Norwegian School of Economics
Bergen, Spring 2020

How Do Different Evaluation Methods Affect Outcomes in Procurement?

Monica Benonisen & Marianne Strand

Supervisor: Malin Arve

Master's Thesis in Economics and Business Administration

Major in Business Analysis and Performance Management

NORWEGIAN SCHOOL OF ECONOMICS

Preface

Writing this thesis has been an educational and inspiring process. We have challenged ourselves both in the choice of language and method, while at the same time learning a great deal about an exciting and relevant area of procurement today.

We want to express our gratitude to our supervisor, Malin Arve, for all her support throughout this writing process. Her critical insights and vast knowledge of evaluation methods has been pivotal to our process of writing this master's thesis.

Our gratitude also extends to our family and friends who have provided us with feedback and general support during this semester.

Finally, we would like to thank Sykehusinnkjøp for providing us with detailed information regarding their procurement practice. This information was helpful when choosing our topic and beneficial when we carried out our study. We would like to thank Hanna Udnæs Hoel and Per-Marthin Karlsen in particular, for informative discussions and explanations regarding public procurement and evaluation methods.

Norwegian School of Economics

Bergen, June 2020

Monica Benonisen — Marianne Strand

Monica Benonisen — Marianne Strand

Abstract

This thesis uses simulation and regression analysis to investigate how different evaluation methods affect outcomes in procurement. In order to simulate the data, we have made our own algorithm in R Studio to answer our proposed questions. This algorithm can easily be adapted by others who want to simulate similar data or run simulation with other assumptions and parameters. Most procurement in Norway involves evaluating tenders based on both price and quality aspects. Price is evaluated by using scoring rules, while quality aspects are evaluated by expert panels and, in some cases, adjusted by the use of normalisation. By first investigating scoring rules, we find that the relative scoring rules recommended by the Norwegian Digitalisation Agency (NDA), and the most commonly used in practice, have serious drawbacks, suggesting that they are not the most suitable. In addition, we know from previous literature that these rules are unpredictable for bidders to use. In this thesis, we therefore provide additional insights, showing that these relative scoring rules also weigh quality relatively less compared to price during evaluation. Finally, we prove that normalisation has adverse effects on outcomes in procurement. The NDA recommends procurers to adjust, or normalise, the quality scores assigned by expert panels. In this thesis, we show that normalisation changes the relative weight of quality in a tender evaluation, leading to arbitrarily and unpredictable outcomes. By rather recommending expert panels to evaluate quality aspects relatively, normalisation can be avoided.

Keywords: Simulation, R, Public Procurement, MEAT, Scoring Rules, Contract Awarding Methods, Scoring Auction

Contents

1 Introduction ... *1*

1.1 Outline ... **4**

2 Legal Framework ... *5*

2.1 The EU Directive on Public Procurement ... **5**

2.2 Norwegian Legislation ... **6**

2.3 The Norwegian Digitalisation Agency ... **7**

3 Evaluation Methods ... *8*

3.1 Reflection of the Procurer´s Preferences ... **9**

3.2 Evaluation of Price ... **12**

3.2.1 Absolute Scoring Rules ... 13

3.2.1.1 Linear Rules ... 13

3.2.1.2 Parabolic Rule ... 15

3.2.1.3 Comparison between the Absolute Rules ... 16

3.2.2 Relative Scoring Rules ... 17

3.3 Evaluation of Quality ... **18**

4 Simulation of Data ... *23*

4.1 The Model Environment ... **23**

4.2 The Simulation and its Output ... **25**

4.3 Limitations ... **27**

5 Logistic Regression ... *28*

6 Analysis ... *31*

6.1 Overview of Different Outcomes ... **31**

6.2 Implications of Different Scoring Rules ... **33**

6.2.1 Different Scoring Rules Assign Price Different Weights ... 34

6.2.2 When Does the Different Rules Provide Different Winners? ... 37

6.2.3 Conclusion – Impact of Different Scoring Rules ... 40

6.3 Implications of Normalisation ... **41**

6.3.1 Regressions and Findings ... 42

6.3.2 Implications of Findings ... 48

6.3.2.1 Normalisation Have a High Impact when Quality is Low ... 48

6.3.2.2 Normalisation Reflects a Higher Willingness to Pay for Low Quality ... 50

6.3.2 Conclusion – Impact of Normalisation ... 51

7 Concluding Remarks and Reflections *52*

7.1 Conclusion **52**

7.2 Reflections and Future Research **54**

References *55*

Appendix *58*

A1 Regressing the Pairwise Differences from Chapter 6.1 ..... **58**

A2 Matrices Used to Analyse Patterns in 0/1-Matrix between Method with and without Normalisation **62**

A3 Robustness of Logistic Regressions in Chapter 6 ..... **65**

A3.1 Model with all Predictors 65

A3.2 Testing for Strong Multicollinearity 66

A3.3 Model Accuracy 67

A4 Algorithms in R-Script **71**

A4.1 Main Algorithm (copied directly from R-script) 71

A4.2 Algorithm for the Overview (copied directly from R-script) 81

List of Figures

3.1 Evaluation Methods and Three Types of Scoring Rules 8
3.2 Linear Rule with and without a Price Threshold 14
3.3 Graph Parabolic Rule 15
3.4 Graphs Absolute Rules 16
5.1 Linear Regression vs. Logistic Regression 28
6.1 Percentage of Different Winners 31

List of Tables

3.1 Example of Evaluation Scheme 11
3.2 Absolute Scale without Normalisation 19
3.3 Absolute Scale with Absolute Normalisation 19
3.4 Absolute Scale with Relative Normalisation 20
3.5 Case 1 - Relative Scale with Relative Normalisation 21
3.6 Case 2 - Relative Scale with Relative Normalisation 22
4.1 Overview of Different Evaluation Methods 24
4.2 Score Matrix 25
4.3 Rank Matrix 26
4.4 0/1 Matrix 26
6.1 Price Scores and Difference in Price Scores for Different Price Combinations and Different Scoring Rules 34
6.2 Rules Compared with Regards to the Difference in Price Score between Bidders ... 36
6.3 Quality Difference that Lead to Different Outcomes in the Procurement 37
6.4 Percentage Amount of Different Winners for Different Price Combinations Comparing Different Scoring Rules 39
6.5 Regression Results Finding 1 43
6.6 Regression Results Finding 2 43
6.7 Regression Results Finding 3 44
6.8 Regression Results Finding 4 45
6.9 Regression Results Finding 5 46
6.10 Regression Results Interaction Terms 47
A1.1 Overview of Different Evaluation Methods 58
A1.2 1&2 as Reference Group 59
A1.3 1&4 as Reference Group 60
A1.4 2&4 as Reference Group 60
A1.5 3&4 as Reference Group 60
A1.6 Differences Due to Category of Scoring Rules and Normalisation 61
A2.1 Price Matrix 62
A2.2 Quality Matrix 63
A3.1.1 Implications of Different Bid Characteristics on Probability 65
A3.2.1 VIF for the Model in Table A3.1.1 67
A3.3.1 Confusion Matrix 68
A3.3.2 Confusion Matrix Belonging to Our Model 70
A3.3.3 Key Measures 70

1 Introduction

The Norwegian public sector procures goods and services for about 500 billion NOK annually (Avdeling for offentlige anskaffelser, 2020a). Therefore, public procurers are subject to strict regulation ensuring fairness and predictability in the procurement process. The procurer uses evaluation methods to evaluate the various dimensions of a tender. From the 2014 EU Directive on Public Procurement, there are two main categories of evaluation methods stipulated; lowest price and the most economically advantageous tender (MEAT). In our study, we will only focus on the latter. Here, the procurer not only considers price and minimum requirements during evaluation, but also assess the tenders based on certain quality aspects (Bergman & Lundberg, 2013, p. 74).

When both price and quality aspects are being evaluated, we need to either transform the price into the same unit as quality, or the quality aspect (often a score) into monetary units. We will in this thesis focus on the first method; transforming price into a score and keeping the quality score as it is. Finally, the two scores are added together. The tender with the highest total score is the one that offers the highest quality possible at the best achievable price. This tender is therefore chosen by the procurer.

When transforming price into a score, scoring rules are commonly used. A scoring rule provides mathematical formulas to calculate the price score. There are two main groups of scoring rules; absolute and relative. Absolute scoring rules provide benchmarks on what is seen as a high and/or a low price, allowing the bidders to calculate their own price scores before the bidding phase. On the contrary, the relative scoring rules benchmark the bids relative to each other. Scholars favour the absolute rules as they provide bidders with knowledge of the price/quality preferences of the procurer, thus enabling them to formulate their best offer possible (Dini, Pachini & Valetti, 2006, p. 304-317). This ensures greater predictability for both the bidder and the procurer. Relative scoring rules, however, do not provide bidders with any information regarding preferences and are therefore not predictable. Both national and international legislation states that an important aim of procurement is to ensure predictability in the process. Despite this, relative scoring rules dominate in practice. Bergman & Lundberg (2013, p. 81) find that relative scoring rules are three times more common in practice than absolute scoring rules. Furthermore, even the Norwegian Digitalisation Agency (NDA) recommends using relative scoring rules in procurement (Avdeling for offentlige anskaffelser, 2019).

Therefore, we believe that there are certain misconceptions concerning the use and impact of different scoring rules. More specifically, that there are misconceptions regarding the impact of weights in the scoring rules. As the tender is evaluated on both price and quality, the NDA recommends using a weighing function to assign price and quality their respective weights in each tender (Avdeling for offentlige anskaffelser, 2020b). These weights are supposed to reflect the procurer's preferences in the trade-off between price and quality. However, the different scoring rules themselves also implies different weights of price compared to quality. Our perception is that this latter information is widely overlooked. We will in this thesis, therefore provide the procurer with new insights regarding how preferences are being reflected when using different scoring rules.

There is also an ongoing debate regarding the quality evaluation. Typically, an expert panel assess the perceived quality of a product and score the tenders according to certain criteria. Afterwards, it is surprisingly common to adjust, or *normalise,* the score. Despite none of the bidders having a quality perceived as a maximum by the expert panel, normalisation involves awarding the bidder with the highest perceived quality a maximum score anyway. The quality score of the other bidders are then adjusted accordingly. This practice has been subject of discussion and criticism in Norway in recent months. Some practitioners claim normalisation may change the weight of quality during evaluation, thereby resulting in arbitrary and less predictable outcomes (Ellingsen & Haukeli, 2020).

This study aims to investigate how different evaluation methods affect outcomes in procurement and we will address both the issue of relative and absolute scoring, as well as the question of normalisation. More specifically, we seek to examine how both different scoring rules and normalisation may have an impact on the preferences of the procurer, subsequently leading to a change in outcome of the procurement. By investigating this, we want to be able to comment on whether the widespread use and recommendations of relative scoring rules are actually rational. We also want to use our findings to draw some conclusions about normalisation. Is the recent criticism reasonable?

To answer our research questions, we have cooperated with Sykehusinnkjøp, a public enterprise solely responsible for handling the procurement processes on behalf of all health trusts in Norway. At the beginning of 2020, the enterprise held a procurement portfolio consisting of

1860 agreements worth about 4.5 billion NOK.[1] Access to these agreements make the health trusts able to purchase a wide variety of equipment necessary to ensure safe an efficient care of patients; from toilet paper and syringes, to X-ray machines and ambulance helicopters. Sykehusinnkjøp has given us access to their evaluation methods, enabling us to produce our own algorithm to simulate tender evaluations. By simulation, we are able to examine how different methods may have an impact on the preferences of the procurer, subsequently leading to a change of outcome in procurement. Furthermore, by conducting a regression analysis, we can investigate when normalisation have an impact on the quality score of a tender and how this practice may affect the overall outcome in procurement.

Our motivation to write this thesis is threefold. First, we would like to understand evaluation methods in more depth because of its necessity in procurement today. Every tender in procurement, both in public and private sector, is subject to assessment through evaluation methods. Being able to cooperate with Sykehusinnkjøp, one of the biggest procurers in Norway, provides an unique opportunity to get a thorough understanding on how the different methods are being used in practice, thus bridging the gap between theory and practice.

Second, a user study conducted by the NDA shows that many procurers display little awareness when choosing evaluation method. They often use the same methods for all procurements without considering the type of product or service being procured (Difi, 2015, p. 30). Thus, investigating whether different methods have an effect on the outcomes, are relevant as it enables procurers to make more informed decisions.

Finally, we perceive this thesis as being an opportunity to provide new insights to evaluation methods. Most theory focuses on discussing benefits and drawbacks of using the different scoring rules in evaluation. In this thesis, we therefore want to go beyond that. What is not as clearly stated in literature, is that different scoring rules themselves may lead to different weighing in the price/quality trade-off. Also, literature hardly mention normalisation. As this practice lately has been subject of debate in Norway, we want to investigate this in more depth. Is it true that normalisation may arbitrarily change the outcomes in procurement? Thus, an important motivation, is to provide procurers with additional information when choosing between evaluation methods.

[1] Information provided by Hanna Udnæs Hoel, Head of Department for Planning and Project Support at Sykehusinnkjøp's National Services Division

1.1 Outline

This thesis will be organised as follows: In Chapter 2, we present international and national legal framework procurers need to adhere to. Chapter 3 first presents what MEAT involves, narrowing our focus. Secondly, the different scoring rules used for calculating the price score of tenders are presented. Lastly, we will explain how quality scores are assigned, and how normalisation affects these scores. Chapter 4 and 5 outlines our methodological approaches, hereby the simulation being done, the model environment and the regression theory to be used in the analysis. In Chapter 6, we first provide an overview of the findings from our simulation. We then examine differences between scoring rules in part one of the analysis, before analysing normalisation in part two. Finally, in Chapter 7 we provide some concluding remarks and reflections upon the validity of our model.

2 Legal Framework

The aim of this chapter is to provide a better understanding of the international and national procurement legislation Norwegian procurers are subject to. Being member of the European Economic Area (EEA) Agreement, Norway is obliged to ensure that the national legislation is in line with EU legislation. This involves incorporating international law into national law. Furthermore, Norwegian procurers are also subject to governmental recommendations when choosing among evaluation methods.

2.1 The EU Directive on Public Procurement

The Norwegian procurement legislation is primarily based upon EU directives that Norway is legally bound to implement through the EEA agreement. The most recent directive is the EU Directive on Public Procurement launched in 2014. It repealed and replaced the previous directive from 2004. The new directive was hailed by the European Parliament as a tool for ensuring the best value for money (rather than the lowest price) and better quality of goods and services (Hobson, 2016). It was therefore upgraded to enable a greater use of quality criteria when awarding public contracts. Up to this point, there had been a heavy reliance on price as the predominant award criteria, which had the unfortunate effect of frequently limiting innovation and encouraging short-term thinking (RIF, 2020, p. 6).

The 2004 Directive on Public Procurement stipulated that contracts were to be awarded by using one out of two criteria, either (i) lowest priced tender or (ii) the most economically advantageous tender (MEAT). With the lowest price method, there are minimum requirements bidders will have to satisfy when submitting bids. The bids received will then be evaluated solely based on price. With MEAT, however, the procurer not only considers the price and minimum requirements, but also evaluates the tenders based on some quality aspects (Bergman & Lundberg, 2013, p. 73-74). This method is preferable for procurers when they do not know for certain what level of quality they prefer, as their preferences depend on the prices of different quality levels.

Bergman & Lundberg (2013, p. 74 & 79) have studied the extent of how these two methods are being used in practice. They performed a study consisting of a sample of 189 Swedish public procurements. Here, they found that the lowest priced tender was used in more than one-third of the procurements, while evaluation methods including both price and quality aspects (MEAT) were used in more than half of the procurements. We do not have any knowledge of similar studies performed in Norway. However, also Norwegian public procurers most commonly use evaluation methods assessing both price and quality aspects (Bjørnstad, 2019).

These trends are in line with EU recommendations, as the 2014 Directive places a much greater emphasis on evaluation of quality criteria other than simply the price (SIGMA, 2016, p. 21). Article 67 (2) states that public procurers are now obliged to award pubic contracts to the "most economically advantageous tender" (MEAT), which is explained as follows:

"The most economically advantageous tender from the point of view of the contracting authority shall be identified on the basis of price or cost, using a cost-effectiveness approach…and may include the best price-quality ratio, which shall be assessed on the basis of criteria, including qualitative, environmental and/or social aspects linked to the subject of matter of the public contract in question".

Based upon this definition, it is clear that the criterion considers the quality of the goods or services being procured, as well as the price (European Parliament, 2020). Although it is still possible to base an award solely on price, one interprets the directive as a strong recommendation of using MEAT, employing criteria other than, or in addition to, price (RIF, 2020, p. 6).

2.2 Norwegian Legislation

The 2014 EU Directive on Public Procurement has been transposed into Norwegian law by the Procurement Act of 17th June 2016 and the Procurement Regulations of 12th August 2016. Both the law and the regulations apply to the procurement of goods and services made by state authorities and public enterprises. The main purpose of the act is to promote an efficient use of society's resources. This implies ensuring that public enterprises act with integrity, so that the society have confidence in that public procurement will take place in a socially beneficial way (Regjeringen, 2017, p. 18-19).

The Procurement Act and the Procurement Regulations specify the guidelines Norwegian public procurers have to adhere to during the procurement process. Public procurement must be done in accordance to certain basic principles to ensure that the purpose of the law is fulfilled. These are enshrined in § 4:

I. competition
II. equal treatment
III. predictability
IV. verifiability
V. proportionality

These principles are the cornerstone throughout the Norwegian procurement legislation. Therefore, procurers have to keep the basic principles in mind when evaluating tenders and choosing among evaluation methods.

2.3 The Norwegian Digitalisation Agency

To get a better understanding on how the legislation is to be followed and interpreted, public procurers have access to support through the Norwegian Digitalisation Agency (NDA). This is the government's foremost tool in providing guidance to public enterprises on how to prepare and manage a procurement process.[2] NDA have the responsibility of overseeing the Norwegian Division for Public Procurement (NDPP). This division provides information about current legislation, the procurement process itself, and different evaluation methods to use in procurement (Avdeling for offentlige anskaffelser, 2020c). The resources are free and available online, and provide important guidelines when enterprises have questions regarding the procurement process and evaluation of tenders.

[2] This responsibility will be transferred to The Norwegian Agency for Public and Financial Management (DFØ) 01.09.2020.

3 Evaluation Methods

In this chapter, we will first present the different evaluation methods that exist within MEAT. We will then narrow our focus to one of these main groups of methods and explain the different aspects of it; how the price and quality of the tender is evaluated and how they are combined and weighted in accordance to each other.

When using MEAT, one is assessing both monetary values (like prices) and technical aspects (like quality) of tenders. It is therefore necessary to make the procurer able to evaluate the tenders on the basis of both dimensions. As a consequence, the use of evaluation methods under MEAT traditionally requires the procurer to adopt scoring rules. According to Bergman & Lundberg (2013, p. 75), a scoring rule can be defined as "a function that assigns a numerical value to different quality levels in a particular dimension or that transforms a value measured on one scale (price or quality) into a measure on another scale (price score or quality score, respectively)".

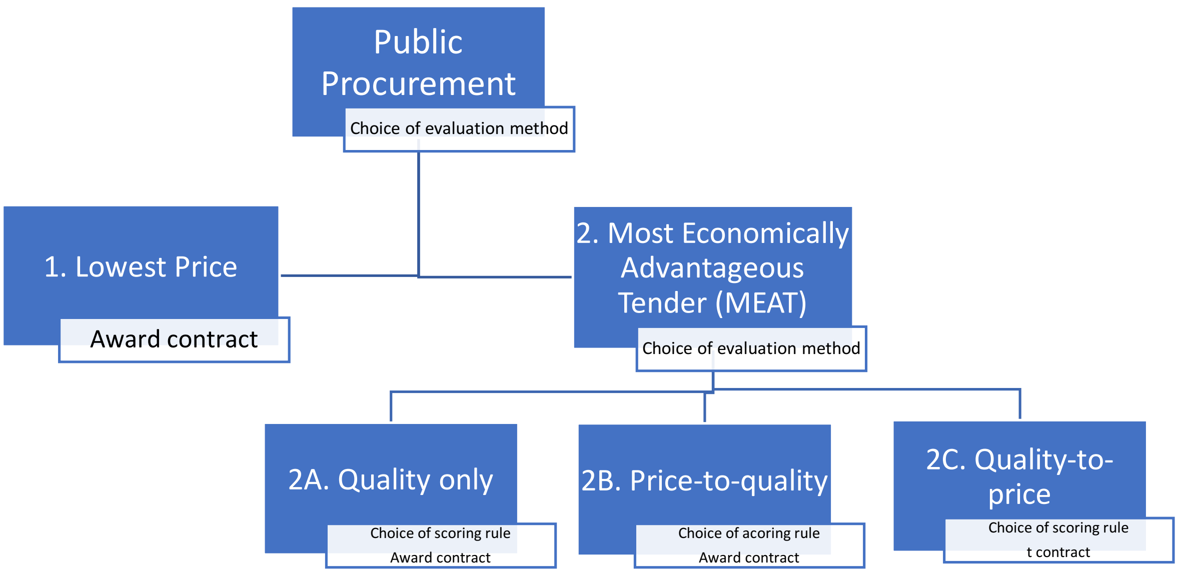

Figure 3.1: *Evaluation Methods and Three Categories of Scoring Rules. Our own illustration based upon Bergman & Lundberg (2013, p. 75).*

As illustrated in Figure 3.1 above, we categorize scoring rules into three main categories; quality-only (2A, also called beauty contest), price-to-quality (2B) and quality-to-price (2C) (Bergman & Lundberg, 2013, p. 75).

