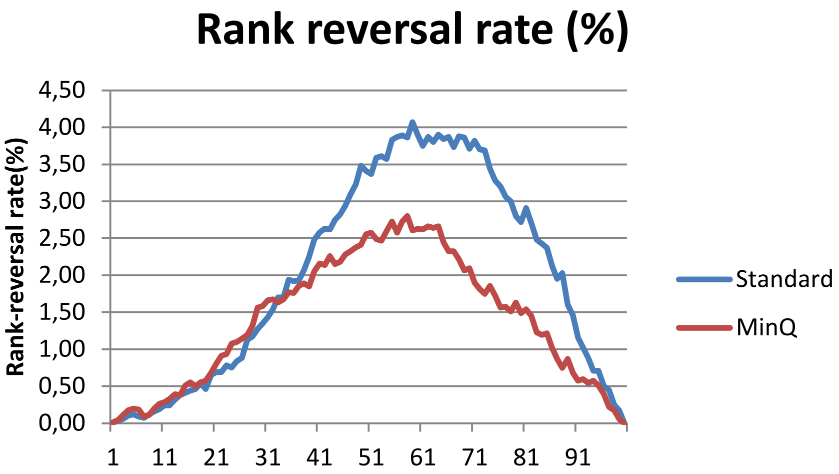

Characteristics of rank-reversal

Yoran Nijenhuis

University of Twente

Uitvoeringsorganisatie
Bedrijfsvoering Rijk
Ministerie van Binnenlandse Zaken en Koninkrijksrelaties

Supervisors:

Fredo Schotanus
Jan Telgen

UNIVERSITY OF TWENTE.

Preamble

In front of you lies the bachelor thesis *Characteristics of rank-reversal.* This report includes research and conclusion about the usage of relative scoring methods in tenders in the Dutch public sector and the factors which influence rank-reversal: a term closely related and associated with relative scoring methods. During most of 2017 I worked on the preparation, writing and required research for this report which has been written as final assignment for the bachelor Industrial Engineering and Management at the University of Twente.

This main research question of this report came on my path by a short introduction by Jan Telgen. By this I became interested in the world of procurement and tendering on which I learnt a lot during my time working on this thesis. Another person from whom I learned a lot about the way tenders are conducted is Richard Lennartz, CEO of UBR|HIS. I would like to thank him for his openness and willingness to share knowledge about relative scoring and allowing for data gathering. The latter was conducted together with my fellow student Gijsbert van den Engh and under supervision of Dick de Waart.

During the whole process Fredo Schotanus has been very supportive as supervisor. I would like to think him for this and especially for helping me during the period I was struggling with my research.

Yoran Nijenhuis,

Enschede, November the 30th 2017

Management summary

Research motivation

In recent years, tender procedures for selecting the best supplier in the Netherlands have been using a best value approach more and more often. Under Dutch law it is now almost always compulsory to evaluate suppliers by the *Economically Most Advantageous tender* method. In practice this implies evaluating suppliers on both price and quality or on quality alone. This can be done by using relative and/or absolute formulas. Relative scoring is used often for the price component; for determining the scores on price suppliers are judged towards each other with the lowest price of all being the reference. Evaluating in this way opens up the possibility of rank-reversal. Rank-reversal is a changed order after removal or entrance of a supplier. If the number of participants changes there is the possibility of a new reference price. Hence, the scores on price could have to be recalculated.

Though the possibility of occurrence of rank-reversal has been proven mathematically and possible problems with this method have been acknowledged, not much research has been conducted on the real world occurrence of rank-reversal. This bachelor assignment aims to fill this gap by answering the following main research question: *How often does rank-reversal occur in practice and what are the characteristics and situations under which it occurs?* Supporting sub questions focused on the motives for using relative scoring methods and its (dis)advantages, a what-if analysis of historic tenders and a simulation to test different situations.

Methodology

Data of historic tenders was gathered at UBR|HIS in The Hague, this yielded 252 governmental tenders. Additionally, 51 tenders where available from a previous bachelor report and tender support platform Negometrix. These historic tenders had information about the number of participating suppliers, the weight for both quality and price, the number of sub-criteria, the obtained quality scores and the offered prices.

In the what-if analysis the supplier with the lowest price was removed from all tenders in our dataset to see whether rank-reversal could have occurred. For the simulation the historic data on quality and the number of suppliers was used to derive data distributions which could be used as input. The simulation allowed changing the number of to be generated tenders, the weight for price and quality, whether a minimum quality threshold was applied and if the latter is the case how high this should be. Additionally, the number of suppliers in a tender and the standard deviation of the tender could be changed.

Besides the two analyses to retrieve rank-reversal rates for different situations a literature study was conducted to find motives for using relative scoring methods. In addition the found advantages and disadvantages of the method where stated.

Results

Relative scoring methods are mainly used due to its easiness and the fact that no predefined scoring tables are required. The what-if analysis showed rank-reversal would have happened in one out of fifty real world cases after removal of the supplier with the lowest bid in all analysed tenders. The simulation showed that the rank-reversal rate follows are parabolic pattern with rates converging to zero at low weights for either price or quality as can be seen in Figure A.

