalibrated data
uLT = uLT / (looping - 1)
uUT = uUT / (looping - 1)
uAP50 = uAP50 / (looping - 1)
uBP50 = uBP50 / (looping - 1)


' Calculate the average of out-of-sample calibrated data
CLT = CLT / (looping - 1)
cUT = cUT / (looping - 1)
cAP50 = cAP50 / (looping - 1)
cBP50 = cBP50 / (looping - 1)
```

```vba
' Calculate the average of plotting probability applicable to the out-of-sample
    sample size
ppLT = ppLT / (looping - 1)
ppUT = ppUT / (looping - 1)
ppAP50 = ppAP50 / (looping - 1)
ppBP50 = ppBP50 / (looping - 1)


' Calculate absolute difference
cLTabs = cLTabs / (looping - 1)
cUTabs = cUTabs / (looping - 1)
aP50abs = aP50abs / (looping - 1)
bP50abs = bP50abs / (looping - 1)

Range("CE3").Value = cLTabs
Range("CE4").Value = cUTabs
Range("CE5").Value = aP50abs
Range("CE6").Value = bP50abs

uLTabs = uLTabs / (looping - 1)
uUTabs = uUTabs / (looping - 1)
uaP50abs = uaP50abs / (looping - 1)
ubP50abs = ubP50abs / (looping - 1)

Range("CD3").Value = uLTabs
Range("CD4").Value = uUTabs
Range("CD5").Value = uaP50abs
Range("CD6").Value = ubP50abs


Application.ScreenUpdating = True


' Print the average amount of data in the tails in the worksheet
Range("CA3").Value = uLT
Range("CA4").Value = uUT
Range("CA5").Value = uAP50
Range("CA6").Value = uBP50

Range("CB3").Value = CLT
Range("CB4").Value = cUT
Range("CB5").Value = cAP50
Range("CB6").Value = cBP50

Range("CC3").Value = ppLT
Range("CC4").Value = ppUT
Range("CC5").Value = ppAP50
Range("CC6").Value = ppBP50

'Range("CE5").Value = CountFalse
'Range("CD5").Formula = "=If(IsOdd(CE5), ""PS! Odd number of data"", """")"
'Range("CD5").Value = Range("CD5").Value
'Range("CE5").Value = ""


'Determine how many seconds code took to run
  MinutesElapsed = Format((Timer - StartTime) / 86400, "hh:mm:ss")

'Notify user in seconds
  MsgBox "This code ran successfully in " & MinutesElapsed & " minutes", _
      vbInformation


End Sub
```