With quality-only (2A), the procurer has set a fixed price and the evaluation is based only on the quality offered (Bergman & Lindberg, 2013, p. 75). The quality is measured in more than one dimension, and a scoring rule is used to assign quality scores to these quality dimensions.[3] In quality-to-price scoring (2C), the price bid is kept in monetary terms, while the quality criteria are given monetary values by the procurer (Bergman & Lundberg, p. 80). In price-to-quality scoring (2B), all award criteria are converted into numerical points to be able to make a comparison of the submitted tenders. A scoring rule is therefore used to transform the price bids (in monetary terms) into points (numerical values) (Bergman & Lundberg, p. 75).

In this thesis, we will focus on price-to-quality scoring as some features of this method are subject of the ongoing debate in recent months.

3.1 Reflection of the Procurer´s Preferences

When evaluating both price and quality in a procurement, it is necessary for the procurer to be able to reflect his true preferences on the two dimensions, respectively. This implies that he must address what he is willing to pay for quality, meaning what price-quality combinations should be equivalent when assigning a score to price and quality (Dini et al., 2006, p. 296). This is important with regards to the basic principles of predictability and equal treatment stated in the public procurement law. However, it will also increase the procurer´s chances of receiving the best tender possible, given his preferences (Dini et al., 2006, p. 296).

What is common in practice, is to use a *weighing function* to combine the price score and the quality score. This function combines price and quality into a single value so that the different tenders can be compared to one another and ranked (Bergman & Lundberg, 2013, p. 75). For example, one could imagine that the procurer wanted price and quality to count equally in the evaluation. Then, the weighing function would require the price and quality scores to be multiplied with 50 % respectively. If there are several prices or quality aspects to be assessed, the weighing function can combine two or more price scores into a single overall price score or combine two or more quality scores into a single quality score (Bergman & Lundberg, 2013, p. 75). Here, one could imagine that there were two quality aspects to evaluate. If the price and quality is weighted equally, the two quality aspects have to be weighted within those 50 %

[3] «Choice of scoring rule» under Quality only (2A) refers to a scoring rule with different quality parameters and no price parameters (Bergman & Lindberg, 2013, p. 75).

"belonging" to the quality weighing. If the procurer wants both quality aspects to count equally in the evaluation, both quality scores have to be multiplied with 25 %. These percentage weights are used when the price and quality dimension have the same scale (both have a maximum score of 50 points for instance). However, one could also weigh the quality dimensions by designing the scales differently. If price is to be weighted 50 % it could be given a maximum of 50 points out of 100, and the two quality dimensions could be designed to range up to 25 points each. Therefore, giving the dimensions 50 points each and thereafter the weights, are redundant.

However, when using price-to-quality scoring, the scoring rule that transforms price into points are also weighting price in comparison to quality. Therefore, the choice of scoring rule also has an impact on how the procurer´s preferences are reflected. We have a perception that this is not quite understood in practice, both from what we see in NDA´s recommendations and the ongoing discussion today regarding normalisation. Hence, our focus in this thesis is to try to point out these misunderstandings by explaining and investigating this from a new point of view.

In order to understand how the scoring rules will indicate different preferences, we need to define and explain one important concept; the *monetary value of a point* (MVP) (Dini et al., 2006, p. 296). The MVP is the monetary discount necessary for a bidder to be able to obtain one additional point in the evaluation (Dini et al., 2006, p. 296). Knowing the MVP before submitting a bid, is of value for both the procurer and the bidder. When the bidders are able to structure their bids optimally, the procurer increases his chances of receiving the best tender possible given his preferences (Dini et al., 2006, p. 296-299). As this is crucial to understand how scoring rules reflects different preferences, we will provide an example to illustrate this.

We will use a simplified version of an actual tender conducted by Sykehusinnkjøp in our analysis. Sykehusinnkjøp is to procure a hearing implant. In our simplified version, they evaluate the tenders based on price and one quality aspect. The water column of the implant represents the quality aspect. Tenders will be awarded points according to the scheme in Table 3.1 below. In this scheme, the points are assigned linearly on the price dimension, so for instance if the bidder provides a price bid of 9500 NOK, he will get 2,5 price points. The bidder gets the maximum price points with a price of 2000 NOK. Hence, there is no point in decreasing the price further as he will only lose money. The highest price accepted is 10 000 NOK, which will be awarded zero points. On the quality dimension, we assume that there are only four different water columns available in the market. An increase in the level of quality is assigned a value of 20 points. But what are these 20 points worth in monetary terms?

The MVP is calculated by dividing the price range (10 000-2 000 = 8 000) by the total points awarded, which is 40. Hence, the MVP in this example is 8 000/40 = ***200 NOK per point***. This indicates that a reduction in the price by 200 NOK will result in one extra point for the bidder. It also implies that the 20 points awarded for increasing the level of quality, is worth 4 000 NOK (200 * 20). This reflects that the procurer is willing to pay 4000 NOK for one additional level of quality. Furthermore, it also informs the bidder that in order to gain 20 points, he can either increase quality by one level (example from 5 000 mm to 10 000 mm in water column) or reduce the price by 4 000 NOK. What the bidder will choose to do, depends on how much it costs the bidder to increase the quality from 5 000 mm to 10 000 mm.

Price		Quality	
Bid (NOK)	Points	Water column	Points
10000	0	5000	0
9000	5	10000	20
8000	10	15000	40
7000	15	20000	60
6000	20		
5000	25		
4000	30		
3000	35		
2000	40		

Max points = 40 + 60 = 100

Table 3.1: *Example of Evaluation Scheme*

If it costs more than 4 000 NOK, for instance 4 500 NOK, the bidder will be better off by decreasing the price bid by 4 000 NOK instead of increasing the quality by one level.

Therefore, if the procurer´s true preferences are not reflected in the scoring rule (scheme) and the MVP, he could lose out on otherwise better opportunities. Let´s assume the procurer keeps the same awarding scheme as above, but his actual monetary value of a higher level of quality is 5 000 NOK instead of 4 000 NOK. We can then assume that the bidder has a budget of 5 000 NOK to optimize his tender. Since the scheme reflects that the value of increasing the quality level by one, is worth less than it costs to provide (4 000 NOK versus 4 500 NOK), he provides a quality of 5 000 mm (which is a minimum requirement) and use the rest of his budget to reduce the price from 10 000 NOK to 5 000 NOK. However, if he knew that the real value of one level of quality was 5 000 NOK instead, he would have been better off by increasing the quality as it only costs him 4 500 NOK. Thus, the bidder would have offered 10 000 mm in water column (using 4 500 NOK of his budget) and a price of 9 500 NOK (using the last 500 of the budget to reduce his price). The first bid is worth 5 000 NOK for the procurer, while the last is worth 5 500 NOK (as the bidder are able to both provide a higher level of quality which is worth 5 000 NOK, and reduce the price by 500 NOK). The procurer, therefore, loses out on 5 500 – 5 000 = 500 NOK.

3.2 Evaluation of Price

In this section, we will present different scoring rules used to transform the price into a score. The price can be evaluated based on either the relative offer or in absolute terms (Dini et al., 2006, p. 304). A relative scoring rule can include the highest and/or the lowest bid as a base price, while an absolute scoring rule specifies benchmarks that are determined independently of the submitted bids. In this study, we have chosen to present four scoring rules. The first three are considered absolute scoring rules, while the last one is a relative rule:

- Linear rule with price threshold
- Linear rule without price threshold
- Parabolic rule
- Lowest bid scoring

As mentioned, national legislation require that evaluation methods are in accordance with certain basic principles. Dini et al. (2006, p. 314) assess different scoring rules based on four key features; *simplicity, predictability, competition* and *sensitivity to bid distribution*. Both *predictability* and *competition* are mentioned as two of the five basic principles in the Norwegian Procurement Act. It is therefore important to have an understanding on how the different scoring rules perform during evaluation with regards to these two principles. Furthermore, one can argue that both the *simplicity* and the *sensitivity to bid distribution* of a rule are important parts of the rule's predictability. Therefore, assessing how well the different scoring rules perform on these key features, has implications on to what extent they are in accordance with national law. We will in this chapter explain what is presented in present literature, and provide additional implications in the first part of the analysis in Chapter 6.

3.2.1 Absolute Scoring Rules

Absolute scoring rules do not compare and benchmark different bidder's price bids against each other when calculating the price score. Hence, they are not sensitive to bid distribution. Instead, they set some absolute benchmarks beforehand. Thus, it is possible to calculate the MVP before submitting a tender and the bidders are therefore able to optimize their tenders. These properties make the absolute scoring rules very predictable, since it reflects the preferences of the procurer and thereby the weight of price and quality. Therefore, the absolute rules are recommended in the literature (Dini et al., 2016, 304-315). With regards to simplicity and price competition, we will observe that there are some differences within this category of scoring rules.

3.2.1.1 Linear Rules

The scholars present two types of linear rules; one with a price threshold and one without a price threshold. First, we present the linear scoring rule with a price threshold as proposed by Dini et al. (2006, p. 305):

$$Price\ score\ = \mathrm{nn} * \frac{(\text{Reserve price} - \text{Price bid})}{(\text{Reserve price} - \text{Price threshold})}$$

In this formula, *nn* represents the maximum number of points available to be awarded to bidders for their price bids. We will use 4 points here, as this is the maximum points after weighting price 40 % (10*0.4 =4). The *reserve price* is defined as the highest bid allowed (Dini et al., 2006, p. 305). A price equal to and above the reserve price, will therefore lead to no points for the bidder. The *price threshold* indicates the lower limit for which the price bids are awarded points. A price equal to the price threshold will award the maximum amount of price points, and prices beneath this point will not lead to an improved price score for the bidder. The scoring rule used in the example in section 3.1 is a linear rule with a price threshold of 2000 NOK and a reserve price of 10 000 NOK.

The linear rule without a threshold is presented below. The rule now only awards the maximum score, nn, if the good is offered for free.

$$Price\ score\ = \mathrm{nn} * \frac{(\text{Reserve price} - \text{Price bid})}{\text{Reserve price}}$$

The graphs in Figure 3.2 below, present the price score as a function of the price bid for the two rules graphically. This is useful to understand the implications of the rules. First of all, it implies that the linear rule without a price threshold will award lower price score for all price bids than the rule with a price threshold. In addition, we observe the gap is quite big for the middle range of prices.

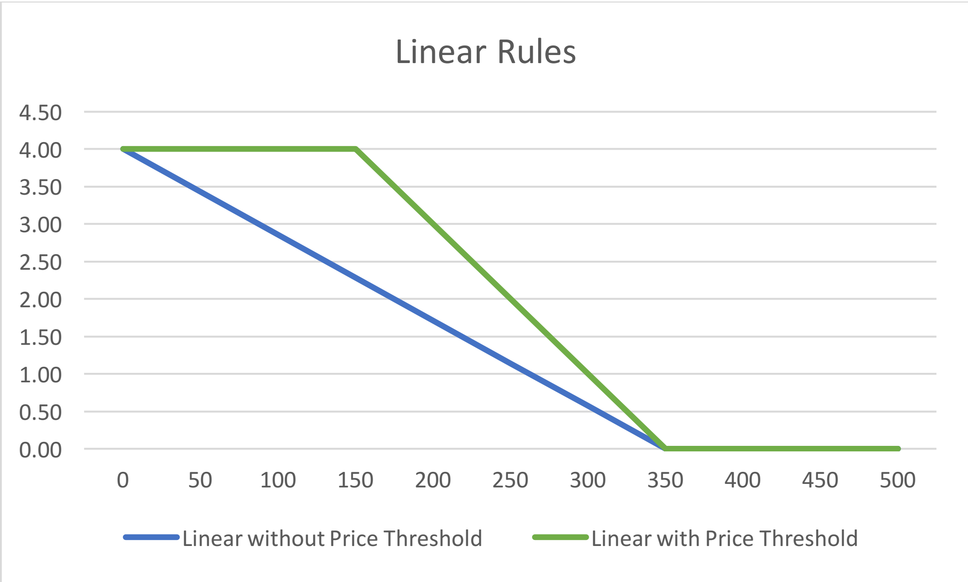

Figure 3.2: *Linear Rule with and without a Price Threshold*

For both versions of this rule, the MVP is constant, which makes both versions of the linear rule very simple. The MVP can be obtained from the following formula[4]:

$$MVP = \frac{(\text{Reserve price} - \text{Price threshold})}{\text{nn}}$$

However, the level of the MVP is quite different with the two rules. When using the linear rule with a price threshold, the MVP will be smaller compared to the rule without a price threshold. This is because the same amount of points is awarded along a smaller range of prices. Without a price threshold the MVP is 350 / 4 = 87.5 NOK per point, and with a price threshold the MVP is (350-150)/4 = 50 NOK per point. As a lower MVP makes it cheaper for the bidders to reduce

[4] Where price threshold = 0 for the rule without price threshold.

their prices, the linear rule with a price threshold increases the price competition among the bidders.

This also follows from the slope of the curves in Figure 3.2. The steeper the curve, the more points are awarded for a small change in price. Therefore, the steeper the slope, the more aggressive price competition is implied by the rule. Hence, the introduction of a price threshold increases the price competition between the bidders.

3.2.1.2 Parabolic Rule

The parabolic scoring rule is perceived as a bit more complicated. Here, the price score increases with lower bids, but at a diminishing rate (Dini et al., 2006, p. 307). Below, we present the parabolic rule proposed by Dini et al.:

$$Price\ score = nn * \left(1 - \left(\frac{Price\ bid}{Reserve\ price}\right)^2\right)$$

An important aspect of the parabolic rule is that it stimulates aggressive price bidding when price bids are close to the reserve price, while it does not incentivise further reduction for already low prices. This follows from the shape of the curve shown in Figure 3.3 below.

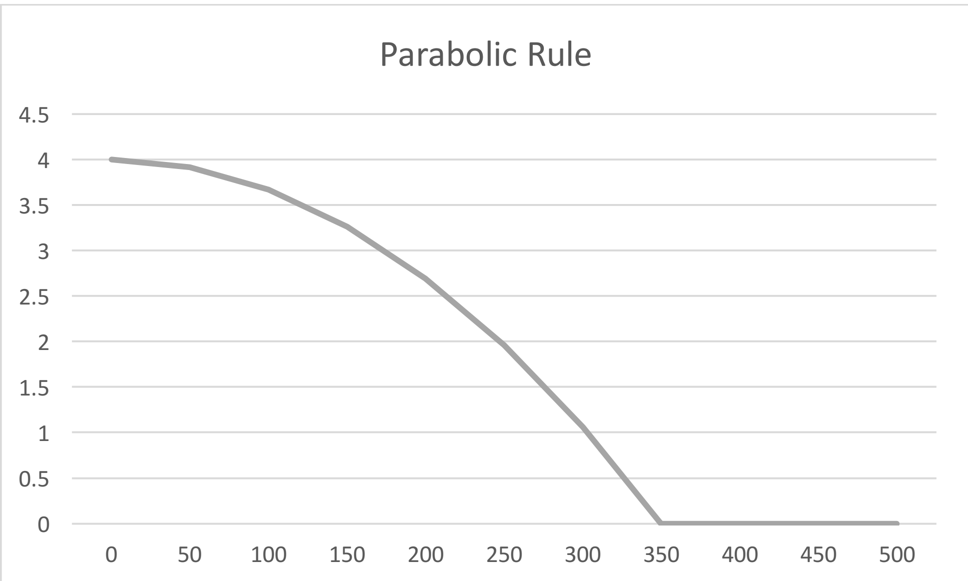

Figure 3.3: *Graph Parabolic Rule*

The curve is concave, being quite flat for the lower prices and steeper for the prices close to the reserve price. In fact, this makes the parabolic rule more alike the linear rule with a price threshold compared to the linear rule without a price threshold. Due to this non-linearity, the MVP is not constant and needs to be calculated for each price bid. However, it will, similarly to the linear rule with a price threshold, be lower for the higher price range and converge against infinity for the lowest prices where the curve is flat. Therefore, it stimulates more aggressive price competition in the area close to the reserve price.

3.2.1.3 Comparison between the Absolute Rules

The graphs of the three absolute rules are shown in Figure 3.4 below. We observe that the parabolic rule is quite similar to the linear rule with a price threshold. Both are awarding higher price scores partitioned on a smaller range compared to the linear rule without a price threshold. Therefore, they similarly stimulate to price competition for the relevant price range, while the linear rule without price threshold does this to a much smaller extent.

Figure 3.4: *Graphs Absolute Rules*

Furthermore, the two linear rules are simpler compared to the parabolic rule. This is due to the linearity and constant MVP for the two former rules. All together, we can understand why the linear rule with a price threshold is favoured in literature. It is highly predictable, not sensitive to bid distribution, simple to use and stimulates aggressive price competition.

3.2.2 Relative Scoring Rules

Common for relative scoring rules, are that the calculation of the price scores depends on the distribution of the other price bids submitted. Hence, they are all sensitive to bid distribution. Therefore, it is not possible to calculate the MVP in advance of the bidding phase, causing the procurers´ preferences to change depending on the bid distribution. Therefore, relative scoring rules are not predictable (Dini et al., 2006, p. 308). However, they also vary when it comes to simplicity and price competition.

Dini et al. (2006) presents three different types of relative scoring rules; average, highest bid-lowest bid and lowest bid. Only the latter will be included in our study, as this rule is recommended by the Norwegian Digitalisation Agency and is the most commonly used in practice.

3.2.2.1 Lowest Bid Rule

With the lowest bid rule, each bidder's price score is dependent on the lowest price bid submitted. Dini et al. (2006, p. 309) presents the formula of the lowest bid scoring rule as shown below:

$$Price\ score = nn * \frac{Lowest\ bid}{Price\ bid}$$

Since each bidder's price bids depend on the lowest price bid submitted, none of the bidders are able to calculate their price scores or the MVP in advance of submitting their bids. The lowest bid rule is therefore considered less predictable compared to the absolute rules. This also explains why this rule is so sensitive to the bid distribution. Thus, an abnormally low tender may change the ranking if rejected (Dini et al., 2006, p. 309). However, the rule is quite simple. It also provides the bidders with incentives of aggressive bidding, as the likelihood of receiving a high score increases when bidders offer a low price. In addition, submitting very low bids may at the same time reduce the other bidder's price scores.

3.3 Evaluation of Quality

The scoring rules presented in chapter 3.2, are not applicable when assigning a score to the quality aspects. Typically, the procurer uses expert panels to assess the quality offered in tenders. This panel review how well the offered quality performs with regards to the procurer's preferences, and award points accordingly. Points can be awarded by using an absolute or a relative scale (Ellingsen & Haukeli, 2020). By using an absolute scale, points are assigned based upon objective factors when reviewing the quality offered. Sykehusinnkjøp uses an absolute scale. Here, quality aspects in one tender are not compared to quality aspects in another tender. With a relative scale, however, points are awarded by relatively comparing the quality offered in the different tenders. The tender with the quality perceived as the best, is offered a maximum number of points and the other tenders are assigned points accordingly.

In all procurement, procurers need to make use of either an absolute or relative scale to assess the quality offered in the tenders. Still, there is a major difference between procurers on what they choose to do afterwards. Some procurers use the quality points awarded and weigh them according to the decided weighting function. The sum of these points constitutes the tender's total quality score. Other procurers choose to adjust, or normalise, the quality points before weighting them. The latter practice has been the subject of debate in recent months and will therefore be explained further.

Normalisation involves awarding the tender with the highest assigned quality points a maximum score. This implies that one of the tenders will receive a maximum score, even when none of the tenders are perceived by the experts as providing the maximum quality. The NDA recommends using normalisation when relative scoring rules are used in the evaluation method. Since the tender offering the lowest price is rewarded with the maximum number of price points, they recommend that the highest quality also is rewarded with the maximum quality points. They state this is important to ensure that the original weighting between price and quality remains the same (Avdeling for offentlige anskaffelser, 2019). This statement has been criticized by Ellingsen & Haukeli (2020), who claim that normalisation can arbitrarily change the weight between award criteria in evaluations.

They explain that there is a difference between performing an absolute normalisation and a relative normalisation. With absolute normalisation the procurer adjusts the quality points with the same absolute value, so the difference in quality points between tenders remains the same.

This is not a problem, but highly unnecessary, according to Ellingsen & Haukeli (2020). They provide an example to illustrate; imagine there is a competition consisting of six bidders, price and quality is given an equal weight in the evaluation (50 % each) and both are assessed on a scale from 0 to 10. The bidder offering the lowest price is awarded 10 points, while the others are evaluated relatively (lowest bid rule). Quality is evaluated on the basis of an absolute scale. The resulting scores are the following:

Bidder	*Quality (50 %)*	*Price (50 %)*	*Total score*
Bidder A	5	3	4
Bidder B	4	4	4
Bidder C	3	6	4.5
Bidder D	2	5	3.5
Bidder E	**1**	**10**	**5.5**
Bidder F	0	9	4.5

Table 3.2: *Absolute Scale without Normalisation. Own illustration, source: Ellingsen & Haukeli (2020)*

None of the bidders provide a quality that is perceived as especially high. However, Ellingsen & Haukeli (2020), show that an absolute normalisation does not affect the outcome of the competition, as the difference between the quality points remain the same:

Bidder	*Quality (50 %)*	*Price (50 %)*	*Total score*
Bidder A	10	3	6.5
Bidder B	9	4	6.5
Bidder C	8	6	7
Bidder D	7	5	6
Bidder E	**6**	**10**	**8**
Bidder F	5	9	7

Table 3.3: *Absolute Scale with Absolute Normalisation. Own illustration, source: Ellingsen & Haukeli (2020)*

In this example, the absolute normalisation involves adjusting all quality scores with 5 points. We observe that the weighting between price and quality remain unchanged and the outcome is the same. Hence, absolute normalisation has no function and is unnecessary. This has an important implication: The weight is not affected by where the quality points are located on the scale. It is the difference in the quality points between the bidders that is decisive. The same happens if the original quality scores are given by a relative scale and an absolute normalisation is performed.