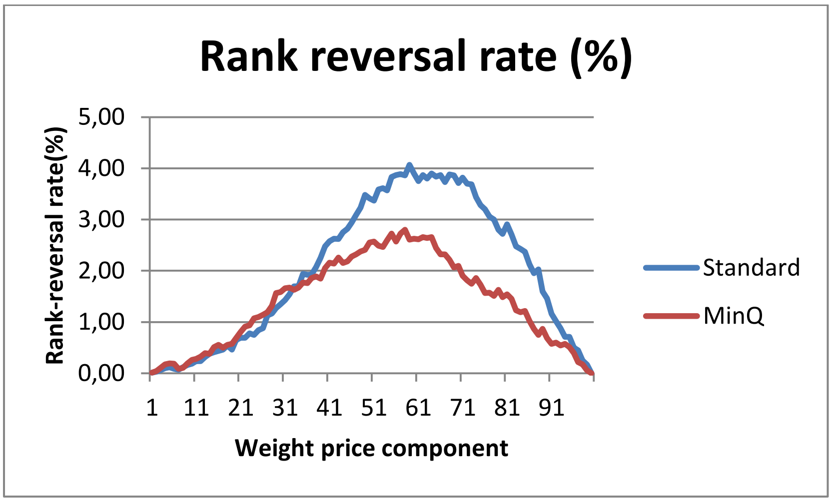

Figure A: Rank-reversal rates

The highest rate was 4.07%, with comparable values for price weights between forty and sixty percent. In most tenders the weight of the price criterion falls in this range. The appliance of a minimum quality threshold reduced the rank-reversal rate, but the peak still lied at 2.80%.

Rank-reversal rates are close to zero towards the edges of graph. In these cases either quality or price is of high importance. At high price weights the tender procedure almost follows the lowest price rule. If the supplier with the lowest price is removed from the tender, the supplier with the second lowest price almost always wins the tender as barely any points can be scored on quality. A same line of reasoning is applicable for high quality weights. Therefore the rank-reversal converges toward zero at both ends of graph.

When more suppliers participate in a tender the rank-reversal rate increases. An increase was also found by increasing the standard deviation of the bids leading to more variance between offers.

Further research can focus on finding a more appropriate distribution for the submitted offers, the impact of entrance of supplier and whether other relative scoring methods yield comparable results and patterns.

As the mathematically proven possibility of rank-reversal occurrence is confirmed by both of the conducted analyses, it is now up to contracting authorities to determine whether they want to change the rules and methods currently in place to evaluate suppliers in a tender. As there are strong options both in favour as against relative scoring, this may become a challenging process.

Chapter 1: Introduction

During a tender procedure to determine the best bid, many different rules can be used to score suppliers on selected criteria. First of all it is decided under which selection approach the tender falls. A choice has to be made whether to award solely on price or to also take quality into account. Awarding scores can be done in an absolute or relative way or a combination of these two. In case of a relative method the supplier's scores depend on the bids of other suppliers; scores on such criterion are calculated relative to the best price or quality. Therefore, the entrance or removal of a supplier with the best score on price or quality possibly influences the scores of all other suppliers. Additionally, relative scoring could give a lesser reckoning on the bidder and could be prone to manipulative bids. In this chapter relative scoring methods, the phenomena rank-reversal and the real world impact will first be introduced. Secondly, the research goal and the corresponding research questions will be stated.

1.1 Relative scoring

In recent years procurement has shifted towards adopting a *best value* approach or *Economically Most Advantageous tender*; bids are both evaluated on price and quality (Dimitri, 2013). Under the Dutch Aanbestedingswet (Article 2.114, 2016) using best price/quality ratios is almost always compulsory for purchasing governmental departments or agencies. This however does not mean all tenders are evaluated in the same manner as multiple scoring methods fall under the EMAT definition. Under the tender law (Aanbestedingswet 2012, Article 2.114, 2016) the following three selection methods imply an EMAT-tender:

1. Best price-quality ratio
2. Lowest total costs based on the cost efficiency of the whole life cycle
3. Lowest price

This report will focus on tenders in whom suppliers are scored on both quality and price and therefore fall under the first definition of EMAT. Whereas quality can be scored in a more systematic way based on a predetermined rule set, doing this for price can difficult. Transforming prices into points is especially difficult when the expected price range is unclear. To counter these difficulties relative scoring is applied; in the Netherlands about half of the public tenders use a relative scoring method.

To evaluate bids in a relative way, many different methods can be used. In this research the focus is on a relative method widely used in practice, in which bids are evaluated on the best submitted bid. The applied formula is specified in Equation 1.

$$Score_i = Score_{max} * Price_{best} / \ Price_i \ (1)$$

As the price-component of a tender usually consists of only one criterion, $Score_{max}$ represents the weight factor for price. This factor represents how much points of a tender can be obtained based on the offered price.