Relative normalisation, however, involves adjusting bidders´ quality score according to the following formula:

$$10 * \frac{Quality\ score\ tender\ x}{Quality\ score\ best\ tender}$$

This type of normalisation changes the difference in quality points between tenders, both when an absolute and a relative scale is used for assigning the original score. This is problematic. Using the same example as above, with initially having absolute scores as in Table 3.3, the scores after a relative normalisation is shown in Table 3.4.

Bidder	*Quality (50 %)*	*Price (50 %)*	*Total score*
Bidder A	10	3	6.5
Bidder B	8	4	6
Bidder C	6	6	6
Bidder D	4	5	4.5
Bidder E	**2**	**10**	**6**
Bidder F	0	9	4.5

Table 3.4: *Absolute Scale with Relative Normalisation. Own illustration, source: Ellingsen & Haukeli (2020)*

Comparing the total score in Table 3.4 with those originally given in Table 3.3, we observe how relative normalisation may change the outcome of a competition in an arbitrary way. In this example, bidder E wins without normalisation, while bidder A wins with relative normalisation. As the difference in total score between the bidders has changed, the weight of quality change. If the procurer does not inform the bidders that they are performing a relative normalisation, Ellingsen & Haukeli (2020) claim this would be contrary to the basic principles of predictability and equal treatment, thereby under certain circumstances being illegal.

Furthermore, Ellingsen & Haukeli (2020) do not recommend relative normalisation if the experts have already assessed the quality of the tenders by using a relative scale. They explain this by presenting two different cases where the procurer receives three tenders.

Tender	*Quality points*	*Difference in quality points*	*Quality points after relative normalisation*	*Difference in quality points*
Tender 1	4		6.7	
Tender 2	5	1	8.3	1.7
Tender 3	6	1	10	1.7

Table 3.5: *Case 1 - Relative Scale with Relative Normalisation. Own illustration, source: Ellingsen & Haukeli (2020)*

If the experts perceive the quality offered in tenders 1-3 as being 4, 5 and 6, respectively, this imply that the procurer (and the experts) means that this difference in quality points reflects the relative differences in quality between them. However, if the procurer chooses to normalise the points relatively, the differences in points increases to 1.7.[5] One can therefore pose the question of what price difference is really reflecting the quality differences between the tenders? Is it 1 or 1.7? An even more interesting question is why didn't the procurer (or experts) award quality points with a relative difference of 1.7 in the first time? Performing a relative normalisation after experts have already assessed and compared the quality, is therefore slightly confusing. In this example, however, the difference ***between*** the tenders are the same and does not change the winner.

[5] It is important to recognise that it is only the absolute difference in quality points that changes in this case, not the relative difference.

However, this is not always the case. If the tenders had been evaluated differently, the difference in quality points between the bidders can also change using relative normalisation. This can be illustrated in Table 3.6 by a second case:

Tender	*Quality points*	*Difference in quality points*	*Quality points after relative normalisation*	*Difference in quality points*
Tender 1	4		5	
Tender 2	5	1	6.25	1.25
Tender 3	8	3	10	2

Table 3.6: *Case 2 - Relative Scale with Relative Normalisation. Own illustration, source: Ellingsen & Haukeli (2020)*

The only change from Table 3.5 to this second case, is that tender 3 is awarded 8 points instead of 6. Nevertheless, the differences in quality points between the bidders have now changed. According to Ellingsen & Haukeli (2020), this appears to be arbitrary. Instead of using relative normalisation, they advise the procurer to rather ensure that the quality points awarded initially reflects the relative quality difference between the tenders, e.g. use a relative scale.

The arguments provided by Ellingsen & Haukeli (2020), is valuable to keep in mind when we in later chapters will evaluate what impact different evaluation methods may have on the outcome of procurements.

4 Simulation of Data

The purpose of this study is to investigate how different evaluation methods affect the preferences of the procurer and, subsequently, may change the outcome of procurements. In order to do this, we have used R Studio to make an algorithm for simulating our data. More specifically, we are simulating the outcome of a procurement using the different evaluation methods presented in Chapter 3. Our aim is to provide general evidence of the more specific examples provided earlier in literature and debates. In this chapter, we will first present the model environment, then explain the reasoning behind the model setup, before describing some of the limitations regarding our approach. The algorithms are presented in Appendix A4 and the files can be provided upon request.

4.1 The Model Environment

Our model environment is a reverse auction environment, where the auction is about procuring, rather than selling a good or a contract. There is only one procurer. In practice, the number of bidders will vary among auctions. However, to be able to perform our study and analyse the methods within a reasonable scale, there are only two bidders in our model.

Moreover, price-to-quality scoring involves the evaluation of both a price and quality criteria. Typically, a tender is evaluated on the basis of several quality criteria and prices are in practice continuous. Quality and price may therefore form millions of combinations. However, for the simplicity of this analysis, we have chosen that the model environment only consists of two quality criteria and five different prices (100, 200, 300, 400, 500).[6] We have chosen to use an absolute quality scale, ranging from 0 to 10, where 10 is the best quality and 0 indicates that the tenders do not offer more than the minimum requirement required to participate in the procurement.

Our model environment consists of two bidders, whom each offer one price and a degree of quality on two types of quality criteria. We will present the tenders as the following: (200,3,8), where 200 represents the price, and 3 and 8 are the original score that the bidder is assigned for quality criteria one and two, respectively. These parameters give us a total of 605 possible

[6] These numbers could be in thousand or million and in NOK or any other currency. We will only refer to the numbers without a currency, for easier reading. However, it does not change the interpretation of our results.

combinations of tenders and 366 025 (605 * 605) possible bid combinations between the two bidders.[7]

In the next step, different evaluation methods are included in the model environment. All scoring rules are modelled in accordance with the formulas presented in Chapter 3. We have chosen to investigate five different evaluation methods, which are presented in Table 4.1 below. For simplicity, we have used numbers instead of their names in the algorithm.

Method	Name
1	Linear rule with a price threshold, no normalisation
2	Linear rule without a price threshold, no normalisation
3	Parabolic rule, no normalisation
4	Lowest bid rule without normalisation
5	Lowest bid with normalisation

Table 4.1: *Overview of Different Evaluation Methods*

We have chosen not to normalise the quality score when using the three absolute scoring rules, as normalisation is only a topic when using relative scoring rules. In addition, this enable us to investigate the differences due to different scoring rules later in our analysis. Furthermore, we have chosen to simulate the lowest bid rule without normalisation and with a relative normalisation. The reason for this is that an absolute normalisation will provide the same results as without. As we want to investigate whether the ongoing criticism of common practice, namely method 5, is reasonable, this division is practical. In the rest of the thesis, we will refer to relative normalisation only as normalisation.

Furthermore, we need to define the weights that are used for assessing the price and the two quality criteria. Since the scale is the same for all three dimensions, ranging from 0 to 10, they will have equal weight if we keep the scores as they are. In most cases, Sykehusinnkjøp operates with a 40 % weighting of price and a 60 % weighting of the quality dimensions in total. We have therefore chosen to do the same, meaning that the price score will be weighted by 40 %, and the two quality dimensions have an equal weight of 30 % each.

[7] Half of the combinations are the same but opposite when it comes to what bidder offer what tender.

For the absolute rules, we have chosen a reserve price of 350 and a price threshold of 150.[8]

4.2 The Simulation and its Output

To compare the different methods, we have made three types of matrices which all were exported to Microsoft Excel and analysed further. The row and column names represent the combinations that bidder 1 and 2 could offer, making each cell one bid combination. The matrices are therefore symmetric along the diagonal.

The first matrix we named "score matrix", as this shows the total scores of the two bidders for different combinations of price and quality. We then made one matrix for each evaluation method. Table 4.2 below shows part of the score matrix for the lowest bid rule with normalisation. As an example, we observe that cell B13 contains the vector (7, 4). This means that when bidder 1 has offered the bid (100,1,0) and bidder 2 has offered (100,0,0). Bidder 1 gets a total weighted score of 7 while bidder 2 get a total weighted score of 4.

	A	B	C	D	E	F	G	H
1		(100, 0, 0)	(100, 0, 1)	(100, 0, 2)	(100, 0, 3)	(100, 0, 4)	(100, 0, 5)	(100, 0, 6)
13	(100, 1, 0)	(7, 4)	(7, 7)	(7, 7)	(7, 7)	(7, 7)	(7, 7)	(7, 7)
14	(100, 1, 1)	(10, 4)	(10, 7)	(8.5, 7)	(8, 7)	(7.75, 7)	(7.6, 7)	(7.5, 7)
15	(100, 1, 2)	(10, 4)	(10, 5.5)	(10, 7)	(9, 7)	(8.5, 7)	(8.2, 7)	(8, 7)
16	(100, 1, 3)	(10, 4)	(10, 5)	(10, 6)	(10, 7)	(9.25, 7)	(8.8, 7)	(8.5, 7)
17	(100, 1, 4)	(10, 4)	(10, 4.75)	(10, 5.5)	(10, 6.25)	(10, 7)	(9.4, 7)	(9, 7)
18	(100, 1, 5)	(10, 4)	(10, 4.6)	(10, 5.2)	(10, 5.8)	(10, 6.4)	(10, 7)	(9.5, 7)
19	(100, 1, 6)	(10, 4)	(10, 4.5)	(10, 5)	(10, 5.5)	(10, 6)	(10, 6.5)	(10, 7)
20	(100, 1, 7)	(10, 4)	(10, 4.43)	(10, 4.86)	(10, 5.29)	(10, 5.71)	(10, 6.14)	(10, 6.57)
21	(100, 1, 8)	(10, 4)	(10, 4.38)	(10, 4.75)	(10, 5.12)	(10, 5.5)	(10, 5.88)	(10, 6.25)
22	(100, 1, 9)	(10, 4)	(10, 4.33)	(10, 4.67)	(10, 5)	(10, 5.33)	(10, 5.67)	(10, 6)
23	(100, 1, 10)	(10, 4)	(10, 4.3)	(10, 4.6)	(10, 4.9)	(10, 5.2)	(10, 5.5)	(10, 5.8)

Table 4.2: *Score Matrix*

We also made a second type of matrix; the "rank matrix". We have one rank matrix for each evaluation method, showing the ranking between the two bidders for each combination of bids instead of the total scores. Table 4.3 below, shows the part of the rank matrix corresponding to the score matrix above. The bidder with the highest score gets the value 1, and the other bidder gets a value of 2. If the scores are equal, both get the number 1.5 (average of 1 and 2).

[8] The algorithm is coded dynamically, so all the parameters can easily be changed. The algorithm can be provided by request.

	A	B	C	D	E	F	G	H
1		(100, 0, 0)	(100, 0, 1)	(100, 0, 2)	(100, 0, 3)	(100, 0, 4)	(100, 0, 5)	(100, 0, 6)
13	(100, 1, 0)	(1, 2)	(1.5, 1.5)	(1.5, 1.5)	(1.5, 1.5)	(1.5, 1.5)	(1.5, 1.5)	(1.5, 1.5)
14	(100, 1, 1)	(1, 2)	(1, 2)	(1, 2)	(1, 2)	(1, 2)	(1, 2)	(1, 2)
15	(100, 1, 2)	(1, 2)	(1, 2)	(1, 2)	(1, 2)	(1, 2)	(1, 2)	(1, 2)
16	(100, 1, 3)	(1, 2)	(1, 2)	(1, 2)	(1, 2)	(1, 2)	(1, 2)	(1, 2)
17	(100, 1, 4)	(1, 2)	(1, 2)	(1, 2)	(1, 2)	(1, 2)	(1, 2)	(1, 2)
18	(100, 1, 5)	(1, 2)	(1, 2)	(1, 2)	(1, 2)	(1, 2)	(1, 2)	(1, 2)
19	(100, 1, 6)	(1, 2)	(1, 2)	(1, 2)	(1, 2)	(1, 2)	(1, 2)	(1, 2)
20	(100, 1, 7)	(1, 2)	(1, 2)	(1, 2)	(1, 2)	(1, 2)	(1, 2)	(1, 2)
21	(100, 1, 8)	(1, 2)	(1, 2)	(1, 2)	(1, 2)	(1, 2)	(1, 2)	(1, 2)
22	(100, 1, 9)	(1, 2)	(1, 2)	(1, 2)	(1, 2)	(1, 2)	(1, 2)	(1, 2)
23	(100, 1, 10)	(1, 2)	(1, 2)	(1, 2)	(1, 2)	(1, 2)	(1, 2)	(1, 2)

Table 4.3: *Rank Matrix*

Thereafter, we used the rank matrices to compare the different evaluation methods. By comparing the ranking inside each cell between two rank-matrices, we made a third type of matrix; "the 0/1-matrix" for each pair of evaluation methods. This matrix shows the value 1 for the bid combinations where the ranking between the bidders are different, and 0 if they are equal. This matrix is used to analyse the pattern of where the different evaluation methods provide different winners of the procurement. A part of the 0/1-matrix, between the lowest bid rule with and without normalisation, is shown in Table 4.4 below.

	A	B	C	D	E	F	G	H
1		(100, 0, 0)	(100, 0, 1)	(100, 0, 2)	(100, 0, 3)	(100, 0, 4)	(100, 0, 5)	(100, 0, 6)
13	(100, 1, 0)	0	0	1	1	1	1	1
14	(100, 1, 1)	0	0	1	1	1	1	1
15	(100, 1, 2)	0	0	0	1	1	1	1
16	(100, 1, 3)	0	0	0	0	1	1	1
17	(100, 1, 4)	0	0	0	0	0	1	1
18	(100, 1, 5)	0	0	0	0	0	0	1
19	(100, 1, 6)	0	0	0	0	0	0	0
20	(100, 1, 7)	0	0	0	0	0	0	0
21	(100, 1, 8)	0	0	0	0	0	0	0
22	(100, 1, 9)	0	0	0	0	0	0	0
23	(100, 1, 10)	0	0	0	0	0	0	0

Table 4.4: *0/1 Matrix*

4.3 Limitations

There are mainly two factors making our model environment not as realistic as in practice. The first is the number of bidders included, while the second is the possible prices the bidders can offer. In collaboration with Sykehusinnkjøp, we were given access to data on several procurements, most of them consisting of more bidders, several quality dimensions and very different price ranges. In our simplified version of a procurement made by Sykehusinnkjøp in Chapter 3, there were actually about 30 quality dimensions and four different bidders.

We do not scale up the model environment due to the complexity and limitations of the programs we used. First, R Studio is using a very long time processing the data when we make the combination of bids larger. However, the simulation is possible to conduct and the algorithm provided in Appendix A4, can easily be adapted to include several prices, quality dimensions or other ranges.

Secondly, Microsoft Excel is, even with this data set, struggling with handling the matrices by shutting down at regular intervals. With more computer power this limitation might be possible to overcome. However, if the model environment and the matrices are scaled up, the analysis will be even more challenging to conduct and not give further results. Our purpose is to observe how different evaluation methods can provide different rankings for the same bids. If we observe different rankings for this dataset, we therefore argue that it will be transferable at a scaled up data set with more bidders and more price combinations.

5 Logistic Regression

We use regression analysis to investigate if our findings can be proven statistically. Our aim is to verify that certain variables, e.g. certain characteristics of the tenders and the combinations of tenders in the procurement, statistically impact who is the winner of procurement contracts, when comparing two different evaluation methods. We will in this chapter shortly present the logistic regression and how to interpret the results from such regressions.

The logistic regression is the most suitable method for our purpose. This is due to our binary response variables, which takes the value of 0 or 1. In the following, we will explain the concept of logistic regression and see how it is compared to the more well-known linear regression.

When the response variable is binary, we are predicting the probability of Y=1, given X. This can be written as p(Y=1 | X). For simplicity, let p(X) = p(Y=1 | X) (James et al, 2013, p. 131). When using linear regression[9], we then have

$$p(X) = \beta_0 + \beta_1 \tag{1}$$

were β_1 is interpreted as the average change in Y associated with a one unit change in X. If X is a dummy variable, which all of our variables are, β_1 is the change in probability of Y=1, when the dummy variable X is 1. A drawback with linear regression is that it may provide p(X) < 0 and p(X) > 1, which is not very sensible (James et al, 2013, p. 132). This can be observed in the left-hand panel of Figure 5.1.

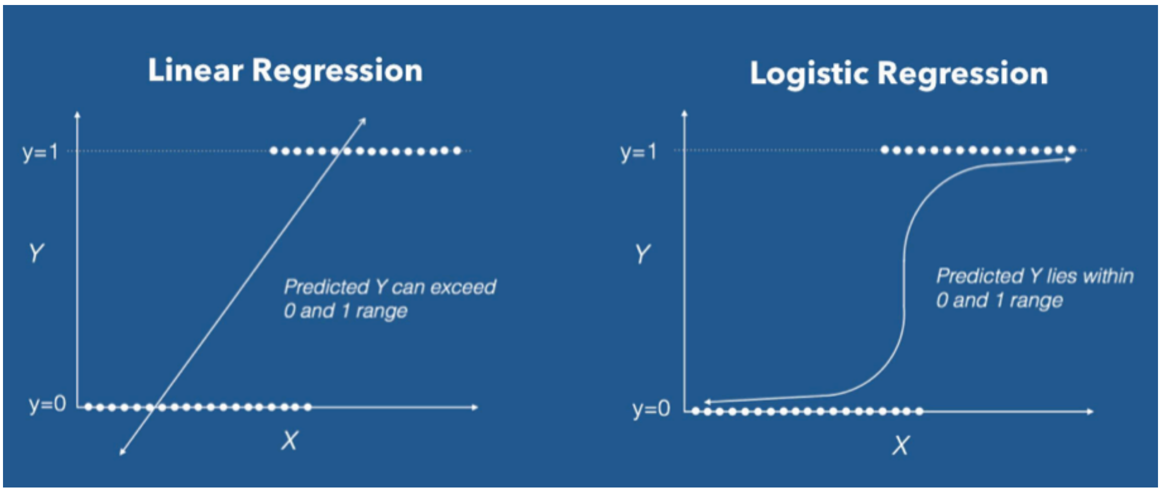

Figure 5.1: *Linear Regression vs. Logistic Regression, Source: Le, (2018).*

[9] The Linear Probability Model (LPM) with an Ordinary Least Square (OLS) estimator.

With logistic regression, however, we avoid this problem by using the logistic function

$$p(X) = \frac{e^{\beta_0+\beta_1 X}}{1+\ e^{\beta_0+\beta_1 X}} \tag{2}$$

This logistic function is a Sigmoid function, forming a S-shaped curve as illustrated in the right-hand panel of Figure 5.1. By using this function, we ensure a sensible prediction, taking values in the range between 0 and 1, regardless of the value of the predictors (James et al, 2013, p. 132).

By manipulating this equation, taking the logarithm of both sides, we obtain the logit or log-odds

$$\ln\left(\frac{p(x)}{1-p(x)}\right) = \ \beta_0 + \beta_1 X \tag{3}$$

The left-hand side of the equation is called the log-odds, or logit, and is linear in the predictors (James et al, 2013, p. 132). Thus, we can now interpret β_1 directly, as being the change in log-odds when changing X by one unit, or for a dummy when X=1. The term, odds, is often used in horse racing and reflects the likelihood that an event will occur. It is the ratio of success to non-success. As an example, if the probability of winning a race is 20 %, you have the odds of ¼ of winning (James et al, 2013, p. 132). However, we will mostly use the probability of Y=1, given X in our analysis.

Still, the reason for presenting the log-odds is due to the importance of not mixing this interpretation with the commonly used linear regression. The beta coefficients from a logistic regression are not interpreted as change in p(X) as with linear regression. With logistic regression the amount that p(X) changes, due to a one unit change in X, will depend on the current value of X (James et al, 2013, p. 132-133). Nevertheless, regardless of the value of X, we can interpret the direction of the impact on Y directly. If β_1 is positive, then an increase in X will be associated with an increase in p(X). Therefore, a negative β_1 can be interpreted as a decrease in p(X) (James et al, 2013, p. 132-133). This will be the most important feature for our purpose.

The coefficients β_0 and β_1 are with logistic regression, estimated by using the maximum likelihood technique. In short, the method seeks to estimate the coefficients such that the resulting probabilities are closest to either 1 or 0. This intuition can be formalized as the likelihood function (James et al, 2013, p. 133)[10]:

$$\ell(\beta_0, \beta_1) = \prod_{i:y_i=1} p(x_i) \prod_{i':y_{i'}=0} 1 - p(x_{i'}) \qquad (4)$$

[10] To run logistic regression in R, the function glm(), with argument "family =binomial", is used.

6 Analysis

In this chapter, we will study the impact of using different evaluation methods. First, we will present an overview of the percentage amount of bid combinations resulting in different winners, when pairwise comparing the different evaluation methods. Based on these findings, we have chosen to divide our further analysis into two parts. First, we will examine the differences between the scoring rules from a new angle, investigating what the differences really imply for the weighting of price and quality. In the last part, we will look further into the implications of using normalisation during the quality evaluation.

6.1 Overview of Different Outcomes

Figure 6.1 below, present an overview of the percentage amounts of bid combinations that results in different winners between the evaluation methods compared.[11] For a better visualisation of the overview, we will refer to the linear rule with a price threshold as linear rule 1, the linear rule without a price threshold as linear rule 2, and lowest bid rule without normalisation as just lowest bid.

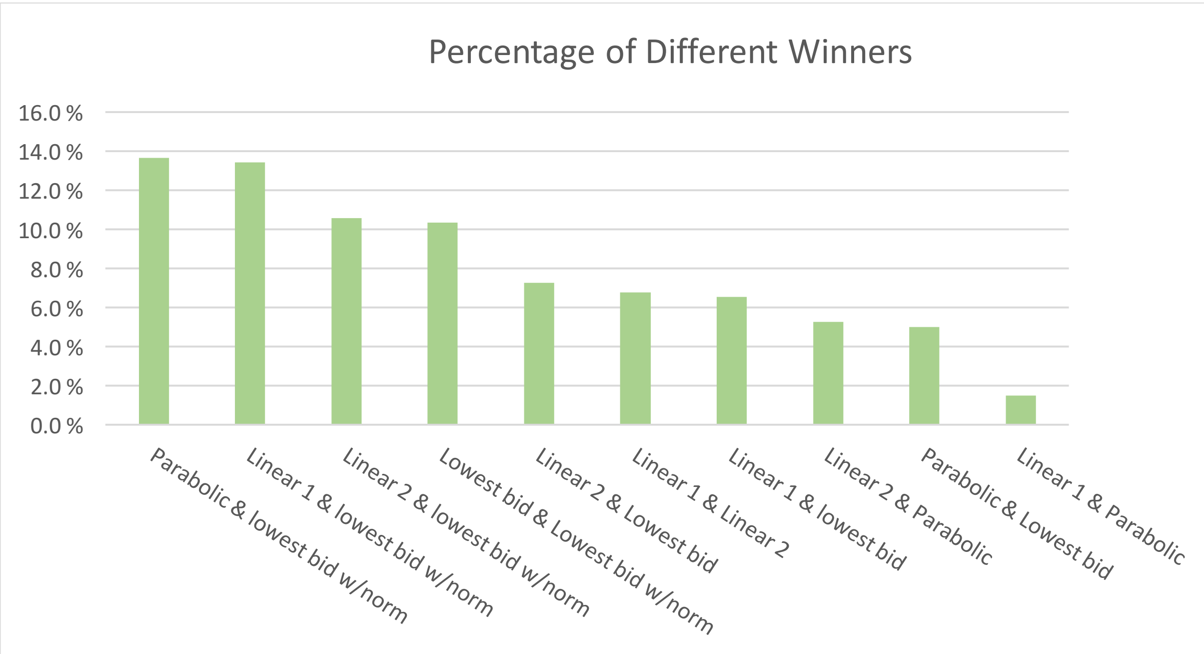

Figure 6.1: *Percentage of Different Winners*

[11] In other words, we have counted the amount of 1's occurring in the 0/1-matrices in our data.

From Chapter 3, we know that different scoring rules reflect different procurer preferences. Therefore, we expect there to be some difference between evaluation methods that use different scoring rules. We especially expect there to be differences between the two main groups of scoring rules, absolute and relative, as they are quite different with regards to predictability and sensitivity to bid distribution. Hence, the relatively high percentage of difference between all the absolute rules (linear rule 1, linear rule 2 and the parabolic rule) and the lowest bid rule are very well expected.

However, what might be surprising is the relatively large amount of differences between some of the absolute scoring rules. We observe that the differences between the linear rules (1 and 2) and the difference between the parabolic rule and linear rule 2 are about the same size as the comparisons between the absolute and relative rules. In fact, the difference of 6.8 % between the two linear rules, is higher than between linear rule 1 and lowest bid rule, which is 6.5 %. We ran a regression to test if the differences are large enough to be perceived as statistically different. The results are shown in Table A1.2 in Appendix A1, and shows that the difference between the linear rules are in fact significantly higher. However, the coefficient is small, indicating a small difference.

Furthermore, what might be surprising is that it is the two linear rules among the absolute scoring rules that differ the most. In comparison, the parabolic rule and the linear rule with a price threshold is very much alike. We have elaborated on this in Chapter 3, but the literature does often explain the two linear rules altogether, as if they were more alike (Dini) et al., 2006, p. 305). Therefore, it might be surprising that the differences within the family of linear rules are of such a size. By testing the difference between the two pairwise comparisons, we find that the odds of getting different winners are 1.55 higher in the comparison between the linear rules, than for the comparison of linear rule 1 and the parabolic rule.[12] This implies that the use of a price threshold has a great impact on the outcome of the procurement.

Furthermore, the most striking observation is that the three combinations with the highest percentage of different outcomes (winners), are comparisons between methods using an absolute scoring rule without normalisation and a relative scoring rule with normalisation. In other words, both the use of different scoring rules as well as differences in the way of evaluating the quality, results in different outcomes. For these comparisons, the percentage of

[12] Results are provided in Appendix A1, Table A1.2.

different winners ranges from 10.6 % to 13.7 %. What is striking, is that the percentage of different winners when comparing an absolute rule and a relative rule (expect from the combinations including linear rule 2), is at least doubled when the relative rule also normalises the quality score. As an example, we have 6.5 % outcomes with different winners when comparing linear rule 1 and the lowest bid rule, while the percentage is 13.4 % when comparing the linear rule 1 and lowest bid with normalisation. The difference between these pairwise comparisons are also statistically significant. See Appendix A1 Tables A1.3-A1.5. This is observed more clearly when comparing the lowest bid rule with and without normalisation, where the only difference is the normalisation. The combination has the fourth highest percentage, with 10.3 % of the bid combinations providing different results. This implies that normalisation has a great impact on who the winner of the procurement is.

Based on these observations we will divide our further analysis into two parts. In the first part, we will investigate evaluation methods that only differ with regards to the scoring rule being used. We do this to achieve a deeper understanding of why they provide such varying results. Our aim is to make the procurers more aware of the differences between scoring rules. In the last part of our analysis, we will investigate the differences occurring due to difference in quality evaluation, that is, using normalisation or not. We will therefore look further into the differences occurring between the lowest bid rule with and without normalisation. This will provide insights on for what bid combinations normalisation impacts the results the most, and hopefully clarify some of the ongoing discussion regarding this. Our aim of both analysis is to provide a better understanding regarding the differences between evaluation methods, making procurers able to make more informed decisions on what method to use.

6.2 Implications of Different Scoring Rules

In chapter 6.1, we observed a relatively high degree of differences occurring between the absolute rules in those comparisons where linear rule 2 were involved. We also observed differences among the absolute rules and the relative rule. Therefore, we want to investigate these differences further. In the first section, we will examine how the scoring rules differ with regards to the *weight* of price and quality. Thereafter, we will look into *when* the scoring rules change the winner of the procurement.

6.2.1 Different Scoring Rules Assign Price Different Weights

In Table 6.1 below, we present the weighted price scores of two bidders for different price combinations, using the four different scoring rules. When we now consider different bid combinations, we are also able to calculate the price score given by the lowest bid rule. The column “diff bidders” shows the difference in price score between the two bidders for each bid combination for a specific rule. For instance, if both bidders offer a price of 100 and are evaluated by the linear rule 1, they both get 4 points and “diff bidders” are zero. This trait is observed among all scoring rules; If both bidders offer the same price, they will be awarded the exact same price score.

If they do not offer the same prices, they are assigned different price scores. Then, the lowest price always gets the highest price score. This is a common trait of all the scoring rules, as procurers generally appreciate a low price. In this study, however, we investigate evaluation methods that also assess quality. Therefore, it is typically a trade-off between high quality and higher prices, and lower prices and lower quality. But how is this trade-off when using different scoring rules? We will see that for some scoring rules, a higher price is valued more than by others.

Different price combinations from the bidders		Linear rule with price threshold (1)			Linear rule without price threshold (2)		
Price bidder 1	Price bidder 2	Score bidder 1	Score bidder 2	**Diff bidders**	Score bidder 1	Score bidder 2	**Diff bidders**
100	100	4	4	**0**	2.86	2.86	**0**
200	100	3	4	**1**	1.71	2.86	**1.14**
300	100	1	4	**3**	0.57	2.86	**2.29**
400	100	0	4	**4**	0.00	2.86	**2.86**
500	100	0	4	**4**	0.00	2.86	**2.86**
300	200	1	3	**2**	0.57	1.71	**1.14**
400	200	0	3	**3**	0.00	1.71	**1.71**
500	200	0	3	**3**	0.00	1.71	**1.71**
400	300	0	1	**1**	0.00	0.57	**0.57**
500	300	0	1	**1**	0.00	0.57	**0.57**
500	400	0	0	**0**	0.00	0.00	**0**
Different price combinations from the bidders		**Parabolic Rule (3)**			**Lowest bid rule, without normalization (4)**		
Price bidder 1	Price bidder 2	Score bidder 1	Score bidder 2	**Diff bidders**	Score bidder 1	Score bidder 2	**Diff bidders**
100	100	3.67	3.67	**0**	4	4	**0**
200	100	2.69	3.67	**0.98**	2	4	**2**
300	100	1.06	3.67	**2.61**	1.33	4	**2.67**
400	100	0.00	3.67	**3.67**	1	4	**3**
500	100	0.00	3.67	**3.67**	0.8	4	**3.20**
300	200	1.06	2.69	**1.63**	2.67	4	**1.33**
400	200	0.00	2.69	**2.69**	2	4	**2**
500	200	0.00	2.69	**2.69**	1.6	4	**2.40**
400	300	0.00	1.06	**1.06**	3	4	**1**
500	300	0.00	1.06	**1.06**	2.4	4	**1.60**
500	400	0.00	0.00	**0**	3.2	4	**0.80**

Table 6.1: *Price Scores and Difference in Price Scores for Different Price Combinations and Different Scoring Rules*

The difference between the price points awarded the two bidders, reflects how much higher the quality score of the bidder with the highest price (the lowest price score) needs to be if his tender is to be assessed as equally good as the tender with the lowest price (and highest price score). The higher the difference in price score, the more quality points are needed from the bidder with the highest price in order to win. We illustrate this by looking at a numeric example from Table 6.1. Here, the linear rule with a price threshold (linear rule 1) is applied, and bidder 1 and bidder 2 offer the prices of 400 and 100, respectively. Bidder 1 will get 0 points as the price is above the reserve price, while bidder 2 gets 4 points as the price is below the price threshold (10 points weighted 40 % = 4 points). The difference in price scores between the bidders are then 4 points. Therefore, it is required that bidder 1 at least offers a quality that assigns him more than 4 quality points (after weighting) in order to win. Since quality is weighted 60 % in total and 30 % on each dimension, one point on one of the dimensions is equal to 0.3 point weighted. He must therefore have an original quality score of 14 points more than bidder 2 in order to win (14 * 30% = 4.2 quality points). This is quite a big difference in quality. However, the difference in prices point is as large as it can be, as one is above the reserve price and the other below the price threshold. Therefore, it makes sense that bidder 1 would have to compensate quite a lot on the quality dimension for the price to be acceptable.

For the same price combinations, different scoring rules result in different price scores. If we look at the same price combination as above (400 and 100), but uses the lowest bid rule, the difference in price points between the bidders are 3 points instead of 4. This implies that for the scoring rule providing the lowest difference in price points (lowest bid in this example), the bidder with the highest price (and lowest price score) will to a larger extent be able to offset his high price (and low price score) by offering quality. In other words; With the same price, the bidder with the highest price needs fewer quality points to win with the lowest bid rule compared to the linear rule with a price threshold. Therefore, when comparing different scoring rules, ***the scoring rule with the lowest difference in price points among the bidders, are actually weighting price relatively more and quality relatively less*** than the other rules.

Table 6.2 below, compares the scoring rules pairwise. It shows which of the scoring rules that result in the lowest difference in price points between the bidders for each possible bid combination. The colours indicate the different rules, and will also be used in tables presented later in this chapter.

Different price combinations from the		Linear 1 & Linear 2	Linear 1 & Parabolic	Linear 2 & Parabolic	Linear 1 & Lowest bid	Linear 2 & Lowest bid	Parabolic & Lowest bid
Price bidder 1	Price bidder 2	Rule with lowest difference in score between the bidders in pairwise comparison					
100	100	Equal	Equal	Equal	Equal	Equal	Equal
200	100	Linear 1	Parabolic	Parabolic	Parabolic	Linear 2	Parabolic
300	100	Linear 2	Parabolic	Linear 2	Lowest bid	Linear 2	Parabolic
400	100	Linear 2	Parabolic	Linear 2	Lowest bid	Linear 2	Lowest bid
500	100	Linear 2	Parabolic	Linear 2	Lowest bid	Linear 2	Lowest bid
300	200	Linear 2	Parabolic	Linear 2	Lowest bid	Linear 2	Lowest bid
400	200	Linear 2	Parabolic	Linear 2	Lowest bid	Linear 2	Lowest bid
500	200	Linear 2	Parabolic	Linear 2	Lowest bid	Linear 2	Lowest bid
400	300	Linear 2	Linear 1	Linear 2	Equal	Linear 2	Lowest bid
500	300	Linear 2	Linear 1	Linear 2	Parabolic	Linear 2	Parabolic
500	400	Equal	Equal	Equal	Parabolic	Linear 2	Parabolic

Table 6.2: *Rules Compared with Regards to the Difference in Price Score between Bidders*

We observe that for almost all price combinations, the linear rule 2 results in the lowest difference in price points compared to the other rules. As this rule results in the overall lowest price scores[13], it might lead us to believe that this rule value price the least, relatively to the other rules. However, the opposite is actually the case. As the linear rule 2 results in lower ***differences*** in price points for the same prices, relatively less quality needs to be offered by the bidder with the highest price (and the lowest price score) to offset his higher price. This implies that with the linear rule without a price threshold, price is actually weighted relatively more and quality relatively less, than the other scoring rules.

Furthermore, when comparing the linear rule 1 and the parabolic rule against the lowest bid rule, we observe that for most of the price combinations, the lowest bid rule results in the lowest difference in price scores between the bidders.[14] Therefore, the bidder with the highest price, will favour the lowest bid rule as it takes less quality to win with a higher price compared to if the bidder is evaluated using one of the other two rules (the linear rule 1 and the parabolic rule). In other words, even though the bidders get high price points with the linear rule 1 and the parabolic rule, the difference between their price scores are also big. With these rules, the bidder with the highest price (and lowest price score) is therefore forced to offer a relatively higher degree of quality in order to win.

To sum up, the linear rule without a price threshold and thereafter the lowest bid rule, provide the lowest differences in price scores between the bidders. Hence, it is therefore easier for the

[13] Observed in Figure 6.1.
[14] When comparing linear rule 1 and the parabolic rule we observe that the latter has the lowest difference in price scores. However, for these two rules the difference between the difference in price score, are very small. Therefore, there are only a few bid combinations that results in different winners among the two rules.

bidder with the highest price to win when offering a low quality, than it is when using the linear rule with a price threshold or the parabolic rule. Therefore, the two former rules are weighting quality relatively less and price relatively more, compared to the other two rules. In addition, this implies that these two scoring rule reflects a higher willingness to pay for a relatively lower quality compared to the linear rule with a price threshold and the parabolic rule. Furthermore, these differences among scoring rules will, for some bid combinations, lead to different winners when comparing different rules. In the next section, we will take a closer look at this.

6.2.2 When Does the Different Rules Provide Different Winners?

We have already shown that different scoring rules require different amounts of quality points for the bidder with the highest price (and lowest price score) to win, e.g. for the winner of the procurement to change. Therefore, ***different differences in quality points among bidders are required to change the outcome (winner)*** of the procurement. Table 6.3 below shows for which original quality differences a change of outcome between the two rules will occur.[15]

Different price combinations from the		Linear 1 & Linear 2	Linear 1 & Parabolic	Linear 2 & Parabolic	Linear 1 & Lowest bid	Linear 2 & Lowest bid	Parabolic & Lowest bid
Price bidder	Price bidder 2	For what orignial quality differences the different winners occur					
100	100	-	-	-	-	-	-
200	100	-	-	-	4,5,6	4,5,6	4,5,6
300	100	8,9,10	9,10	8	9,10	8	-
400	100	10,11,12,13	13	10,11,12	10,11,12,13	10	10,11,12
500	100	10,11,12,13	13	10,11,12	11,12,13	10	11,12
300	200	4,5,6	6	4,5	5,6	4	5
400	200	6,7,8,9,10	9,10	6,7,8	7,8,9,10	6	7,8
500	200	6,7,8,9,10	9,10	6,7,8	8,9,10	6,7,8	8
400	300	2,3	-	2,3	-	2,3	-
500	300	2,3	-	2,3	4,5	2,3,4,5	4,5
500	400	-	-	-	0,1,2	0,1,2	0,1,2

Table 6.3: *Quality Difference that Lead to Different Outcomes in the Procurement*

To understand how these differences are found, we need to look back at Table 6.1 and the differences in price score for different price combinations for the different rules. In other words, we need to compare the columns "diff bidders" for different rules. As an example, we compare the linear rule 1 and the lowest bid rule for the price combination of 200 and 100.[16]

[15] The colours reflect the same as in Table 6.2 above, what rule that has the lowest difference, e.g. what rule that first will change the winner or first let the bidder with highest price win.
[16] In Table 6.3 this example refers to the second row in column 1&4.

The difference in price scores between the bidders is 1 weighted point when using the linear rule 1, and 2 weighted points with the lowest bid rule. ***All quality combinations that result in a quality difference in the range of these differences (between 1 to 2, after being weighted), will lead to different winners with the two rules.*** The reason for this, is that the scoring rule with the lowest difference will experience a change of outcome with a quality difference of 1 weighted point. For the other scoring rule however, there will not be a change of winner until the quality difference is 2 weighted points. Therefore, for quality difference between 1 and 2 weighted points, the two rules will result in different winners. To get a weighted quality difference of 1-2 points the original quality difference need to be of 4, 5 or 6 points (4*0.3 = 1.2, 5*0.3 = 1.5, 6*0.3=1.8).[17] When the quality difference is higher than 2 weighted points, the bidder with the highest price wins, when using either of the rules. Hence, the rules do not result in different winners anymore.

There are two factors explaining where the differences occur; the level of the two "diff bidders" in the comparison, and the difference between "diff bidders" for the two rules. From Table 6.3, we observe that for some bid combinations only one quality difference (example a difference of 10 points) change the outcome, while for other areas, several quality differences will change the outcome (example both 6, 7, 8 and 9 points in quality difference). The number of quality differences result from how big the difference, in the difference in price score between bidders ("diff bidders"), are. Obviously, there are more combinations the more quality differences that appears in Table 6.3.

However, the amount of differences in outcome also depends on the level of "diff bidders" for the two rules. The lower the difference in price scores between both bidders ("diff bidders"), the lower quality difference is needed to change the outcome. Subsequently, when the quality difference is small (for example 2 points), the different quality combinations that hold this difference are higher. Likewise, if the quality difference needed to change the outcome is high (for instance 10 points), the quality combinations that hold this large quality difference are fewer.

The percentage amount of bid combinations resulting in different winners for different price combinations with different pairwise comparisons of scoring rules, are shown in Table 6.4. The table shows for what price combinations the difference between the rules are the greatest.

[17] Quality points of 3 and 7 are outside the range; 3*0.3= **0.9**, 7*0.3 = **2.1**

Different price combinations from the		Linear 1 & Linear 2	Linear 1 & Parabolic	Linear 2 & Parabolic	Linear 1 & Lowest bid	Linear 2 & Lowest bid	Parabolic & Lowest bid
Price bidder	Price bidder	Percentage amount of different winners					
100	100	-	-	-	-	-	-
200	100	-	-	-	14 %	14 %	14 %
300	100	7 %	4 %	3 %	4 %	3 %	-
400	100	5 %	1 %	5 %	5 %	2 %	5 %
500	100	5 %	1 %	5 %	3 %	2 %	3 %
300	200	14 %	4 %	10 %	9 %	5 %	4 %
400	200	15 %	4 %	11 %	11 %	4 %	7 %
500	200	15 %	4 %	11 %	7 %	11 %	3 %
400	300	11 %	-	11 %	-	11 %	-
500	300	11 %	-	11 %	10 %	21 %	10 %
500	400	-	-	-	18 %	18 %	18 %

Table 6.4: *Percentage Amount of Different Winners for Different Price Combinations Comparing Different Scoring Rules*

For the comparison of the linear rule 1 and 2, we observe the most differences when one bidder offers a price between the price threshold and reserve price, and the other offers a price close to the reserve price or above. This is consistent with the observation from examining the graphs of the two rules in Figure 3.2 in Chapter 3. The two curves are the furthest apart for the price ranges in the middle. Moreover, the differences between the linear rule 1 and the parabolic rule, also confirms the theory presented in Chapter 3; that the two rules are alike and thus not lead to many differences in outcome. Therefore, we observe the same patterns when comparing the linear rule without a price threshold and the parabolic rule, as for the two linear rules. The curve of the linear rule 2 is the furthest apart from the curve of the parabolic rule in the middle to high range of prices.