Supplier	Price	Score price
A	€ 1500	30
B	€ 2000	22.5
C	€ 5000	9

Table 1: Calculation of price scores in a relative way

Table 1 shows an example in which the price component is scored using the formula of Equation 1. The maximum score which can be obtained for price is 30 points; supplier A has the lowest price of all bidders and therefore receives the maximum number of points. All other suppliers' prices are evaluated relative to the bid of supplier A. Hence, supplier B obtains 30*€ 1500/€ 2000 = 22.5 points and supplier C 30*€ 1500 / €5000 = 9 points. The scores and prices are plotted against each other in Figure 1.

Figure 1: Plot of the offered price against the obtained scores

Scores calculated using Equation 1 roughly follows an exponential distribution. Therefore, scores will convert towards zero if there is a large deviation towards the lowest price. Other relative methods mostly follow a linear pattern, sometimes using a price threshold as minimum to obtain scores.

1.2 Rank-reversal

When relative scoring is applied the entrance or removal of one supplier could influence the outcome of tender. The scores on the relative criterion have to be recalculated when a supplier with the lowest price offer or best quality enters or withdraws from the tender. Only in this scenario the removal or entrance of a certain supplier influences the total score and therefore possibly the tender outcome.

Suppli-er	Price	Price score	Quality score	Total
A	**1700**	**41.2**	**42**	**83.2**
B	1400	50	30	80
C	2200	31.8	50	81.8

Entrance

Removal

Suppli-er	Price	Price score	Quality score	Total
A	1700	23.5	42	65.5
B	1400	28.6	30	58.6
C	**2200**	**18.2**	**50**	**68.2**
D	800	50	10	60

Figure 2: Fictional tender in which rank-reversal occurs

The fictional tender in Figure 2 shows a situation in which rank-reversal occurs. Price and quality are assigned the same weight of fifty percent and to score suppliers on price the formula in Equation 1 is applied. In the situation on the left hand side three suppliers have put in a bid, supplier B has lowest price and is assigned the maximum fifty points on that criterion. On the right supplier D has entered the tender. Though this supplier is non-competitive because of its low quality, its low price of 800 influences the outcome. Since the entered supplier has the lowest price, all scores have to be re-calculated. Result of the entrance of supplier D is a new winner in supplier C instead of supplier A: rank-reversal has occurred.

Though rank-reversal could occur at all positions, it is most problematic if it involves the winner of the tender. In that case, the (non-)participation of one supplier could partly determine which supplier wins the contract. As tenders could involve large contracts, rank-reversal could lead to a large amount of additional costs for the contractor and arbitrary results. Rank-reversal in the public sector sometimes occurs afterwards as it is found a supplier did not submit all required deliverables or a judge rules a bid manipulative or unrealistic. Entrance of a supplier in the second running of a tender opens up the possibility of collusion as bids can be tailored to the offers of the first running.

1.2.1 Possibility of occurrence

By using relative methods, the scoring rule itself guarantees that there is the possibility of rank-reversal as scores are interdependent. However, whether and when rank-reversal can or will occur will always depend on the combination of the used scoring mechanism and the submitted offers (Telgen & Timmer, 2016, p.3).

1.3 Real world impact

As stated earlier in this introduction relative scoring is a widely used method in governmental tender procedures. In the majority of call for tenders at UBR|HIS relative scoring methods are used for determining the suppliers' scores on the price criterion. UBR|HIS, short for *Uitvoeringsorganisatie Bedrijfsvoering Rijk | Haagse Inkoop Samenwerking,* is the primary contractor for managing procurement of six governmental departments. Additionally, they are responsible for purchasing ICT, cleaning and mobility services for the government as a whole (UBR|HIS, 2017).

The government itself states that they should act in a righteous way (Rijksoverheid, 2015, p.10). It must be seen in which situation scoring methods, whether absolute or relative, fulfil this requirement (Planjer, L. & Lennartz, R., Rechtvaardig gunnen is de kern, om het even of het via absolute of relative beoordeling is, February 14 2017). As the existence of the ranking paradox is acknowledged and most tenders managed by UBR|HIS use relative scoring, the real world impact of relative scoring will be analysed. For gaining a good view on the real world occurrence of rank-reversal UBR|HIS gave permission to gather historic tender data. This yielded around 250 tenders which were open to suppliers from the whole European Union. These tenders have a contract value of at least €134,000 depending on the type of contract (The European Parliament and the Council of the European Union, 2014, Article 4c).

As contracts can be worth millions of euros selecting the right supplier is of high importance. A wrong choice, possibly influenced by the applied scoring method, can lead to higher cost and a lower service level.

1.4 Available data

Besides 250 governmental tenders, fifty other tenders were available for analysis. Almost thirty of these were obtained from Negometrix, an online portal facilitating the submission of tenders. The remaining tenders were obtained from a previous bachelor report also focussing on relative scoring (Merckel, 2015, p.36-41).

1.5 Research goal

Though recent research has focussed under which conditions rank-reversal can and will occur (Telgen & Timmer, 2016), not much is known about how often rank-reversal really occurs. Since relative scoring methods are applied in a majority of governmental tenders with contracts worth millions of euros, the occurrence of rank-reversal on a large scale is undesirable. To gain a good view on the appropriateness of relative scoring, the research goal is to identify the real world occurrence of rank-reversal and the characteristics under which it occurs.