For the comparisons of an absolute and a relative rule, we observe the opposite; namely that there are more differences when both prices are high or both prices are low. The reasoning behind this is that the absolute rules consider both 100 and 200 as low prices, as they benchmark them to the price threshold of 150. Similarly, they consider both 300 to 500 as higher prices, as they benchmark them to the reserve price of 350. However, the lowest bid rule only compares the prices with each other when deciding if the prices are high or low. Therefore, using this rule, a price of 200 in comparison to 100, becomes much higher than compared to 150 which is the benchmark in the linear rule with a price threshold. For the comparisons between the linear rule 1 and the parabolic rule against the lowest bid rule, the absolute rules have the lowest difference between the price scores among bidders (green and blue areas) for these bid combinations. Hence, for the areas where there exist the most differences, the bidder with the highest price will win more often if using the absolute rule. However, this is because prices are considered very similar by the absolute rules for these combinations.

6.2.3 Conclusion – Impact of Different Scoring Rules

We find that both the linear rule without a price threshold and the lowest bid rule, to a larger extent reflects a higher willingness to pay for a relatively lower quality compared to the linear rule with a price threshold and the parabolic rule. Therefore, when a procurer chooses among scoring rules, he must be aware of the fact that these scoring rules themselves reflect different weighing in the price-quality trade-off. While the linear rule without a price threshold and the lowest bid rule places a relatively higher emphasis on price relative to quality, the linear rule with a price threshold and the parabolic rule places a relatively larger emphasis on quality relative to price. As the development of MEAT have placed a larger emphasis on the quality aspects in procurement, this argument should weigh heavily when choosing among rules.

Therefore, we do not find evidence in our study that explain why the relative scoring rule is so widely recommended and utilized in practice. We have shown that absolute scoring rules, as the linear rule with a price threshold and the parabolic rule, display traits that by our opinion would be more favourable for the procurer. According to Dini et al., (2006, p. 307 & 318) the linear rule with a price threshold is to a larger extent easier to use and understand by the bidders, implying that the former is the most favourable. However, as this rule and the parabolic rule are in fact quite similar, they would to a large extent be equivalent to one another.

Compared to the relative scoring rule, the absolute scoring rules are in general more predictable as well. By making each bidder able to calculate their own price score independently, each bidder would to a larger extent be able to customize their tenders in compliance with the procurer's preferences. This is beneficial for both the bidder and the procurer. Paradoxically, the NDA do not even mention absolute scoring rules in their recommendations. We believe this practice is based upon a perception that setting a reserve price and a price threshold may be a challenging task for the procurer. As shown in Chapter 3, however, choosing a price threshold larger than zero would be more beneficial than not having any at all. In addition, if the procurer has just a little knowledge of his operating market, one could argue that he would be able to set a reserve price after all. This does not have to represent his budget constraint, but should rather be used as a helpful tool assuring price competition in an area the procurer finds acceptable.

6.3 Implications of Normalisation

As presented in Chapter 3.3, Ellingsen & Haukeli (2020) claim that *relative* normalisation can arbitrarily change the weight between the award criteria being evaluated. This may in turn change the outcome of the procurement. Due to these arbitrarily changes, they state that practicing normalisation will be contrary to the criterion given by law regarding predictability and equal treatment (Ellingsen & Haukeli, 2020).

We have in our overview shown that there are differences occurring when comparing the lowest bid rule without normalisation and the lowest bid rule with relative normalisation. As these two relative methods only differ on how to evaluate quality, all of the different outcomes arise as a result of normalisation. Hence, it will change the outcome of the procurement for certain bid combinations. Thus, to investigate if these differences are arbitrarily or not, we will examine *when* the differences occur. In order to do so, we investigate the pattern in the 0/1-matrix between the two methods to get a better understanding of what causes the differences. We then run several regressions, testing if our findings are statistically significant. Hence, we will in this part of the analysis present our findings and the regressions, before drawing some conclusions on the results.

The Norwegian Digitalisation Agency (NDA) does not directly state that the quality score they recommend being normalised, are assigned on the basis of an absolute scale. They neither state how the normalisation is to be done, absolute or relative. However, they state the following (Avdeling for offentlige anskaffelser, 2019, translated to English):

> *"The quality criteria are often evaluated against certain evaluation criteria, and there is a risk that none of the tenders will be given the maximum points on these criteria"*

Thus, we assume that the NDA's recommendation is meant for the use of an absolute scale, as with a relative scale the bidder with the highest quality would actually be given the maximum points on these criteria. Regarding what type of normalisation the NDA refer to, it is evident that using an absolute normalisation will not change the weight or the outcome of the competition.[18] According to Ellingsen & Haukeli (2020), absolute normalisation would therefore be unnecessary. However, for the relative normalisation we observe that the relative

[18] Chapter 3.3, p. 19-20

differences between the quality points of the bidders, changes. In addition, this is the method often used in practice, leading up to the ongoing discussion today (Ellingsen & Haukeli, 2020). Therefore, *we assume that NDA recommends relative normalisation when using an absolute scale*. In the following, we will refer to relative normalisation as just normalisation.

6.3.1 Regressions and Findings

For all our regressions in this part of the analysis, we have the same binary response variable, *diff_rank_45*. The variable takes the value 1 for all bid combinations that results in different ranking (winners) between the lowest bid rule without normalisation (method 4) and the lowest bid rule with normalisation (method 5), and 0 otherwise. Furthermore, all our predictors are dummy variables. In addition, when referring to quality score or points in the bid characteristics, we are referring to the original score before weighting or normalising. Our purpose of the regressions is to investigate if certain characteristics of the bid combinations, will increase the probability of observing different winners of the procurement when using the two evaluation methods. We will then be able to draw conclusions regarding these characteristics. Do the characteristics provide reasons to conclude that the outcome is changing arbitrarily? Or is there a reasoning behind the differences?

We have already explained how differences in price scores between two bidders will affect how many additional quality points the bidder with the highest price (and lowest price score) needs in order to win. When two bidders offer the same price, they will always be awarded the same price score, regardless of what scoring rule being used in the procurement. Thus, when prices are equal, the decisive factor of the outcome is the quality evaluation. Methods without normalisation always ensures that the bidder with the highest total quality wins the procurement. However, if the quality score is normalised, this is not always the outcome. As our first finding, we observe different outcomes when the bidders offer the same prices and the same total quality, even if the two quality dimensions are weighted equally. This is a striking observation. Should there really be a difference between bidders when they offer the exact same price and equal total quality?

The predictor, *equal_prices,* takes the value 1 if the prices are equal and 0 if they are different. Similar, the predictor *equal_totalQ* takes the value 1 if the total original quality score offered by the bidders are the same, 0 otherwise. With these predictors we run a logistic regression to test our first finding. The results are shown in Table 6.5 below. Both predictors are positive and highly significant, implying that the probability of finding different winners increase when both prices and total quality is equal. The probability of differences is 21.4 % if the combination of these two characteristics occur.

	Finding 1
(Intercept)	-2.27 ***
	(0.01)
equal_totalQ	0.73 ***
	(0.02)
equal_prices	0.24 ***
	(0.01)
N	366 025
AIC	241 709

*** p < 0.001; ** p < 0.01; * p < 0.05.

Table 6.5: *Regression Results Finding 1*

The observation above is caused by the fact that we have more than one dimension of quality. This lead us to our next finding. We observe many differences when the two bidders perform the best on one quality dimension each. This happens when both the price and their total quality are the same, but also when the prices are the same and their total quality is different. We run a logistic regression testing finding 2; that differences occur when the bidders perform the best on one quality dimension each. Table 6.6 below shows the results of the logistic regression.

	Finding 2	Finding 2.1
(Intercept)	-2.32 ***	-2.33 ***
	(0.01)	(0.01)
equal_prices	0.24 ***	0.24 ***
	(0.01)	(0.01)
highest_one_q	0.24 ***	0.15 ***
	(0.01)	(0.01)
equal_totalQ		0.65 ***
		(0.02)
N	366 025	366 025
AIC	242 607	241 532

*** p < 0.001; ** p < 0.01; * p < 0.05.

Table 6.6: *Regression Results Finding 2*

The predictor, *highest_one_q*, is 1 if they perform the best on one quality dimension each, 0 otherwise. We can conclude that this variable has a significant and increasing impact on the probability of finding differences between the evaluation methods, both when including only *equal_prices* and when adding *equal_totalQ* in the model. However, when including *equal_totalQ*, we observe that *highest_one_q* explains less, but it explains more than finding 1. The probability of observing a difference is 13.7 % when prices are equal and the bidders performs the best on one quality dimension each, while it is 21.6 % if they in addition offers an equal total quality. What might be surprising is the greater probability of a difference when prices and total quality is equal (21.6 %), compared to the situation where total quality is not equal (13.7%). However, the reasoning behind this, is that we need to specify the characteristics for the situation where total quality is not equal, more than we already have. We will revisit this, after shortly examining the situation when prices are different.

We also observe differences in outcome between the methods when prices are different. However, in these cases, we only find differences when the bidder with the highest price also offers the highest total quality. This imply that even with normalisation, the bidder with the highest price will not be able to win if he offers the lowest total quality. The predictor, *highest_p_q*, is 1 if one of the bidder has the highest price and total quality at the same time, 0 otherwise. The results from the logistic regression using this predictor are shown in Table 6.7 below. The predictor is highly significant and positive, indicating a higher probability of finding differences with these bid characteristics. The probability of difference in this case is 19.8 %.

	Finding 3
(Intercept)	-3.03 ***
	(0.01)
highest_p_q	1.63 ***
	(0.01)
N	366 025
AIC	222 556

*** p < 0.001; ** p < 0.01; * p < 0.05.

Table 6.7: *Regression Results Finding 3*

Our next findings occur regardless of the price combination. However, as we will observe afterwards, it will be possible to explain the differences more accurate when separating between equal prices and not equal prices. First, we observe a common feature for all combinations; differences occur when the difference in total quality between the bidders are low. We have defined three variables for "low" difference in total quality; *diff_Q_low* which is 1 if the total quality is below 5, and 0 otherwise, *diff_Q_lav*[19] which is 1 if the difference in total quality is below 7 points, and 0 otherwise, and *diff_Q_under11* which is 1 if the total quality is below 11 points, and 0 otherwise. The results of the regressions with each of these predictors are shown in Table 6.8. The results show that the difference has to be lower than diff_Q_under11, as this predictor is not significant. Furthermore, if low quality is defined as under 5 or 7 ("low" or "lav"), it is highly significant and positive, indicating a higher probability of finding different outcomes for these bid combinations. Since quality under 5 explains less compared to when it is defined as below 7, we will use the predictor *diff_Q_lav* in the further analysis.

	Finding 4	Finding 4.1	Finding 4.2
(Intercept)	-2.70 ***	-18.57	-2.49 ***
	(0.01)	(34.50)	(0.01)
diff_Q_lav	0.73 ***		
	(0.01)		
diff_Q_under11		16.52	
		(34.50)	
diff_Q_low			0.50 ***
			(0.01)
N	366 025	366 025	366 025
AIC	240 133	235 187	241 511

*** p < 0.001; ** p < 0.01; * p < 0.05.

Table 6.8: *Regression Results Finding 4*

Moreover, as normalisation involves adjusting the quality dimension to a larger numeric value, we expect the effect of normalisation to be greater when the highest quality offered among the two bidders, is low. We define a variable for each quality dimension (q1 and q2) for each bidder (t1 and t2), being 1 if the q1 or q2 for t1 or t2 is lower than 4 (being in the very low range of 0-3 out of 10), and 0 otherwise. The predictors are the following; low_q1_t1, low_q1_t2, low_q2_t1 and low_q2_t2. When running a logistic regression with these predictors, we

[19] We use the Norwegian word for "low" = "lav", as we have used "low" in the previous version

observe the same positive and significant coefficient of 0.43 for all four variables. The regression results are shown in Table 6.9.

	Finding 5
(Intercept)	-2.86 ***
	(0.02)
low_q1_t1	0.43 ***
	(0.01)
low_q2_t1	0.43 ***
	(0.01)
low_q1_t2	0.43 ***
	(0.01)
low_q2_t2	0.43 ***
	(0.01)
N	366 025
AIC	237 450

*** p < 0.001; ** p < 0.01; * p < 0.05.

Table 6.9: *Regression Results Finding 5*

After having investigated the data in detail, we are able to explain the bid characteristics that results in different outcomes more specifically. Therefore, we will now summarize all the findings by running two regressions; the first will explain the part of the data where prices are equal, while the second explains the differences occurring when prices are different. In both regressions, we use the exact combination of predictors (or characteristics) in one interaction term that we believe explains the pattern of differences in each part of the data.

All the variables remain the same, except from *highest_p_q*, which is removed. Instead we have the variables *t1_highest_price* and *t1_highest_totalquality*, where t1 stands for bidder 1. The variables are 1 if the bidder has the highest price or total quality, respectively, and 0 otherwise. All together they are catching up the same as *highest_p_q*, but only for bidder 1. However, the data are symmetric, so the results from this will be the same if we instead examined bidder 2. In addition, the variables catch up which bidder has the highest price and total quality. We need to know this, as we observe that the bidder who has the highest price and quality, also offers a low quality in the cases where the differences occur. Therefore, when prices are different, we also include the variables *low_q1_t1* and *low_q2_t1*. Lastly, we also include the variable *diff_Q_lav*. The difference between using this variable and the two variables, *low_q1_t2* and *low_q2_t2,* is that the quality for bidder two does not have to be as low as under 4 on both dimensions. With *diff_Q_lav,* only the total quality difference needs to be under 7. Therefore, we chose to use this variable as it explains more of the differences. These variables are run together in Model 7 in Table 6.10 below.

For equal prices we find that, equal total quality or not, the difference occur when the bidders perform the best on one quality dimension each, the difference in total quality is low[20], and both have low quality on one of the dimensions.[21] With the variable, *diff_Q_lav*, we ensure that the second dimension can be high, but not very different compared to the other bidder. These variables are run together in Model 6 in Table 6.10.

	Model 6	Model 7
(Intercept)	-2.21 ***	-2.18 ***
	(0.01)	(0.01)
equal_prices*highest_one_q*low_q1_t1*low_q1_t2*diff_Q_lav	2.74 ***	
	(0.04)	
t1_highest_price*t1_highest_totalquality*low_q1_t1*low_q2_t1*diff_Q_lav		2.53 ***
		(0.06)
N	366 025	366 025
AIC	238 409	241 793

*** p < 0.001; ** p < 0.01; * p < 0.05.

Table 6.10: *Regression Results Interaction Terms*

We observe from the regression results in Table 6.10, that both interaction terms are highly significant and have a higher positive coefficient, than when running the variables separately. This means that there is a much higher probability for different outcomes with the two rules when having these characteristic of the bids. In fact, the probability of finding a difference between the two methods when prices are equal and the bids have the characteristics as in Model 6, is 62.9 %. The probability when prices are different and characterised as in Model 7, is 58.7 %. These probabilities are much higher than when the variables are tested separately, and they are quite high despite a very specific situation. For robustness of our results, we have in Appendix A3 included all variables in one model. The results are to a great extent the same, especially the interaction terms are still positive and highly significant. Hence, we can conclude that a procurer should be especially careful using normalisation when the bids are characterised according to the interaction terms in Model 6 and 7.

[20] A difference in total quality below 7 is low as there is 20 points possible in total (10 points on each dimension)
[21] Low quality on one dimension = under 4 points

6.3.2 Implications of Findings

By performing regressions, we were able to indicate what bid characteristics resulted in different outcomes. We will now discuss what these findings imply for the use of normalisation and illustrate with some examples.

6.3.2.1 Normalisation Have a High Impact when Quality is Low

We observe significantly more differences in outcome when each bidder performs the best on one quality dimension each and they offer the same prices and total quality. We can therefore conclude that the *combination* of quality has an important impact on the results when normalising, despite that the two quality dimensions are equally weighted. Without normalisation, we know that the difference in total quality points between the bidders are the decisive factor to whom win the competition. This makes intuitively sense, as the two dimensions are weighted the same and has the same scale from 0 to 10. Therefore, when we now observe that combinations like (100,0,2) & (100,1,1) can make the latter bidder win (instead of performing equally), normalisation seems to favour offering quality on both dimensions, rather than performing better on one. Hence, the weight of the two quality dimensions is changed, resulting in a higher weight of quality for bidder two in this case.

However, we have also proved a significant higher amount of differences when prices are equal and total quality are not, but the bidders perform the best on one dimension each. With the tender combinations; (100,0,1) & (100,10,0), the last tender offers a quality of 9 points higher than the first. Still, the tenders are perceived as equally good by the method with normalisation, while without normalisation the second tender wins the procurement. We do not understand the reasoning behind such a practice, and agree with Ellingsen and Haukeli that normalisation seems to change the weights of quality and the outcome arbitrarily.

The last example is somehow extreme, as each bidder are assigned a score of zero on the opposite dimension of each other. Therefore, a difference in quality higher than 7 is easier to achieve.[22] However, in the regressions this is caught up by the variables for low quality. Another example can better explain why these differences especially occur when we have zeros in one dimension; (100,2,2) & (100,0,8). Also, in this case, we observe that the second bidder has a

[22] diff_Q_lav was defined as under 7, and we proved this was significant, but not when quality where under 11 (diff_Q_under11).

higher quality in total than bidder one. Therefore, the second bidder wins when using the method without normalisation. However, using the method with normalisation, he loses; The first bidder wins the first dimension (2 > 0), and thus gets 10 points here. The same does bidder 2 with regards to the second dimension (8 > 2). However, since the first bidder has achieved some points on the second dimension, he gets 2/8 of the 10 points[23] that bidder two (winner of the second dimension) is awarded. Bidder two on the other hand, get zero points on the first dimension since 0/2 * 10 is still zero. Therefore, when having zero quality points on one dimension, the normalisation will actually not adjust that score relative to the other. Hence, the difference in quality points changes from being 2 points originally, to 10 points in favour of bidder 1. Therefore, the relative difference between the bidders increases much more on this dimension. On the other dimension, the relative difference only increases by 0.5 point. Thus, normalisation "punishes" especially harshly if a bidder is not able to offer any quality above the minimum requirement (0 in quality score). In addition, also in this example, the changes seem arbitrarily as the results depend on a "right" combination of quality in the two dimensions; Even if the two dimensions are weighted equally and each point of quality are originally intended to be equally important, the bidder with lowest total quality wins with normalisation.

Common for all these examples, is that the quality is low in either both dimensions for both bidders, or low in one dimension for both bidders. In addition, the total quality difference is in most cases low. However, the low quality difference alone cannot explain the outcomes as it did for different scoring rules in section 6.2. This can be illustrated by the following bid combination; (100,5,8) & (100,10,7). In this example, we do not observe different outcomes, even if the difference in total quality are low (4 points) and they perform the best on one dimension each.[24] Hence, normalisation probably has the highest impact when quality is low, which is reasonable as the adjustment to 10 points is greater for those cases. Nevertheless, the combination of quality is important, as if both bidders have the exact same and low quality on both dimensions, normalisation does not change the outcome. This coincide with the highly significant and positive coefficients we observe in Model 6 and 7 when the interaction of the predictors is tested.

[23] Score after normalisation: 2/8 * 10 = 2,5 points

[24] Scores with normalisation: Bidder 1: 5 + 10 = 15, Bidder 2: 10 + 8,75 = 18,75. Without normalisation the bidder with the highest total quality always wins, e.g. bidder 2.

6.3.2.2 Normalisation Reflects a Higher Willingness to Pay for Low Quality

We have until now, examined combinations where prices are equal in order to comment on the effect of low quality. For different prices, the bids also have to be characterized by low quality for normalisation to have an impact. However, now we observe that the quality has to be low for both bidders. This make sense, as we have seen that normalisation has a higher impact on increasing the quality score when the quality is adjusted from a low value. However, this implies that the impact of normalisation is even stronger as it also changes the outcome when prices are different and a higher quality difference is needed to change the outcome.

We remember from chapter 6.2.1, that when bidders offer different prices, their price scores are also different. The greater the difference in price scores, the higher degree of quality the bidder with the highest price (and lowest price score) needs to offer to be able to win the procurement. In other words, the higher the price difference, the more difficult it is for the bidder with highest price to win. With normalisation, however, we observe that the bidder with the highest price (and the lowest price score), are able to win even if he only offers a tiny point of higher quality. Let us use an example to illustrate this quite striking observation; (400,1,1) & (100,0,0). In this example bidder one has 1 point more on both quality dimensions, in total two points out of 20 possible. With normalisation, the quality points (before weighting) will be adjusted so the bids will change to (400,10,10) & (100,0,0). The relative difference between the tenders change from 2 points to 20 points on the quality dimension. Hence, with normalisation, the first bidder actually wins, even with a price 4 times higher than the other.[25] We remember that the original score is given objectively using an absolute scale, reflecting a small difference in quality between the bidders. In addition, it is important to remember that having a score of 0, does not mean they are offering no quality at all, only no quality above the minimum requirements. One can therefore ask; Are the normalised scores really reflecting the relative difference of the bids correctly, when changing so drastically? Still, whether this is intentionally or not, normalisation increase the willingness to pay for low quality.

[25] Total score for bidder 1 with normalisation is 1+6 = 7. Total score for bidder 2 with normalisation is 4.

6.3.2 Conclusion – Impact of Normalisation

We have by the performed regressions proved that certain characteristics of bid combinations have a significantly higher probability of changing the outcome of the procurement. Normalisation changes the outcome both when prices are equal and when they are not. Common for both groups are the feature of low quality and the importance of a "correct" combination of the quality in order to win. Hence, we can conclude that normalisation causes unreasonable and arbitrarily changes in the weights of low qualities. This arbitrarily changes the outcome of the procurement, but in favour of the bidder with the highest price. Hence, it changes the procurer´s preferences to become more willing to pay for relatively low quality.

7 Concluding Remarks and Reflections

The aim of this thesis was to investigate how different evaluation methods affect outcomes in procurement. In this final chapter, we will therefore provide concluding remarks to our findings, reflect on the validity of the model assumptions and suggest potential future extensions.

7.1 Conclusion

In our master study, we have performed bid simulations using our own algorithm. This provided observations on *when* different evaluation methods lead to different outcomes. We have proved that price is weighted relatively more than quality when using the linear rule without a price threshold or the lowest bid rule, compared to the parabolic rule and the linear rule with a price threshold. This is caused by the difference in price points assigned by the different rules. As the two former methods leads to smaller differences in price points between bidders, a lower degree of quality needs to be offered to compensate for a higher price. Therefore, the bidder with the highest price (and lowest price score) would prefer using one of these rules as it would be easier for him to offset his relatively high price. These findings also have implications for the procurer. If he prefers that quality should be weighted relatively more compared to price, we would recommend using one of the latter rules. However, if he prefers price to be weighted relatively more than quality, one of the former rules could be used. The important lesson to keep in mind from this analysis, is that not only weighting functions reflect the procurer's preference in the trade-off between price and quality. Scoring rules themselves also lead to different weighing in the price-quality trade-off. As both national and international procurement legislation places a larger emphasis on quality in procurement, we argue that this would be an argument for using either the linear rule with a price threshold or the parabolic rule.

Furthermore, there is a major difference between absolute and relative scoring rules. All of the three absolute scoring rules ensures the basic principle of predictability as each bidder are able to calculate their own price scores in advance of evaluation. This characteristic is pivotal in procurement, as it enables bidders to offer tenders more closely in line with the procurer´s preferences. This is not possible with the relative scoring rules as each bidder's price score depends on the other bidders' price bids. Therefore, the bidder has no knowledge of the procurer's preferences and the trade-off between price and quality. This difference is widely discussed in literature, and constitutes an argument for not using the relative rules. Along with the observation that the relative lowest bid rule weigh price relatively more than quality, we wonder why the NDA recommend using this evaluation method? Actually, none of the scoring rules presented by the NDA are absolute rules. We have not found any logical reasons to this practice and would recommend the NDA to revise their recommendations, away from the relative rules. These rules are unpredictable, they do not reflect the buyer´s preferences and are actually weighting quality less than the preferred absolute rules.

In our regression analysis we have proved that performing normalisation on certain characteristics of bid combinations, have a significantly higher probability of changing the outcome of procurement. This happens both when prices are equal and when they are not. Common for both groups are the feature that normalisation has its greatest impact when qualities are low, highlighting the importance of a "correct" combination of the quality in order to win. Hence, we can conclude that normalisation causes unreasonable and arbitrarily changes in the weights of low qualities. These changes are in favour of the bidder with the highest price, changing the procurer´s preferences to become more willing to pay for low quality. We believe that the NDA has not foreseen these consequences, as predictability is an important principle in both national and international procurement legislation today. Therefore, one could even argue that practicing normalisation is illegal, as it contradicts basic legal principles.

In total, we agree with the critics, claiming normalisation is highly unnecessary and even wrong. When quality is evaluated by an expert panel, they assign each quality dimension a score based upon their perception of the offered quality. This can be done based upon an absolute or relative scale. If an absolute scale is used, all bidders are able to compete for the maximum amount of points. Hence, if none of them are able to provide the maximum quality, we argue that it is wrong to change the scores by relative normalisation. We have shown that this kind of normalisation are able to change the outcome of procurement, thereby contradicting procurement legislation. Moreover, if the quality were assessed based upon a relative scale, one of the bidders would get a maximum score regardless, and there would be no need to adjust the scores afterwards. Hence, in such cases normalisation is unnecessary. In line with Ellingsen & Haukeli (2020), our recommendation to NDA would be to rather advise procurers to use a relative scale for the evaluation of quality aspects if a relative scoring rule is used. In this way, one could prevent these arbitrary and unpredictable outcomes in procurement caused by relative normalisation.

7.2 Reflections and Future Research

In our analysis, we have not focused on the degree of predictability between the absolute scoring rules and the relative scoring rules. It is clearly stated in literature that the absolute scoring rules are highly predictable, compared to the relative scoring rules which are not. These facts are therefore taken for granted, not making this a primary focus in our study. We rather wanted to explain the different scoring rules from a somehow different angle than already being made in literature. Thus, we focused more on the implications of all rules and compared them, trying to explain what the different scoring rules would imply for the procurer's preferences.

Furthermore, we have in our analysis used an absolute scale for the quality dimension. However, the points given in our simulations could just as well have been relative scores. Thereby, our analysis is, just as good, a proof for Ellingsen & Haukeli´s statement regarding arbitrarily changes in outcome when using a relative normalisation on a relative scale (Ellingsen & Haukeli, 2020).

When we initially started our study, we wanted to investigate a new type of evaluation method, just recently launched by Sykehusinnkjøp. This is an evaluation method belonging to the group quality-to-price-scoring, and is called the Monetary Evaluation Method. In our algorithm, we initially performed simulations using this method as well. Discussing this type of scoring rule proved to be beyond the scope of our master's thesis. However, it would be interesting to go further and compare the two categories of scoring rules by using a similar approach as we have performed in this thesis. If procurers or other students would like to experiment and investigate this method further, the algorithm including the monetary method can be provided.

Finally, in our model and analysis we only have five prices. Therefore, there are few prices in the range between the price threshold and the reserve price (only 200 and 300 are inside the range) when using the linear rule with a price threshold. This might have limited our analysis with regards to the interpretation of the effect of a price threshold and a reserve price. However, it was important for our focus to have some prices outside the range. This was done in order to compare the absolute rules with the relative rule, as the latter does not consider minimum or maximum prices. Therefore, it could be interesting for future research, to include more prices in-between these upper and lower limits to examine how the differences between absolute and relative scoring rules are affected by how the price threshold and the reserve price are set.

References

Avdeling for offentlige anskaffelser. (2020a, 27.05). Tilbud steg for steg (leverandører). Retrieved from: https://www.anskaffelser.no/anskaffelsesprosessen/tilbud-steg-steg-leverandorer

Avdeling for offentlige anskaffelser. (2020b, 05.05). Anskaffelsesprosessen steg for steg – Tildelingskriterium. Retrieved from: https://www.anskaffelser.no/anskaffelsesprosessen/anskaffelsesprosessen-steg-steg/avklare-behov-og-forberede-konkurransen/spesifikasjoner-krav-kriterier-og-kontraktsvilkar/tildelingskriterium

Avdeling for offentlige anskaffelser. (2020c, 31.01). Om oss. Retrieved from: https://www.anskaffelser.no/om-oss

Avdeling for offentlige anskaffelser. (2019, 16.08). Anskaffelsesprosessen steg for steg – Relative poengmodellar. Retrieved from: https://www.anskaffelser.no/anskaffelsesprosessen/anskaffelsesprosessen-steg-steg/avklare-behov-og-forberede-konkurransen/spesifikasjoner-krav-og-kriterier/tildelingskriterium/evalueringsmodellar/relative-poengmodellar

Bjørnstad, B. E. (2019, 05.02). En bedre evalueringsmodell? *Anbud365.* Retrieved from: https://www.anbud365.no/regelverk/en-bedre-evalueringsmodell/#_Hlk528572405

Bergman, M. A., & Lundberg, S. (2013). Tender Evaluation and Award Methodologies in Public Procurement. *Journal of Purchasing & Supply Management*, *19*, 73-83.

Difi. (2015, 08.09). *RAS Trøndelag.* Scandic Hell Hotell. Retrieved from: https://www.difi.no/sites/difino/files/20150908_aws_presentasjon_og_oppgaver-print.pdf

Dini, F., Pachini, R. & Valetti, T. (2006). Chapter 12: Scoring Rules. In Dimitri, N., Piga, G., Spagnolo, G. (Eds.), *Handbook of Procurement,* Cambridge University Press, 293-321.

Ellingsen, C. B. & Haukeli, O. E. (2020, 18.03). Normalisering kan vilkårlig endre tildelingskriterienes vekt. *Anbud365*. Retrieved from: https://www.anbud365.no/regelverk/normalisering-kan-vilkarlig-endre-tildelingskriterienes-vekt/?utm_source=Nyhetsbrev+fra+Anbud365&utm_campaign=49d9f93fc8-EMAIL_CAMPAIGN_2019_05_22_01_20_COPY_04&utm_medium=email&utm_term=0_580f3ce305-49d9f93fc8-175435681

European Parliament. (2020, 04). Public Procurement Contracts. Retrieved from: https://www.europarl.europa.eu/factsheets/en/sheet/34/public-procurement-contracts

Hackerearth. (2020). Practical Guide to Logistic Regression Analysis in R. Retrieved from: https://www.hackerearth.com/practice/machine-learning/machine-learning-algorithms/logistic-regression-analysis-r/tutorial/

Hobson, C. (2016, 27.04). Can the EU directive on public procurement encourage value-based procurement in healthcare? *Medtech Views*. Retrieved from: http://www.medtechviews.eu/article/can-eu-directive-public-procurement-encourage-value-based-procurement-healthcare?page=1

James, G., Witten, D., Hastie, T. & Tibshirami, R. (2013). *An introduction to statistical learning with applications in R*. New York: Springer.

Le, J. (2018, 10.04). Logistic Regression in R Tutorial. *Datacamp*. Retrieved from: https://www.datacamp.com/community/tutorials/logistic-regression-R

Regjeringen. (2017, 11). Veileder til reglene om offentlige anskaffelser (Anskaffelsesforskriften). Retrieved from: https://www.regjeringen.no/contentassets/df547bb0f73d43d9b90756002473f680/no/pdfs/veileder-offentlige-annskaffelser.pdf

Rådgivende ingeniørers forening (RIF). (2020, 01). How to derive MEAT criteria. Introducing quality criteria into procurement. Retrieved from: https://www.rif.no/wp-content/uploads/2019/12/MEAT-Criteria_final.pdf

Statistical tools for high-throughput data analysis (STHDA). (2018, 11.03). Logistic Regression Assumptions and Diagnostic in R. Retrieved from: http://www.sthda.com/english/articles/36-classification-methods-essentials/148-logistic-regression-assumptions-and-diagnostics-in-r/

Support for Improvement in Governance and Management (SIGMA). (2016, 09). 2014 EU directives: Public Sector and Utilities Procurement. *Sigma Pubic Procurement Briefs*. Brief 30. Retrieved from: http://www.sigmaweb.org/publications/Public-Procurement-Policy-Brief-30-200117.pdf

The Comprehensive R Archive Network (CRAN). (2020). Collinearity Diagnostics, Model Fit & Variable Contribution. Retrieved from: https://cran.r-project.org/web/packages/olsrr/vignettes/regression_diagnostics.html

Appendix

A1 Regressing the Pairwise Differences from Chapter 6.1

To statistically be able to conclude if the differences among the pairwise comparisons of evaluation methods from section 6.1 are statistically significant, we perform some logistic regressions. For these regressions, we need to simulate the data into a slightly different structure. Instead of having one dependent variable for each pair of evaluation methods (as diff_rank_45 in chapter 6.3), leading to one row of observation for each bid combination, we now have one row for every bid combination for each pairwise comparison.[26] Hence, our dependent variable is called *diff_rank*, and takes the value 1 if the observation is a bid combination with different ranking, 0 otherwise. Furthermore, we have one dummy variable for every pair of evaluation methods, called ruleXX, where XX is the number of the two methods compared. We present Table 4.1 from Chapter 4 to remind you of the numeration from the simulation:

Method	Name
1	Linear rule with a price threshold, no normalisation
2	Linear rule without a price threshold, no normalisation
3	Parabolic rule, no normalisation
4	Lowest bid rule without normalisation
5	Lowest bid with normalisation

Table A1.1: *Overview of Different Evaluation Methods*

The dummy variables take the value 1 if the observation belongs to the given pair of methods, 0 otherwise. All observation belongs to one of these dummies, so we must exclude one at a time from the regression to avoid perfect multicollinearity. In order to compare two of the dummy variables, we must leave out one of them as a reference group, as a regression is only testing whether the predictors are statistically different from the reference group, not if they are different from each other. In addition, we need to include all other dummies, so the reference group only consist of the variable of our interest.

Table A1.2 shows the comparison between the linear rules (1&2) as the reference group, testing our first observation that there are more differences between the two linear rules than between

[26] The Algorithm is presented in Appendix A4.2

the first linear (1) and the lowest bid rule (4). We also test if the difference between 1&2, and the second linear (2) and the lowest bid rule (4), is significant. In addition, we test whether there is a statistically higher probability of differences between the linear rules than between linear rule 1 and the parabolic rule (1&3). We observe that all coefficients are highly significant at the 1% level. The negative coefficient of *rule14* prove that the differences between the linear rule 1 and the lowest bid rule (1&4) are significantly lower than for the two linear rules. In addition, the differences in the comparison of linear rule 2 and the lowest bid rule (2&4) are statistically higher than when comparing the two linear rules. However, the coefficients are small, indicating a low change in probability of different ranking between the methods. On the other hand, *rule13* has a much higher negative coefficient, indicating the comparison 1&2 has a much higher probability of different ranking than in the comparison of the linear rule 1 and the parabolic rule (1&3). For the next tables, we have also included all dummy variables, but only the relevant variables will be shown.

	Coefficients
(Intercept)	-2.62 ***
	(0.01)
rule15	0.76 ***
	(0.01)
rule25	0.49 ***
	(0.01)
rule35	0.63 ***
	(0.01)
rule14	-0.04 ***
	(0.01)
rule24	0.07 ***
	(0.01)
rule34	-0.32 ***
	(0.01)
rule13	-1.55 ***
	(0.02)
rule23	-0.27 ***
	(0.01)
rule45	0.46 ***
	(0.01)
N	3 660 250
AIC	1 951 748

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Table A1.2: *1&2 as Reference Group*

The next three tables are testing the statistic relationship between the pairs of absolute and relative rules without normalisation, and the pairs where the method with a relative rule also uses normalisation. The pairs without normalisation is the reference group. Table A1.3, A1.4 and A1.5 show a positive and significant relation for all comparisons. The coefficients are relatively high and indicates that normalisation has a great impact on changing the winner in procurement.

	Coefficients
rule15	0.80 ***
	(0.01)
N	3 660 250
AIC	1 951 748

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Table A1.3: *1&4 as Reference Group*

	Coefficients
rule25	0.41 ***
	(0.01)
N	3 660 250
AIC	1 951 748

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Table A1.4: *2&4 as Reference Group*

	Coefficients
rule35	0.95 ***
	(0.01)
N	3 660 250
AIC	1 951 748

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Table A1.5: *3&4 as Reference Group*

To clarify this, we have performed a regression where we instead of looking at the rules, investigate two variables explaining whether the two rules compared differ in the type of rule (absolute or relative) and if they differ with regards to normalisation. The dummy variable *absolute_relative_diff,* takes the value 1, if one of the rules in the observation is absolute and the other is relative, and 0 otherwise (if both are absolute or both are relative). The dummy variable *normalization_diff* takes the value 1 if one of the methods compared perform normalisation and the other does not, 0 otherwise. The dependent variable is still *diff_rank*.

	Coefficients
(Intercept)	-2.76 ***
	(0.00)
absolute_relative_diff	0.05 ***
	(0.01)
normalization_diff	0.71 ***
	(0.00)
N	3 660 250
AIC	1 984 030

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Table A1.6: *Differences Due to Category of Scoring Rules and Normalisation*

We observe in Table A1.6, that both variables are positive and significant at the 1% level, indicating the probability of observing different outcomes is higher if the methods differ in these two characteristics. In addition, we observe that the difference in normalisation has a much higher impact on the outcome, as the coefficient is larger. We are therefore investigating normalisation further in part two of the analysis, in addition to the differences between scoring rules in the first part.

A2 Matrices Used to Analyse Patterns in 0/1-Matrix between Method with and without Normalisation

In order to get an understanding regarding the pattern in the 0/1-matrix between the two methods; lowest bid rule without normalisation (4) and lowest bid rule with relative normalisation (5), we started our search for pattern in a smaller data set. This smaller data set has the same parameters, only with one quality dimension instead of two.

Table A2.1 is the first matrix we made, and shows the percentage amount of different winners for different price combinations. What is to be noticed from this matrix, is that when having one quality dimension, we do not observe differences when prices are equal. From chapter 6.3, we observe that normalisation in fact changes the outcome of the procurement when prices are equal and quality in total is equal. We argue that this is due to the combination of quality offered in the two dimensions. By looking at the differences between the methods, with only one dimension, we observe that there are no differences when prices are equal. Hence, the observation of different outcomes when prices are equal are dependent on having more than one dimension of quality.

Price	100	200	300	400	500
100		12.4 %	12.4 %	15.7 %	12.0 %
200	12.4 %		9.9 %	12.4 %	14.0 %
300	12.4 %	9.9 %		5.0 %	7.4 %
400	15.7 %	12.4 %	5.0 %		5.8 %
500	12.4 %	14.0 %	7.4 %	5.8 %	

***Table A2.1:** Price Matrix*

Furthermore, we wanted to analyse if there are certain quality combinations among the two bidders that result in more differences than others. Table A2.2 shows the percentage amount of bid combinations where the two methods lead to different winners when there are different quality combinations among the two bidders. The vertical columns show the highest offered quality, while the horizontal rows denote the quality difference between the two bidders in absolute terms. By quality, we here refer to the original quality score assigned by the absolute scale.

We observe that the highest percentages of change occur when the highest quality is in the lower or medium range, especially the low range, and the absolute difference in offered quality is low. Hence, this is absolutely consistent with our findings from the regressions in chapter 6.3; Normalisation has its greatest impact on the outcome when both bidders offer a low degree of quality, which subsequently shows that it is not necessary to provide a high degree of quality in order to win.

	Highest quality										
Difference in quality (in absolute value)	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
0	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %	0 %
1	not possible	40 %	34 %	24 %	12 %	8 %	8 %	4 %	0 %	0 %	0 %
2	not possible	not possible	32 %	32 %	28 %	20 %	16 %	8 %	4 %	4 %	0 %
3	not possible	not possible	not possible	24 %	24 %	20 %	20 %	12 %	8 %	8 %	0 %
4	not possible	not possible	not possible	not possible	16 %	16 %	16 %	16 %	12 %	8 %	0 %
5	not possible	not possible	not possible	not possible	not possible	8 %	8 %	8 %	8 %	8 %	0 %
6	not possible	not possible	not possible	not possible	not possible	not possible	0 %	0 %	0 %	0 %	0 %
7	not possible	not possible	not possible	not possible	not possible	not possible	not possible	0 %	0 %	0 %	0 %
8	not possible	not possible	not possible	not possible	not possible	not possible	not possible	not possible	0%	0 %	0 %
9	not possible	not possible	not possible	not possible	not possible	not possible	not possible	not possible	not possible	0%	0 %
10	not possible	not possible	not possible	not possible	not possible	not possible	not possible	not possible	not possible	not possible	0%

Table A2.2: *Quality Matrix*

The highest percentage of all, occurs when the highest quality is 1 and the difference in quality is 1. Hence, when one of the bidders have 1 quality point and the other has 0 points. For this area, as much as 40 % of the bid combinations result in different outcome. This underlines our argument from the implications of the findings in chapter 6.3, that normalisation “punishes” especially harshly if a bidder is not able to offer any quality above the minimum requirement (0 in quality score).

A3 Robustness of Logistic Regressions in Chapter 6

A3.1 Model with all Predictors

To test our results from the logistic regressions performed in Chapter 6 (where we only included the interactions terms alone), we have performed a regression including all the predictors and the interaction terms in one regression. The results are shown in Table A3.1.

	Coefficients
(Intercept)	-3.19 ***
	(0.02)
low_q1_t1	0.24 ***
	(0.01)
low_q1_t2	0.38 ***
	(0.01)
low_q2_t1	0.34 ***
	(0.01)
low_q2_t2	0.49 ***
	(0.01)
diff_Q_lav	0.76 ***
	(0.01)
highest_one_q	-0.13 ***
	(0.01)
equal_prices	0.04 *
	(0.02)
t1_highest_price	-0.03 **
	(0.01)
t1_highest_totalquality	-0.20 ***
	(0.02)
equal_prices*highest_one_q* low_q1_t1*low_q1_t2:diff_Q_lav	2.24 ***
	(0.04)
t1_highest_price*t1_highest_totalquality*low_q1_t1*low_q2_t1*diff_Q_lav	1.65 ***
	(0.06)
N	366 025
AIC	229 718

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Table A3.1: *Implications of Different Bid Characteristics on Probability*

We observe that all predictors are still significant, most on the 1 % level. We observe that some of the variables alone are now negative. However, when placed together with the interaction terms, the combinations we observed that resulted in changes in outcome, are still positive and highly significant. The coefficients of the interaction term are still high, and the probability of finding difference in ranking when prices are equal, the bidders win in on one dimension each, difference in total quality is low (defined as under 7) and both bidders offer low on quality dimension one, is in fact 99.88 %. For the other combination, when bidder one has the highest price and total quality, the difference in total quality is low and the quality for bidder one is low, the probability is only 39.4 % (compared to 58 % in Chapter 6). However, this is still a relatively high percentage. In addition, if we have a low quality for bidder two, the probability increases to 57 %, providing quite similar results as presented in Chapter 6.