1.6 Research questions

Main question: *How often does rank-reversal occur in practice and what are the characteristics and situations under which it occurs?*

To help answering the main question three research questions are formulated and addressed.

1.) *What are the motives for the usage of relative scoring and what are the advantages and disadvantages of using this method?*
    As relative scoring is applied at large scales there should be a valid line of reasoning supporting the usage of them. Additionally, to make conclusions about the usage of a method the advantages and disadvantages should be clear in advance.
2.) *How often does rank-reversal occur: is it more than a theoretical problem?*
    By analysing the historic data it can be found whether there would be a different ranking and winner in a tender when the supplier with the lowest price would have withdrawn or been discarded. Repeating this procedure for all tenders leads to a rank-reversal rate based on all available historic tenders.
3.) *What are the characteristics under which rank-reversal occurs?*
    By changing parameters like the minimum score on quality and the ratio between quality and price, it can be found under which characteristics the occurrence rank-reversal is more likely

to happen. Additionally, the influence of sector and the number of criteria and bidders can be analysed.

1.7 Report Structure

Each of the three research questions will be addressed in its own chapter. In addition to this the used methodology will be explained and the most interesting results discussed before coming to the main conclusions. On a chapter-by-chapter basis this report has the following structure:

Chapter 1: Introduction
Chapter 2: Methodology
Chapter 3: Motives for using relative scoring methods
Chapter 4: Occurrence of rank-reversal
Chapter 5: Characteristics of rank-reversal
Chapter 6: Discussion of results
Chapter 7: Conclusion

Chapter 2: Methodology

In this chapter the used research approach and methodology will be described. In addition the steps taken to prepare and verify the gathered data will be discussed.

2.1 Research approach

Each research question will be discussed in a separate chapter. The first sub-question was answered using a literature review. The real world occurrence of rank-reversal was addressed by creating a what-if analysis in Excel with the help of Visual Basics for Applications. Lastly, to see which characteristics influence rank-reversal a simulation study was conducted by once again using VBA.

2.2 Research Methodology

In order to answer sub-questions two and three a similar method was used as both focus on the occurrence of rank-reversal. In both situations the outcome of a tender was evaluated twice. First in the initial situation and a second time after a supplier entered or withdrew from a tender.

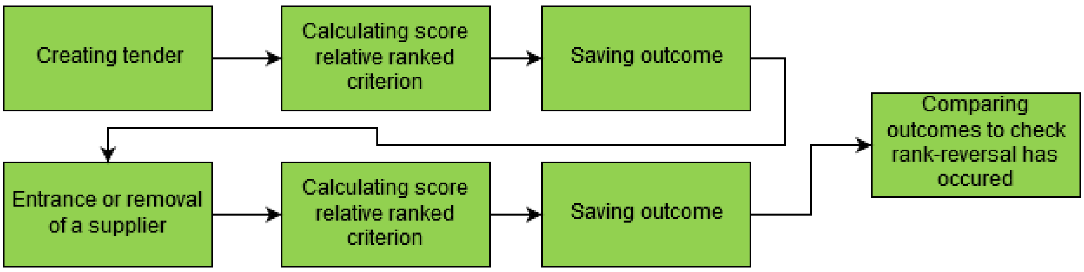

Figure 3: Flow-chart of the simulation process

Figure 3 shows the steps taken in the process. After removing the bid with the lowest price all scores on the relative criterion must be recalculated. It must be seen whether this is also the case when a new supplier enters; the scores only have to be recalculated if the supplier entering the tenders has the lowest price of all.

In order to see whether rank-reversal did occur it must be known which suppliers finished first and second in the original situation. The second is place is stored, because if the supplier with the lowest price also wins the tender and is later discarded you expect the number two to win in the revised situation. Therefore the winning supplier of the revised situation is compared to both the winner and runner-up of the original tender.

2.3 Data preparation

The answering of research questions two and three both required parts of the same dataset including the governmental as well as non-governmental tenders. In order to use the data for analysis most of the tender data had to be verified and/or standardised.

In order to analyse the data a template which could be used for both datasets was created. The used template for standardisation had fifteen categories, not yet including the scores on quality and the offered prices, which were used to store data. After the data collection some of these proved to be irrelevant or negligible. The following categories were seen as relevant for the scope of the project:

- CPV-codes
- Maximum number of points on quality
- Maximum number of points on price
- Number of sub-criteria quality
- Number of sub-criteria price
- Minimum score on quality
- Used scoring method

2.3.1 CPV-codes

CPV, Common Procurement Vocabulary, is a European classification system used for public works, deliveries and services (CPV-codes, n.d.). It consists of 9 digits, each on different level. The first digit shows the main category in which the call for tender is conducted. By including more digits the category becomes more specific. Tenders can fall under multiple CPV-codes when the required works or services have a broad variety of applications. Analyses concerning CPV-codes will only be conducted at the highest level, thus taken only the first digit into account, or at level two when there is a large number of tenders with a recurring CPV-code. By only taking the first digit into account, sector based conclusions can be made. When looking further than the first digit the analyses become very specific and could yield invalid conclusions as the number of tenders in the sub-set becomes smaller. A description of the CPV-codes at level one can be found below in Table 2.

First digit	Description
0	Agriculture, forestry and fishery products
1	Ores and minerals; electricity, gas and water
2	Food products, beverages and tobacco; textiles, apparel and leather products
3	Other transportable goods, except metal products, machinery and equipment
4	Metal products, machinery and equipment
5	Constructions and construction services
6	Distributive trade services; accommodation, food and beverage serving services; transport services; and utilities distribution services
7	Financial and related services; real estate services; and rental and leasing services
8	Business and production services
9	Community, social and personal services

Table 2: CPV-codes with its corresponding categories

2.3.2 Maximum number of points

In most tenders using the best value for money approach, offers are both evaluated on price and quality. For both of the two criteria a maximum number of points can be earned. Together these two values form the relative weight factor between quality and weight. For example when 700 points can be earned on quality and 300 on price, the weight factor for quality is 700/(700+300)=70%. Though the factor itself is an important variable, the maximum number of points to be earned on both criteria is stored. This was done to make indexation of quality scores possible.

Compared to the weight factor, quality scores are an order of magnitude higher in most tenders. For example the weight of quality is 70%, but in total 700 points can be earned on quality. Most of these order of magnitude differences are erased by multiplying or dividing one of the variables by ten. However, sometimes rather arbitrary scores which require more calculations like a 1720-430 split are chosen.

2.3.3 Number of sub-criteria

In many tenders quality is evaluated on multiple sub-criteria. Sometimes the price component is also split in multiple sub-criteria; each of these possibly being evaluated in a relative manner. Additionally, the number of sub-criteria for quality can give an indication of how much scores between suppliers can differ.

2.3.4 Quality threshold

In multiple tenders a quality threshold is applied. In these cases a two-stop approach is used for evaluating each bidder. First of all suppliers are scored on the quality components. If this score is under a pre-determined threshold, mostly 60 or 65 percent of the total of number of points available for quality, the supplier is put aside. If a supplier satisfies the quality minimum, it will be evaluated on price. At this stage only the prices of the suppliers satisfying the minimum quality threshold will be taken into account. Hence, it could be the case that the base price for the relative scoring is not the lowest of all submitted bids. The whole procedure is visualised below in the flowchart in Figure 4.

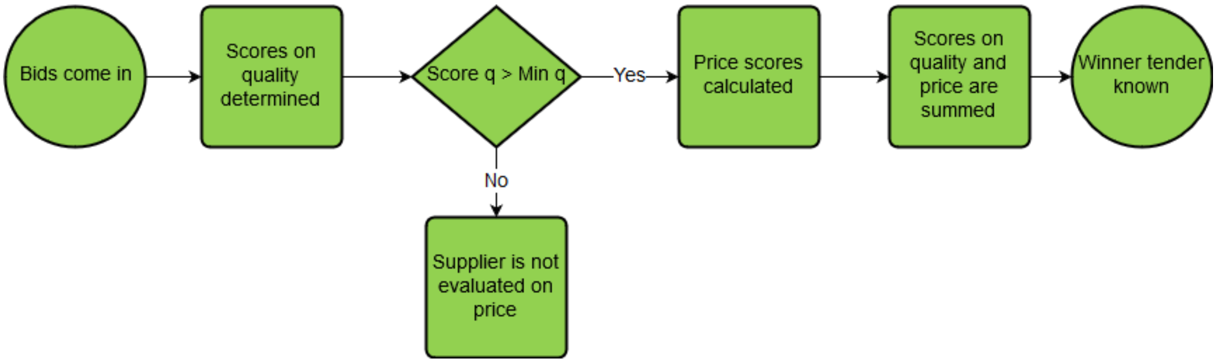

Figure 4: Tender evaluation when a quality threshold is applied

2.3.5 Applied scoring methods

Though most gathered tenders use the formula in Equation 1 to evaluate suppliers on price, other methods can also be found in the data set. These different methods should be treated carefully when analysing certain situations. Suppliers put in bids based on the used scoring method; should another method have been applied they would probably have put in a different offer. When deriving data distributions prices submitted under other circumstances therefore should be excluded. For more general variables like the weight factor and the number of sub-criteria tenders using different scoring rules can however be used.

2.3.6 Data differences

The tenders gathered at HIS have more data richness than the other tenders. The base data each tender has includes the weight factor between quality and price and the scores for each supplier on price and quality. Therefore, for each tender it can be calculated who has won the tender and what happens when the bid with the lowest price is discarded. However, the non-HIS tenders have no information on the sector, tender type and whether a minimum quality threshold was applied.

In the analysis a distinction was made between the tenders gathered at HIS and those who were obtained from other sources. By doing this the Haagse Inkoop Samenwerking will have a clear view on the impact of relative scoring on their way of working and can discuss whether actions should be taken.