A3.2 Testing for Strong Multicollinearity

When running logistic regression there are four important assumptions to be aware of (STHDA, 2018):

1. The outcome is a binary variable, taking values 0 or 1.
2. There is a linear relationship between the logit and the predictors (equation 3 in Chapter 5).
3. There are no influential values, outlier or extreme values in the continuous predictors
4. There is no high correlation (multicollinearity) among the predictors

As we are using only dummy variables as predictors (independent variables), there is no need to test for the second and third assumption, as these are only meaningful for continuous variables. The first assumption is fulfilled, as we have a binary outcome for our response variable. Thus, it is only necessary to investigate the assumption of multicollinearity further.

Multicollinearity refer to the problem that occur when two or more predictors are highly correlated (STHDA, 2018). Two variables can be positive and negative correlated, indicated by 1 or -1 for perfect correlation. If the correlation coefficient is 0, there is no sign of correlation. The problem should be fixed by removing the concerned variables.

To test for correlation among predictors, we use a measure called **Variance Inflation Factor (VIF)**. VIF measures how much the variance of the estimated regression coefficient (betas) is inflated, by the correlation among the predictor variables in the model (CRAN, 2020). A VIF of 1 for a specific predictor, means there is no correlation among the predictor and the rest of the predictor variables. Hence, the coefficient estimate of that predictor is not inflated at all.

The general rule of thumb is that VIF should not exceed 4, and a VIF of more than 10 is a sign of serious multicollinearity (CRAN, 2020).

The VIF-values for our regression in Table A3.1 are presented in Table A3.2.1 below.[27] All values are below the value of 2. Hence, there is no sign of multicollinearity in this model and this assumption for logistic regression holds.

Predictor	VIF
low_q1_t1	1.29
low_q1_t2	1.28
low_q2_t1	1.26
low_q2_t2	1.25
diff_Q_lav	1.19
highest_one_q	1.21
equal_prices	1.30
t1_highest_price	1.22
t1_highest_totalquality	1.80
low_q1_t1*low_q1_t2*diff_Q_lav*highest_one_q*equal_prices	1.19
low_q1_t1*low_q2_t1*diff_Q_lav*t1_highest_price*t1_highest_totalquality	1.09

Table A3.2.1: *VIF for the Model in Table A3.1.1*

A3.3 Model Accuracy

There exists several metrics for evaluating model fit. However, we have focused on two of the most common; Akaike Information Criteria (AIC) and The Confusion Matrix.

Akaike Information Criteria (AIC) is the counterpart of R-squared, that is used as a metric of model fit in linear regressions. However, the rule for AIC is the smaller the better. This is therefore, a bit useless when only having one model, because you would have to compare different models to understand if one is better than the other (Hackerearth, 2020). The AIC for this model is 229 718 as presented in Table A3.1. It is difficult to say whether this is a high number or not, but it is lower than most of the models in Chapter 6, where the variables are explaining the dependent variable alone.

[27] We have also done this for the models using two predictors. However, the results are the same; We do not observe any problem regarding multicollinearity.

Another useful and more informative metrics for testing the model is the **Confusion Matrix.**[28] It is the most crucial metric, commonly used to evaluate classification models (Hackerearth, 2020). The Confusion Matrix has the form shown in Table A3.3.1.

	1 (Actual)	0 (Actual)
1 (Predicted)	True Positive	False Positive
0 (Predicted)	False Negative	True Negative

Table A3.3.1: *Confusion Matrix*

Positive is when Y=1, and negative is Y=0. The table show how many of the observations that are correctly predicting Y=1 (true positive) and correctly predicting Y=0 (true negative). Furthermore, false positive is those who are predicted to be positive but actually was negative, and the opposite for false negative (Hackerearth, 2020).

The overall model **accuracy** can be calculated as

$$\frac{\text{True Positives} + \text{True Negatives}}{\text{True Positives} + \text{True Negatives} + \text{False Positives} + \text{False Negatives}}$$

Sensitivity, also called the **True Positive Rate (TPR)**, indicates how many positive values, out of all the positive values, have been **correctly predicted** (Hackerearth, 2020). The formula to calculate the true positive rate is

$$\frac{\text{True Positives}}{\text{True Positives} + \text{False Negative}}$$

[28] The function confusionMatrix() from the caret package in R Studio will give you both the matrix and these performance metrics.

Specificity, also called **True Negative Rate (TNR),** indicates how many negative values, out of all the negative values, that have been correctly predicted (Hackerearth, 2020). This is calculated as

$$\frac{\text{True Negative}}{\text{True Negative} + \text{False Positive}}$$

The **precision** of the model is how many values, out of all the predicted positive values, are actually positive, given by the following equation (Hackerearth, 2020):

$$\frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

The Confusion Matrix belonging to our model is presented in Figure A3.2.1 below. In order to do make this, we randomly split our data into a test and a training set. The training data consists of 80 % of the data and is used to perform the logistic regression, and thereby providing estimates of the coefficients. The rest is test data (20 %) where the estimates are used to evaluate the model. Hence, the confusion matrix shows what we have predicted using our estimates on the test data, and what the actual values in the test data are. In order to say if the prediction is correct or not, we must define a limit for the predicted value to be considered as Y=1. As we have seen the probabilities of our models are 60 % at the highest, and also 20 % for certain bid characteristics, we define the predicted values to be 1 if they are 40 % or above.

We observe the amount of correctly predicted 0´s are very high, only 348 false positive (positive when they in fact were negative). The amount of correctly predicted positives is much less. However, the amount of positives is also less. We are predicting 7 081 of the one´s to be zeros.

	1 (Actual)	0 (Actual)
1 (Predicted)	510	348
0 (Predicted)	7 081	65 266

Table A3.3.2: *Confusion Matrix Belonging to our Model*

The metrics calculated from the Confusion Matrix is presented in Table A3.3.3.

Metrics	Value
Accuracy	89.85 %
Sensitivity/ TPR	6.72 %
Specificity/ TNR	99.47 %
Precision	59.44 %

Table A3.3.3: *Key Measures*

The overall model accuracy is quite high, 89.85 %. However, the sensitivity or True Positive Rate (TPR) is very low, only 6.72 %. This indicates that the model predicts the zeroes better than when y=1. This is reflected in the Specificity or the True Negative Rate, which is almost 100 %. Nevertheless, the Precision is higher, indicating that a large amount of the predicted positive values, are in fact positive. All in all, we have an acceptable model fit, especially as our focus only is to indicate whether certain characteristics of the bid combinations are more likely to result in different outcome than others.

A4 Algorithms in R-Script

In this chapter, the algorithms for the simulation are provided. We have one main algorithm that is used to create the matrices presented in Chapter 4. In addition, it creates the data for the regressions in Chapter 6.3. This is provided in part A4.1 below. In part A4.2, the algorithm used to create the data used in Chapter 6.1 and Appendix A1 is provided.

A4.1 Main Algorithm (copied directly from R-script)

```
#Input data
max_p_score <- 10
max_q_score <- 10
weight_p <- 0.4
weight_qq <- 0.6
weight_q <- c(0.3,0.3)
price_threshold <- 150
reserve_price <- 350

### Simulation Possible Bid Combinations ###
quality <- seq(0,10)
price <- seq(100,500,by=100)

# making a list of combinations of price and quality (vectors of
 price and quality inside a list)
comb <- vector("list")

t=1
for (p in price) {
  for (q in quality) {
    for (q2 in quality) {
      comb[[t]] <- c(p,q,q2)
      t=t+1
    }
  }
}


### Scoring rules / Evaluation methods ###
lsrule1 <- function(x) {
  if ( x > price_threshold && x < reserve_price) {
    lsrule1_score <- max_p_score * ((reserve_price -
 x)/(reserve_price - price_threshold)) * weight_p
    return (lsrule1_score)
  } else {
    if (x <= price_threshold) {
```

```
      return(max_p_score*weight_p)
    } else {
      return(0)
    }
  }
}

lsrule2 <- function(x) {
  if (x < reserve_price) {
    lsrule2_score <- max_p_score * ((reserve_price -
 x)/reserve_price) * weight_p
    return(lsrule2_score)
  } else {
    return(0)
  }
}

psrule <- function (x) {
  if (x < reserve_price) {
    ps_score <- max_p_score * (1-(x/reserve_price)^2) * weight_p
    return(ps_score)
  } else {
    return(0)
  }
}

lbrule <- function(x) {    # price-score lowest bid method
  lb_score <-  (max_p_score * min(x)/x) * weight_p
  return(lb_score)
}

q11 <- vector("numeric")
lbruleq <- function(x) {  # tranformation rule normalization of
 quality
  if (max(x)>0 && max(x) < max_q_score) {
    for (i in seq_along(x)) {
      q11[i] <- max_q_score * x[i]/max(x)
    }
    return(q11)
  } else {
    return (x)
  }
}

### TENDER EVALUATION ###
# lists containing vectors where each vector contains total score
 for all suppliers,
```

```
# and each vector is the result of one combination of bids from the
 suppliers #
vectorrule1 <- vector("list")
vectorrule2 <- vector("list")
vectorrule3 <- vector("list")
vectorrule4 <- vector("list")
vectorrule5 <- vector("list")

# List of ranking-vectors for the different methods #
rankrule1 <- vector("list")
rankrule2 <- vector("list")
rankrule3 <- vector("list")
rankrule4 <- vector("list")
rankrule5 <- vector("list")

# Variables for the data frame and regression.
diff_rank_15 <- vector("numeric")
diff_rank_25 <- vector("numeric")
diff_rank_35 <- vector("numeric")
diff_rank_14 <- vector("numeric")
diff_rank_24 <- vector("numeric")
diff_rank_34 <- vector("numeric")
diff_rank_12 <- vector("numeric")
diff_rank_13 <- vector("numeric")
diff_rank_23 <- vector("numeric")
diff_rank_45 <- vector("numeric")

price_t1 <- vector("numeric")
price_t2 <- vector("numeric")

q1_t1 <- vector("numeric")
q1_t2 <- vector("numeric")
q2_t1 <- vector("numeric")
q2_t2 <- vector("numeric")

totalQ_t1 <- vector("numeric")
totalQ_t2 <- vector("numeric")

highest_p_q <- vector("numeric")
highest_one_q <- vector("numeric")
```

```
l=1
for (t1 in comb) {
  for (t2 in comb) {
    t11 <- sum(t1[-1]*weight_q)
    t22 <- sum(t2[-1]*weight_q)
    quality <- c(t11,t22)  # weighted quality score both suppliers

    q1 <- c(t1[2],t2[2])
    q2 <- c(t1[3],t2[3])
    qt1 <- c(lbruleq(q1)[1],lbruleq(q2)[1])  #dim1 and dim2 for
bidder 1
    qt2 <- c(lbruleq(q1)[2],lbruleq(q2)[2])  #dim1 and dim2 for
bidder 2
    # Weightning all the quality criteria and sum for the two
bidders
    # then puts the total score of quality for each bidder into the
vector "q_values"
    q_values <- c(sum(qt1 * weight_q), sum(qt2 * weight_q))

    ls1_values <- c(lsrule1(t1[1]), lsrule1(t2[1]))  # weighted
price score rule 1
    vectorrule1[[l]] <- round(quality + ls1_values,digit=2)
    rankrule1[[l]] <- rank(-vectorrule1[[l]])

    ls2_values <- c(lsrule2(t1[1]), lsrule2(t2[1]))
    vectorrule2[[l]] <- round(quality + ls2_values,digit=2)
    rankrule2[[l]] <- rank(-vectorrule2[[l]])

    ps_values <- c(psrule(t1[1]), psrule(t2[1]))
    vectorrule3[[l]] <- round(quality + ps_values,digit=2)
    rankrule3[[l]] <- rank(-vectorrule3[[l]])

    lb_values <- lbrule(c(t1[1], t2[1]))
    vectorrule5[[l]] <- round(q_values + lb_values,digit=2)
    rankrule5[[l]] <- rank(-vectorrule5[[l]])

    vectorrule4[[l]] <- round(quality + lb_values,digit=2)
    rankrule4[[l]] <- rank(-vectorrule4[[l]])

    ## Storing data to be used in regressions ##
    price_t1[l] <- t1[1]
    price_t2[l] <- t2[1]
    q1_t1[l] <- t1[2]
    q1_t2[l] <- t2[2]
    q2_t1[l] <- t1[3]
    q2_t2[l] <- t2[3]
    totalQ_t1[l] <- q1_t1[l] + q2_t1[l]
```

```
    totalQ_t2[l] <- q1_t2[l] + q2_t2[l]
    highest_p_q[l] <- ifelse(price_t1[l] > price_t2[l] &
 totalQ_t1[l] > totalQ_t2[l]
                              | price_t2[l] > price_t1[l] &
 totalQ_t2[l] > totalQ_t1[l], 1 , 0)

    highest_one_q[l] <- ifelse(q1_t1[l] > q1_t2[l] & q2_t1[l] <
 q2_t2[l]
                                | q2_t1[l] > q2_t2[l] & q1_t1[l] <
 q1_t2[l]
                                | q1_t2[l] > q1_t1[l] & q2_t2[l] <
 q2_t1[l]
                                | q2_t2[l] > q2_t1[l] & q1_t2[l] <
 q1_t1[l] , 1 , 0)

    ifelse(rankrule1[[l]] == rankrule5[[l]], diff_rank_15[l] <- 0,
 diff_rank_15[l] <- 1)
    ifelse(rankrule2[[l]] == rankrule5[[l]], diff_rank_25[l] <- 0,
 diff_rank_25[l] <- 1)
    ifelse(rankrule3[[l]] == rankrule5[[l]], diff_rank_35[l] <- 0,
 diff_rank_35[l] <- 1)

    ifelse(rankrule1[[l]] == rankrule4[[l]], diff_rank_14[l] <- 0,
 diff_rank_14[l] <- 1)
    ifelse(rankrule2[[l]] == rankrule4[[l]], diff_rank_24[l] <- 0,
 diff_rank_24[l] <- 1)
    ifelse(rankrule3[[l]] == rankrule4[[l]], diff_rank_34[l] <- 0,
 diff_rank_34[l] <- 1)

    ifelse(rankrule1[[l]] == rankrule2[[l]], diff_rank_12[l] <- 0,
 diff_rank_12[l] <- 1)
    ifelse(rankrule1[[l]] == rankrule3[[l]], diff_rank_13[l] <- 0,
 diff_rank_13[l] <- 1)
    ifelse(rankrule2[[l]] == rankrule3[[l]], diff_rank_23[l] <- 0,
 diff_rank_23[l] <- 1)

    ifelse(rankrule4[[l]] == rankrule5[[l]], diff_rank_45[l] <- 0,
 diff_rank_45[l] <- 1)

    l=l+1
  }
}

### OUTPUT MATRICES ###

# Score Matrices - to show total scores for the two bidders for each
 bid comibnation #
```

```
matrix_tsrule1 <- matrix(data=vectorrule1,nrow=length(comb),
 ncol=length(comb), byrow = TRUE)
matrix_tsrule2 <- matrix(data=vectorrule2,nrow=length(comb),
 ncol=length(comb), byrow = TRUE)
matrix_tsrule3 <- matrix(data=vectorrule3,nrow=length(comb),
 ncol=length(comb), byrow = TRUE)
matrix_tsrule4 <- matrix(data=vectorrule4,nrow=length(comb),
 ncol=length(comb), byrow = TRUE)
matrix_tsrule5 <- matrix(data=vectorrule5,nrow=length(comb),
 ncol=length(comb), byrow = TRUE)

rownames(matrix_tsrule1) <- as.character(comb)
colnames(matrix_tsrule1) <- as.character(comb)
rownames(matrix_tsrule2) <- as.character(comb)
colnames(matrix_tsrule2) <- as.character(comb)
rownames(matrix_tsrule3) <- as.character(comb)
colnames(matrix_tsrule3) <- as.character(comb)
rownames(matrix_tsrule4) <- as.character(comb)
colnames(matrix_tsrule4) <- as.character(comb)
rownames(matrix_tsrule5) <- as.character(comb)
colnames(matrix_tsrule5) <- as.character(comb)

write.table(matrix_tsrule1, file="matrix_tsrule1.csv")
write.table(matrix_tsrule2, file="matrix_tsrule2.csv")
write.table(matrix_tsrule3, file="matrix_tsrule3.csv")
write.table(matrix_tsrule4, file="matrix_tsrule4.csv")
write.table(matrix_tsrule5, file="matrix_tsrule5.csv")

# Rank Matrices - to show the rankings for the two bidders for
 different bid combinations #
matrix_rankrule1 <- matrix(data=rankrule1,nrow=length(comb),
 ncol=length(comb), byrow = TRUE)
matrix_rankrule2 <- matrix(data=rankrule2,nrow=length(comb),
 ncol=length(comb), byrow = TRUE)
matrix_rankrule3 <- matrix(data=rankrule3,nrow=length(comb),
 ncol=length(comb), byrow = TRUE)
matrix_rankrule4 <- matrix(data=rankrule4,nrow=length(comb),
 ncol=length(comb), byrow = TRUE)
matrix_rankrule5 <- matrix(data=rankrule5,nrow=length(comb),
 ncol=length(comb), byrow = TRUE)

rownames(matrix_rankrule1) <- as.character(comb)
colnames(matrix_rankrule1) <- as.character(comb)
rownames(matrix_rankrule2) <- as.character(comb)
colnames(matrix_rankrule2) <- as.character(comb)
rownames(matrix_rankrule3) <- as.character(comb)
colnames(matrix_rankrule3) <- as.character(comb)
```

```
rownames(matrix_rankrule4) <- as.character(comb)
colnames(matrix_rankrule4) <- as.character(comb)
rownames(matrix_rankrule5) <- as.character(comb)
colnames(matrix_rankrule5) <- as.character(comb)

write.table(matrix_rankrule1, file="matrix_rankrule1.csv")
write.table(matrix_rankrule2, file="matrix_rankrule2.csv")
write.table(matrix_rankrule3, file="matrix_rankrule3.csv")
write.table(matrix_rankrule4, file="matrix_rankrule4.csv")
write.table(matrix_rankrule5, file="matrix_rankrule5.csv")

# 0/1 - Matrices - to show the difference in outcome between two
 rules #
matrix_12 <- matrix(data=diff_rank_12,nrow=length(comb),
 ncol=length(comb), byrow = TRUE)
matrix_13 <- matrix(data=diff_rank_13,nrow=length(comb),
 ncol=length(comb), byrow = TRUE)
matrix_23 <- matrix(data=diff_rank_23,nrow=length(comb),
 ncol=length(comb), byrow = TRUE)
matrix_14 <- matrix(data=diff_rank_14,nrow=length(comb),
 ncol=length(comb), byrow = TRUE)
matrix_24 <- matrix(data=diff_rank_24,nrow=length(comb),
 ncol=length(comb), byrow = TRUE)
matrix_34 <- matrix(data=diff_rank_34,nrow=length(comb),
 ncol=length(comb), byrow = TRUE)
matrix_15 <- matrix(data=diff_rank_15,nrow=length(comb),
 ncol=length(comb), byrow = TRUE)
matrix_25 <- matrix(data=diff_rank_25,nrow=length(comb),
 ncol=length(comb), byrow = TRUE)
matrix_35 <- matrix(data=diff_rank_35,nrow=length(comb),
 ncol=length(comb), byrow = TRUE)
matrix_45 <- matrix(data=diff_rank_45,nrow=length(comb),
 ncol=length(comb), byrow = TRUE)
rownames(matrix_12) <- as.character(comb)
colnames(matrix_12) <- as.character(comb)
rownames(matrix_13) <- as.character(comb)
colnames(matrix_13) <- as.character(comb)
rownames(matrix_23) <- as.character(comb)
colnames(matrix_23) <- as.character(comb)
rownames(matrix_14) <- as.character(comb)
colnames(matrix_14) <- as.character(comb)
rownames(matrix_24) <- as.character(comb)
colnames(matrix_24) <- as.character(comb)
rownames(matrix_34) <- as.character(comb)
colnames(matrix_34) <- as.character(comb)
rownames(matrix_15) <- as.character(comb)
colnames(matrix_15) <- as.character(comb)
```

```
rownames(matrix_25) <- as.character(comb)
colnames(matrix_25) <- as.character(comb)
rownames(matrix_35) <- as.character(comb)
colnames(matrix_35) <- as.character(comb)
rownames(matrix_45) <- as.character(comb)
colnames(matrix_45) <- as.character(comb)

write.table(matrix_12, file="matrix_12.csv")
write.table(matrix_13, file="matrix_13.csv")
write.table(matrix_23, file="matrix_23.csv")
write.table(matrix_14, file="matrix_14.csv")
write.table(matrix_24, file="matrix_24.csv")
write.table(matrix_34, file="matrix_34.csv")
write.table(matrix_15, file="matrix_15.csv")
write.table(matrix_25, file="matrix_25.csv")
write.table(matrix_35, file="matrix_35.csv")
write.table(matrix_45, file="matrix_45.csv")


#### CREATING A DATA FRAME WITH VARIABLES FOR THE REGRESSIONS ###

# Variables created (not all are used in the thesis)
diff_q1 <- abs(q1_t2 - q1_t1)
diff_q2 <- abs(q2_t2 - q2_t1)
diff_totalQ <- abs(totalQ_t2 - totalQ_t1)
t1_highest_totalquality <- ifelse(totalQ_t1 > totalQ_t2, 1,0)
t2_highest_totalquality <- ifelse(totalQ_t2 > totalQ_t1, 1,0)
t1_highest_q1 <- ifelse(q1_t1 > q1_t2,1,0)
t1_highest_q2 <- ifelse(q2_t1 > q2_t2,1,0)
t2_highest_q1 <- ifelse(q1_t1 < q1_t2,1,0)
t2_highest_q2 <- ifelse(q2_t1 < q2_t2,1,0)
equal_totalQ <- ifelse(totalQ_t1==totalQ_t2,1,0)

diff_prices <- abs(price_t1 - price_t2)
is_diff_in_prices <- ifelse(price_t1 == price_t2, 0,1)
equal_prices <- ifelse(price_t1 == price_t2, 1,0)
both_prices_low <- ifelse(price_t1 < 300 & price_t2 < 300,1,0)
both_prices_high <- ifelse(price_t1 > 300 & price_t2 > 300,1,0)
t1_highest_price <- ifelse(price_t1 > price_t2, 1,0)
t2_highest_price <- ifelse(price_t2 > price_t1, 1,0)

## Making the data frame ##
data_alt1 <- data.frame(price_t1, price_t2, q1_t1, q2_t1, q1_t2,
 q2_t2,
                        diff_rank_15, diff_rank_25, diff_rank_35,
 diff_rank_45,
                        diff_rank_14, diff_rank_24, diff_rank_34,
```

```
 diff_rank_12, diff_rank_13, diff_rank_23,
                              diff_prices, equal_prices,
                              totalQ_t1,totalQ_t2, diff_q1, diff_q2,
 diff_totalQ, equal_totalQ,
                              t1_highest_price, t1_highest_totalquality,
 t2_highest_price, t2_highest_totalquality,
                              t1_highest_q1, t1_highest_q2, t2_highest_q1,
 t2_highest_q2,
                              both_prices_high, both_prices_low,
 is_diff_in_prices, highest_p_q, highest_one_q)


### Adding more variables to the data frame ###

## Quality variables ##
# bidder 1 and 2: low/high q1 and q2
data_alt1$lav_q1_t1 <- ifelse(q1_t1 < 6, 1, 0)
data_alt1$lav_q2_t1 <- ifelse(q2_t1 < 6, 1, 0)
data_alt1$lav_q1_t2 <- ifelse(q1_t2 < 6, 1, 0)
data_alt1$lav_q2_t2 <- ifelse(q2_t2 < 6, 1, 0)

# bidder 1 and 2: low/med/high q1 and q2
data_alt1$low_q1_t1 <- ifelse(q1_t1 < 4, 1, 0)
data_alt1$med_q1_t1 <- ifelse(q1_t1 < 8 & q1_t1 > 3, 1, 0)
data_alt1$high_q1_t1 <- ifelse(q1_t1 > 7, 1, 0)

data_alt1$low_q2_t1 <- ifelse(q2_t1 < 4, 1, 0)
data_alt1$med_q2_t1 <- ifelse(q2_t1 < 8 & q2_t1 > 3, 1, 0)
data_alt1$high_q2_t1 <- ifelse(q2_t1 > 7, 1, 0)

data_alt1$low_q1_t2 <- ifelse(q1_t2 < 4, 1, 0)
data_alt1$med_q1_t2 <- ifelse(q1_t2 < 8 & q1_t2 > 3, 1, 0)
data_alt1$high_q1_t2 <- ifelse(q1_t2 > 7, 1, 0)

data_alt1$low_q2_t2 <- ifelse(q2_t2 < 4, 1, 0)
data_alt1$med_q2_t2 <- ifelse(q2_t2 < 8 & q2_t2 > 3, 1, 0)
data_alt1$high_q2_t2 <- ifelse(q2_t2 > 7, 1, 0)

# bidder 1 and 2: total Q low/med/high
data_alt1$low_Q_t1 <- ifelse(totalQ_t1 < 11, 1,0)
data_alt1$low_Q_t2 <- ifelse(totalQ_t2 < 11, 1,0)

# differences
data_alt1$diff_q1_low <- ifelse(diff_q1 < 6 ,1,0)
data_alt1$diff_q2_low <- ifelse(diff_q2 < 6 ,1,0)

data_alt1$diff_Q_low <- ifelse(diff_totalQ <= 5,1,0)
```

```
data_alt1$diff_Q_rel_low <- ifelse((diff_totalQ >= 6 & diff_totalQ
 <= 10),1,0)
data_alt1$diff_Q_rel_high <- ifelse((diff_totalQ >= 11 & diff_totalQ
 <= 15),1,0)
data_alt1$diff_Q_high <- ifelse(diff_totalQ >= 16,1,0) # one of
 these must be reference group

data_alt1$diff_Q_lav <- ifelse(diff_totalQ <= 6,1,0)
data_alt1$diff_Q_med <- ifelse(diff_totalQ <= 14 & diff_totalQ >=
 7,1,0)
data_alt1$diff_Q_høy <- ifelse(diff_totalQ >= 15,1,0)

data_alt1$diff_Q_under11 <- ifelse(diff_totalQ < 11,1,0)

## Price variables ##
# prices are divided into high/low - level of prices
data_alt1$lav_pt1 <- ifelse(price_t1 < 300, 1, 0)
data_alt1$lav_pt2 <- ifelse(price_t2 < 300, 1, 0)

# prices divided into high/medium/low - level of prices
data_alt1$low_p_t1 <- ifelse(price_t1 < 300, 1, 0)
data_alt1$high_p_t1 <- ifelse(price_t1 > 300, 1, 0)
data_alt1$medium_p_t1 <- ifelse(price_t1 == 300, 1, 0)
data_alt1$low_p_t2 <- ifelse(price_t2 < 300, 1, 0)
data_alt1$high_p_t2 <- ifelse(price_t2 > 300, 1, 0)
data_alt1$medium_p_t2 <- ifelse(price_t2 == 300, 1, 0)

# low/high diff between prices
data_alt1$diff_prices_lav <- ifelse(diff_prices < 300, 1,0)

# high, medium, low diff - diff between prices
data_alt1$diff_prices_low <- ifelse(diff_prices == 100, 1,0)
data_alt1$diff_prices_medium <- ifelse(diff_prices < 400 &
 diff_prices > 100, 1,0)
data_alt1$diff_prices_high <- ifelse(diff_prices == 400, 1,0)


write.csv(data_alt1, file = "Data frame Alternative 1 100-500
 prices.csv")
```