2.4 Verification

In order to make valid conclusions the dataset had to be verified. A few checks were undertaken to make sure as much as possible of the dataset could be used. First of all, it was checked whether realistic scores were assigned on the quality criteria. The number of points earned cannot be higher than the total number of points available. However, this proved to be the case in 18 tenders where the total number available points were stored as a percentage instead of just the points. These incorrect strings were later corrected after correspondence with HIS.

Secondly, it was found some tender results had only information on the number of obtained points. Though these scores can be derived when the submitted offers and the scoring rules are known, this cannot be done the other way around. By taking an in-depth look at these tenders an assumption could be made on the possible occurrence of rank-reversal. Since this is very time consuming, these price points only tenders are excluded from the analysis.

Chapter 3: Motives for using relative scoring methods

In this chapter the first research question, *What are the main reasons for the usage of relative scoring and what are the advantages and disadvantages of using this method?,* will be discussed. Though relative scoring is seen as a controversial scoring method, it is applied at large scales. To do so you could argue there should be a valid line of reasoning supporting the usage of relative scoring methods. Additionally, to see whether the method is problematic its advantages and disadvantages should be known. The found motives for using relative scoring can be evaluated based on the results of the what-if analysis and simulation.

3.1 Advantages

The main advantage of relative scoring methods is the easiness when it comes to calculating price scores. There is not much work which needs to be done in advance besides determining which method to use and possible formulate restrictions and determine parameters. Afterwards scores on price can be calculated quickly as there are straightforward Excel-templates and formulas available. According to a senior official at the HIS the availability of templates is a main reason why relative scoring is applied in the majority of their call for tenders. Since standard templates for calculating scores and the descriptive document are available, initiating a call for tender can be done quickly. This noteworthy as other scoring methods also have templates available. Hence, this should not be a criterion on which a scoring method can be chosen.

Secondly, relative scoring sort of postpones a budgeting decision. It offers an option when the expected prices are unknown in advance. As scores are judged to each other, no predefined scoring table is required. The usage of relative scoring also yields an advantage for buyers as they do not necessarily need to have market knowledge. When an absolute scoring method is used more knowledge is required as a more precise estimation of the costs must be submitted.

3.2 Disadvantages

One major disadvantage of the usage of relative scoring is the impact which unrealistic offers have. Though strategic bids are allowed as these can be used to gain market share, unrealistic are not. The contracting authority has the possibility to put aside these offers (Aanbestedingswet 2012, Article 2.116, 2016), but only after asking the supplier an explanation on the submitted offer. Excluding a supplier can therefore become a long process for the contracting authority.

Problems can also arise with strategic bids of €0.00. When the descriptive document does not mention that this kind of offers are not allowed, these bids will have to be accepted (ECLI:NL:GHARL:2013:BZ8213, 2013). Mathematically these offers cause problems when using a relative method. For the supplier submitting a bid of zero euros, €0.00 will have to be divided by €0.00 resulting in a division by zero error. As Equation 2 shows, all non-zero offers will result in a score of null on price whatever the weight factor or price may be.

$$Score_i = Score_{max} * \frac{€\ 0.00}{Price_i} = 0\ (2)$$

In an extreme situation closely related suppliers could even determine their bids together to maximise the chances of one of them winning. This is not automatically in contrary to the Dutch Tender and Competition laws (Ongeldige en bijzondere inschrijvingen, n.d.).

According to Cheng (2009) the usage of relative scoring gives the sitting supplier an advantage. Since the new contract will be most likely comparable to the expired one, the supplier servicing the original contract can use its knowledge to optimise its offer. In a tender that is scored on two price criteria, the original supplier has a better insight which bid can yield the most in terms of points and profits. This is especially the case in smaller tenders where a sitting or smaller supplier has a very good insight in the expected order size of the contract.

Though Cheng (2009) clearly states major disadvantages of relative scoring, he questions whether relative scoring should be forbidden. He mentions the large consequences and investments of change and situations in which relative and absolute scoring yields the same results.

Chapter 4: Occurrence of rank-reversal

This chapter will focus on the real word occurrence of rank-reversal. By removing the supplier with the lowest price from all qualified tenders, the incurred risks can be addressed.

4.1 Methodology

In order to see whether rank-reversal would have occurred when the supplier with the lowest price would have been removed, the two-step approach shown in Figure 5 has been used.

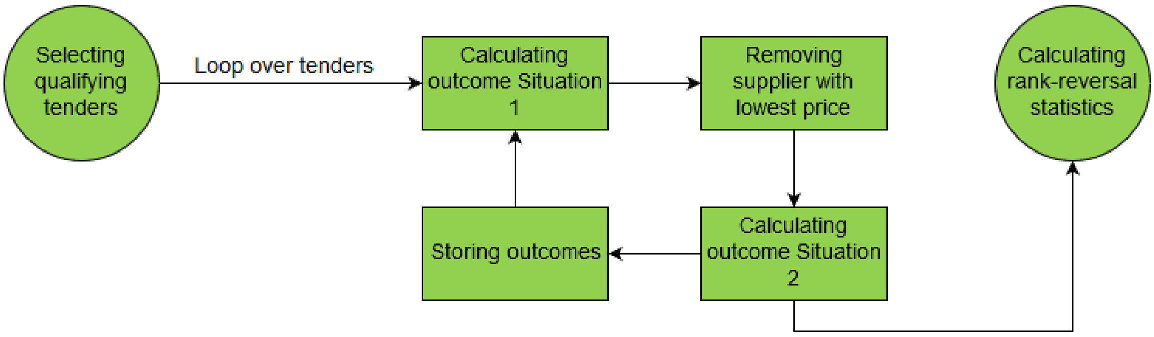

Figure 5: Flow-chart applied on the tender dataset

As mentioned earlier not all tenders were appropriate for this analysis as they used different scoring methods or only two suppliers were interested in the contract. When there are only two bids there is no rank anymore in the second situation as this would consist of just one supplier.

Visual Basic for Applications (VBA) was used to execute all but a few steps required in the analysis. These tasks were as following:

1.) Test whether a tender should be excluded
2.) Calculate price and quality scores for Situation 1
3.) Sort price scores
4.) Calculate total scores Situation 1
5.) Calculate all scores in Situation 2 while ignoring the supplier with the lowest offer
6.) Check whether rank-reversal has occurred

4.2 Results

Table 3 below shows the main outcomes of the what-if analysis. In total 125 tenders were suitable for analysis. This meant over half of the gathered tenders were excluded for using different scoring methods or for having incomplete or invalid data strings.

	Whole dataset	HIS tenders	Non-HIS tenders
No. analysed tenders	125	94	31
No. rank-reversal	3	2	1
Rank-reversal rate	2.40%	2.13%	3.23%

Table 3: General rank-reversal statistics

	0	1	2	3	4	5	6	7	8	9
Whole HIS-dataset	2	8	4	34	22	9	24	123	28	45
No. analysed tenders	0	0	4	4	5	2	5	58	13	14
No. Rank-reversal	0	0	0	0	1	0	0	2	0	0
Rank-reversal rate	N/A	N/A	0.00%	0.00%	20.00%	0.00%	0.00%	3.45%	0.00%	0.00

Table 4: Rank-reversal per CPV-category

As can been seen in Tables 3 and 4 rank-reversal does not prevail in all but three tenders. The corresponding rank-reversal rate is 2.40%. By looking at the ten CPV-categories, rank-reversal did only occur in two categories concerning: *metal products, machinery and equipment* (category 4) and *financial services* (category 7). In six categories the order involving the winner and runner-up did not change after removal of the supplier with the lowest price and two categories had no tenders analysed as they were excluded for reasons mentioned in Chapter 2.

4.2.1 Win on price

	Whole dataset	HIS tenders	Non-HIS tenders
No. analysed tenders	125	94	31
Win on price	74	55	19
Win on price (%)	59.2%	58.5%	61.3%
Win on price + Rank-reversal	3	2	1
Best price + best quality	36	27	9

Table 5: Win on price statistics

In Table 5 the term *win on* price is used multiple times. This term does not imply a supplier wins the tender by virtue of his score on price, but wins the tender and also has the best price offer. This is the case in more than half of the analysed tenders. In most tenders however having the lowest price is not decisive. In half of the tenders in which the winning supplier has the lowest price, it also has the best quality rating. Additionally, since in the analysed tenders weight of the price criterion is mostly thirty or forty percent, as can be seen in Figure 6, a higher price can be compensated with a good quality score.

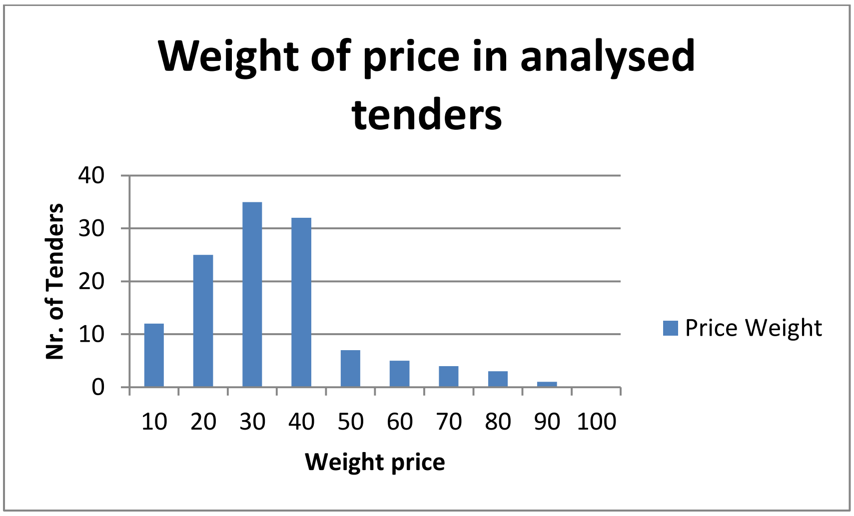

Figure 6: Weight of price in analysed tenders

4.2.2 CPV Category characteristics

It was intended to do an analysis of the characteristics of the CPV Category in which rank-reversal is most occurring. This analysis was omitted due to rank-reversal only occurring in two CPV-categories. Additionally, two specific tenders do not give an insight on the characteristics of a certain category.