A4.2 Algorithm for the Overview (copied directly from R-script)

```
#Input data
max_p_score <- 10
max_q_score <- 10
weight_p <- 0.4
weight_qq <- 0.6
weight_q <- c(0.3,0.3)
price_threshold <- 150
reserve_price <- 350

### Simulation Possible Bid Combinations ###
quality <- seq(0,10)
price <- seq(100,500,by=100)

# making a list of combinations of price and quality (vectors of
 price and quality inside a list)
comb <- vector("list")

t=1
for (p in price) {
  for (q in quality) {
    for (q2 in quality) {
      comb[[t]] <- c(p,q,q2)
      t=t+1
    }
  }
}


### Scoring rules / Evaluation Methods ###
lsrule1 <- function(x) {
  if ( x > price_threshold && x < reserve_price) {
    lsrule1_score <- max_p_score * ((reserve_price -
 x)/(reserve_price - price_threshold)) * weight_p
    return (lsrule1_score)
  } else {
    if (x <= price_threshold) {
      return(max_p_score*weight_p)
    } else {
      return(0)
    }
  }
}

lsrule2 <- function(x) {
  if (x < reserve_price) {
```

```
    lsrule2_score <- max_p_score * ((reserve_price -
 x)/reserve_price) * weight_p
    return(lsrule2_score)
  } else {
    return(0)
  }
}

psrule <- function (x) {
  if (x < reserve_price) {
    ps_score <- max_p_score * (1-(x/reserve_price)^2) * weight_p
    return(ps_score)
  } else {
    return(0)
  }
}

lbrule <- function(x) {    # price-score lowest bid method
  lb_score <-  (max_p_score * min(x)/x) * weight_p
  return(lb_score)
}

q11 <- vector("numeric")
lbruleq <- function(x) {  # scoringregel normalisering kvalitet
  if (max(x)>0 && max(x) < max_q_score) {
    for (i in seq_along(x)) {
      q11[i] <- max_q_score * x[i]/max(x)
    }
    return(q11)
  } else {
    return (x)
  }
}

### TENDER EVALUATION ###

# lists containing vectors where each vector contains total score
 for all suppliers,
# and each vector is the result of one combination of bids from the
 suppliers #
vectorrule1 <- vector("list")
vectorrule2 <- vector("list")
vectorrule3 <- vector("list")
vectorrule4 <- vector("list")
vectorrule5 <- vector("list")

# List of ranking-vectors from the different combination of bids #
```

```
rankrule1 <- vector("list")
rankrule2 <- vector("list")
rankrule3 <- vector("list")
rankrule4 <- vector("list")
rankrule5 <- vector("list")

# Variables for the data frame and regression.
diff_rank <- vector("integer")

price_t1 <- vector("numeric")
price_t2 <- vector("numeric")

q1_t1 <- vector("numeric")
q1_t2 <- vector("numeric")
q2_t1 <- vector("numeric")
q2_t2 <- vector("numeric")

rule14 <- vector("integer")
rule24 <- vector("integer")
rule34 <- vector("integer")
rule15 <- vector("integer")
rule25 <- vector("integer")
rule35 <- vector("integer")
rule12 <- vector("integer")
rule23 <- vector("integer")
rule13 <- vector("integer")
rule45 <- vector("integer")

one_normalization_diff <- vector("integer")
# 1 if one of the rules compared normalize and the other does not, 0
 otherwise
absolutt_relative_diff <- vector("integer")
# 1 if their is a difference in absolut/relative rule, 0 if the
 rules are the same type of rule

l=1
k=1
for (t1 in comb) {
  for (t2 in comb) {
    t11 <- sum(t1[-1]*weight_q)
    t22 <- sum(t2[-1]*weight_q)
    quality <- c(t11,t22)  # weighted quality score both suppliers

    q1 <- c(t1[2],t2[2])
    q2 <- c(t1[3],t2[3])
    qt1 <- c(lbruleq(q1)[1],lbruleq(q2)[1])  #dim1 and dim2 for
 bidder 1
```

```
    qt2 <- c(lbruleq(q1)[2],lbruleq(q2)[2])  #dim1 and dim2 for
bidder 2
    # Weightning all the quality criteria and sum for the two
bidders
    # then puts the total score of quality for each bidder into the
vector "q_values"
    q_values <- c(sum(qt1 * weight_q), sum(qt2 * weight_q))

    ls1_values <- c(lsrule1(t1[1]), lsrule1(t2[1]))  # weighted
price score rule 1
    vectorrule1[[l]] <- round(quality + ls1_values,digit=2)
    rankrule1[[l]] <- rank(-vectorrule1[[l]])

    ls2_values <- c(lsrule2(t1[1]), lsrule2(t2[1]))
    vectorrule2[[l]] <- round(quality + ls2_values,digit=2)
    rankrule2[[l]] <- rank(-vectorrule2[[l]])

    ps_values <- c(psrule(t1[1]), psrule(t2[1]))
    vectorrule3[[l]] <- round(quality + ps_values,digit=2)
    rankrule3[[l]] <- rank(-vectorrule3[[l]])

    lb_values <- lbrule(c(t1[1], t2[1]))
    vectorrule4[[l]] <- round(quality + lb_values,digit=2)
    rankrule4[[l]] <- rank(-vectorrule4[[l]])

    vectorrule5[[l]] <- round(q_values + lb_values,digit=2)
    rankrule5[[l]] <- rank(-vectorrule5[[l]])

    ## Storing the data ##
    # comparison 1&5 #
    ifelse(rankrule1[[l]] == rankrule5[[l]], diff_rank[k] <- 0,
diff_rank[k] <- 1)
    price_t1[k] <- t1[1]
    price_t2[k] <- t2[1]
    q1_t1[k] <- t1[2]
    q1_t2[k] <- t2[2]
    q2_t1[k] <- t1[3]
    q2_t2[k] <- t2[3]

    rule15 [k] <- 1
    rule25 [k] <- 0
    rule35 [k] <- 0
    rule14 [k] <- 0
    rule24 [k] <- 0
    rule34 [k] <- 0
    rule12 [k] <- 0
    rule23 [k] <- 0
```

```
    rule13 [k] <- 0
    rule45 [k] <- 0

    one_normalization_diff[k] <- 1
    absolutt_relative_diff[k] <- 1

    # comparison 2&5 #
    ifelse(rankrule2[[l]] == rankrule5[[l]], diff_rank[k+1] <- 0,
diff_rank[k+1] <- 1)
    price_t1[k+1] <- t1[1]
    price_t2[k+1] <- t2[1]
    q1_t1[k+1] <- t1[2]
    q1_t2[k+1] <- t2[2]
    q2_t1[k+1] <- t1[3]
    q2_t2[k+1] <- t2[3]

    rule15 [k+1] <- 0
    rule25 [k+1] <- 1
    rule35 [k+1] <- 0
    rule14 [k+1] <- 0
    rule24 [k+1] <- 0
    rule34 [k+1] <- 0
    rule12 [k+1] <- 0
    rule23 [k+1] <- 0
    rule13 [k+1] <- 0
    rule45 [k+1] <- 0

    one_normalization_diff[k+1] <- 1
    absolutt_relative_diff[k+1] <- 1

    # comparison 3&5 #
    ifelse(rankrule3[[l]] == rankrule5[[l]], diff_rank[k+2] <- 0,
diff_rank[k+2] <- 1)
    price_t1[k+2] <- t1[1]
    price_t2[k+2] <- t2[1]
    q1_t1[k+2] <- t1[2]
    q1_t2[k+2] <- t2[2]
    q2_t1[k+2] <- t1[3]
    q2_t2[k+2] <- t2[3]

    rule15 [k+2] <- 0
    rule25 [k+2] <- 0
    rule35 [k+2] <- 1
    rule14 [k+2] <- 0
    rule24 [k+2] <- 0
    rule34 [k+2] <- 0
    rule12 [k+2] <- 0
```

```
    rule23 [k+2] <- 0
    rule13 [k+2] <- 0
    rule45 [k+2] <- 0

    one_normalization_diff[k+2] <- 1
    absolutt_relative_diff[k+2] <- 1

    # comparison 1&4 #
    ifelse(rankrule1[[l]] == rankrule4[[l]], diff_rank[k+3] <- 0,
diff_rank[k+3] <- 1)
    price_t1[k+3] <- t1[1]
    price_t2[k+3] <- t2[1]
    q1_t1[k+3] <- t1[2]
    q1_t2[k+3] <- t2[2]
    q2_t1[k+3] <- t1[3]
    q2_t2[k+3] <- t2[3]

    rule15 [k+3] <- 0
    rule25 [k+3] <- 0
    rule35 [k+3] <- 0
    rule14 [k+3] <- 1
    rule24 [k+3] <- 0
    rule34 [k+3] <- 0
    rule12 [k+3] <- 0
    rule23 [k+3] <- 0
    rule13 [k+3] <- 0
    rule45 [k+3] <- 0

    one_normalization_diff[k+3] <- 0
    absolutt_relative_diff[k+3] <- 1

    # comparison 2&4 #
    ifelse(rankrule2[[l]] == rankrule4[[l]], diff_rank[k+4] <- 0,
diff_rank[k+4] <- 1)
    price_t1[k+4] <- t1[1]
    price_t2[k+4] <- t2[1]
    q1_t1[k+4] <- t1[2]
    q1_t2[k+4] <- t2[2]
    q2_t1[k+4] <- t1[3]
    q2_t2[k+4] <- t2[3]

    rule15 [k+4] <- 0
    rule25 [k+4] <- 0
    rule35 [k+4] <- 0
    rule14 [k+4] <- 0
    rule24 [k+4] <- 1
    rule34 [k+4] <- 0
```

```
    rule12 [k+4] <- 0
    rule23 [k+4] <- 0
    rule13 [k+4] <- 0
    rule45 [k+4] <- 0

    one_normalization_diff[k+4] <- 0
    absolutt_relative_diff[k+4] <- 1

    # comparison 3&4 #
    ifelse(rankrule3[[l]] == rankrule4[[l]], diff_rank[k+5] <- 0,
diff_rank[k+5] <- 1)
    price_t1[k+5] <- t1[1]
    price_t2[k+5] <- t2[1]
    q1_t1[k+5] <- t1[2]
    q1_t2[k+5] <- t2[2]
    q2_t1[k+5] <- t1[3]
    q2_t2[k+5] <- t2[3]

    rule15 [k+5] <- 0
    rule25 [k+5] <- 0
    rule35 [k+5] <- 0
    rule14 [k+5] <- 0
    rule24 [k+5] <- 0
    rule34 [k+5] <- 1
    rule12 [k+5] <- 0
    rule23 [k+5] <- 0
    rule13 [k+5] <- 0
    rule45 [k+5] <- 0

    one_normalization_diff[k+5] <- 0
    absolutt_relative_diff[k+5] <- 1

    # comparison 1&2 #
    ifelse(rankrule1[[l]] == rankrule2[[l]], diff_rank[k+6] <- 0,
diff_rank[k+6] <- 1)
    price_t1[k+6] <- t1[1]
    price_t2[k+6] <- t2[1]
    q1_t1[k+6] <- t1[2]
    q1_t2[k+6] <- t2[2]
    q2_t1[k+6] <- t1[3]
    q2_t2[k+6] <- t2[3]

    rule15 [k+6] <- 0
    rule25 [k+6] <- 0
    rule35 [k+6] <- 0
    rule14 [k+6] <- 0
    rule24 [k+6] <- 0
```

```
    rule34 [k+6] <- 0
    rule12 [k+6] <- 1
    rule23 [k+6] <- 0
    rule13 [k+6] <- 0
    rule45 [k+6] <- 0

    one_normalization_diff[k+6] <- 0
    absolutt_relative_diff[k+6] <- 0

    # comparison 1&3 #
    ifelse(rankrule1[[l]] == rankrule3[[l]], diff_rank[k+7] <- 0,
diff_rank[k+7] <- 1)
    price_t1[k+7] <- t1[1]
    price_t2[k+7] <- t2[1]
    q1_t1[k+7] <- t1[2]
    q1_t2[k+7] <- t2[2]
    q2_t1[k+7] <- t1[3]
    q2_t2[k+7] <- t2[3]

    rule15 [k+7] <- 0
    rule25 [k+7] <- 0
    rule35 [k+7] <- 0
    rule14 [k+7] <- 0
    rule24 [k+7] <- 0
    rule34 [k+7] <- 0
    rule12 [k+7] <- 0
    rule23 [k+7] <- 0
    rule13 [k+7] <- 1
    rule45 [k+7] <- 0

    one_normalization_diff[k+7] <- 0
    absolutt_relative_diff[k+7] <- 0

    # comparison 2&3 #
    ifelse(rankrule2[[l]] == rankrule3[[l]], diff_rank[k+8] <- 0,
diff_rank[k+8] <- 1)
    price_t1[k+8] <- t1[1]
    price_t2[k+8] <- t2[1]
    q1_t1[k+8] <- t1[2]
    q1_t2[k+8] <- t2[2]
    q2_t1[k+8] <- t1[3]
    q2_t2[k+8] <- t2[3]

    rule15 [k+8] <- 0
    rule25 [k+8] <- 0
    rule35 [k+8] <- 0
    rule14 [k+8] <- 0
```

```
    rule24 [k+8] <- 0
    rule34 [k+8] <- 0
    rule12 [k+8] <- 0
    rule23 [k+8] <- 1
    rule13 [k+8] <- 0
    rule45 [k+8] <- 0

    one_normalization_diff[k+8] <- 0
    absolutt_relative_diff[k+8] <- 0

    # comparison 4&5 #
    ifelse(rankrule5[[l]] == rankrule4[[l]], diff_rank[k+9] <- 0,
 diff_rank[k+9] <- 1)
    price_t1[k+9] <- t1[1]
    price_t2[k+9] <- t2[1]
    q1_t1[k+9] <- t1[2]
    q1_t2[k+9] <- t2[2]
    q2_t1[k+9] <- t1[3]
    q2_t2[k+9] <- t2[3]

    rule15 [k+9] <- 0
    rule25 [k+9] <- 0
    rule35 [k+9] <- 0
    rule14 [k+9] <- 0
    rule24 [k+9] <- 0
    rule34 [k+9] <- 0
    rule12 [k+9] <- 0
    rule23 [k+9] <- 0
    rule13 [k+9] <- 0
    rule45 [k+9] <- 1

    one_normalization_diff[k+9] <- 0
    absolutt_relative_diff[k+9] <- 0

    l=l+1
    k=k+10
  }
}

## Making the data frame ##
data_alt2 <- data.frame(price_t1, price_t2, q1_t1, q2_t1, q1_t2,
 q2_t2, diff_rank, rule15, rule25, rule35, rule14, rule24, rule34,
 rule12, rule13, rule23, rule45, one_normalization_diff,
 absolutt_relative_diff)

write.csv(data_alt2, file = "Data frame Alternative overview.csv")
```