Chapter 5: Characteristics of rank-reversal

This chapter focuses on the conducted simulation and the characteristics of rank-reversal which are derived from the results. First of all the outline and distributions of the simulation will be given before going to the results.

5.1 Outline of simulation

The process of the simulation follows more of the same procedure as the what-if analysis. It can still be considered as a static process. Therefore, it was decided to once again use Visual Basics for Applications in Excel for the simulation. VBA was chosen over PlantSimulation as the extra possibilities of the latter would not be used and VBA has a simpler syntax.

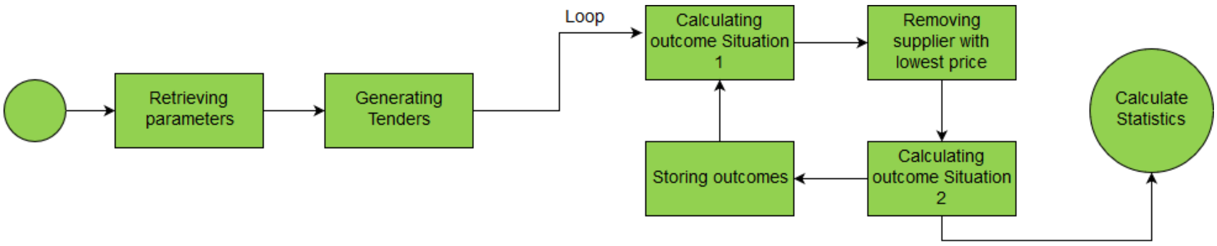

Figure 7: Flowchart of the simulation process

Figure 7 shows the steps taken in the simulation process. The first two steps require user defined input and will be described in the sections below.

5.1.1 Retrieving parameters

Before starting the simulation five parameters can be changed by the user:

1. Number of tenders to be simulated
2. Weight of price criterion
3. Weight of quality criterion
4. Minimum quality threshold applied (1/0)
5. Value of the minimum quality threshold

To get valid results and decrease the chances of strange results, the simulation requires a sufficient amount of tenders to be generated. At first, the number of tenders to be generated was set at 1,000 tenders. This number was however latter increased to 10,000 to counter some outliners.

The four other parameters determine how the simulation will play out. All possible integer combinations between for the weight factor between price and quality were tested to see its influence. The influence of the minimum quality threshold was also tested but only with the value used most. Based on the real world tenders dataset the value for the minimum quality threshold was set at 60 percent. This value was applied in seventy percent of historic tenders in which the minimum quality threshold was applied. The application of other values was lower than ten percent for each occurring value.

If the minimum quality threshold is applied some additional checks and if-statements need to be followed to arrange the appliance. For example, certain calculations have to be rearranged to make sure the simulation does not crash when there are no suppliers in a tender. This is possible as all suppliers can be excluded for not exceeding the minimum quality amount. The tender stats relative to the total of simulated tenders are only calculated towards the number of tenders with at least three tenders, since for rank-reversal to occur at least three suppliers have to put in a bid.

In addition to the five parameters mentioned earlier, the input of simulation was altered manually in the code to allow for a fixed number of participating suppliers and a chanced standard deviation of the bids. These chances could not be made on the dashboard.

5.1.2 Deriving distributions

In order to generate tenders with a certain amount of randomness data distributions had to be derived. To model the input, gathered historic data can be used directly, by using an empirical distribution or applying a statistical probability density function (Robinson, 2014, p.125). The latter option was preferred as it creates the most unique values.

In total three distributions had to be derived: for the number of suppliers participating in a tender, for the scores on quality and for the offered prices. Only for the scores on quality it was possible to use a statistical probability density function as input. Historic data showed a normal distribution was suitable. An empirical distribution was used as input for the number of suppliers participating in a tender. By using a random number each tender was assigned a certain number of suppliers with the selection being based on the histogram of the available dataset.

Deriving a distribution for the offered prices proved to be most difficult of all three. The original plan was to split out all offered prices over multiple orders of magnitudes ranging from below €100 till over a million. This caused problems as some of these classes had not enough offers in it to do a real test of fitness. Additionally, within the classes there was a large variety of offers. In one class this even resulted in a histogram with almost half of the values being concentrated towards the lower bound and the other half towards to upper bound. Because the error for the tested distributions was always higher than the allowed error according to the chi-squared test, all tested statistical probability density functions had to be rejected.

Because no probability density function could be derived from the historic tender dataset an assumption had to be made concerning the prices scores. As the offered prices mostly lie closely to each other, it was assumed that prices were normally distributed with a mean of €10,000 and a standard deviation of €2000. The distributions with its parameters can be found in Table 5.

	Nr. of Tenders	Quality (%)	Price
Type of Distribution	Empirical	Normal	Normal
Mean	N/A	71.31	10000
Standard Deviation	N/A	18.19	2000

Table 5: Parameters of used data distributions

Due to the shape of the normal distribution applied for quality it is possible to receive a score of quality over hundred percent. Since this is not possible, these scores were reset to one. Hence, these suppliers have the best possible quality.

5.2 Results

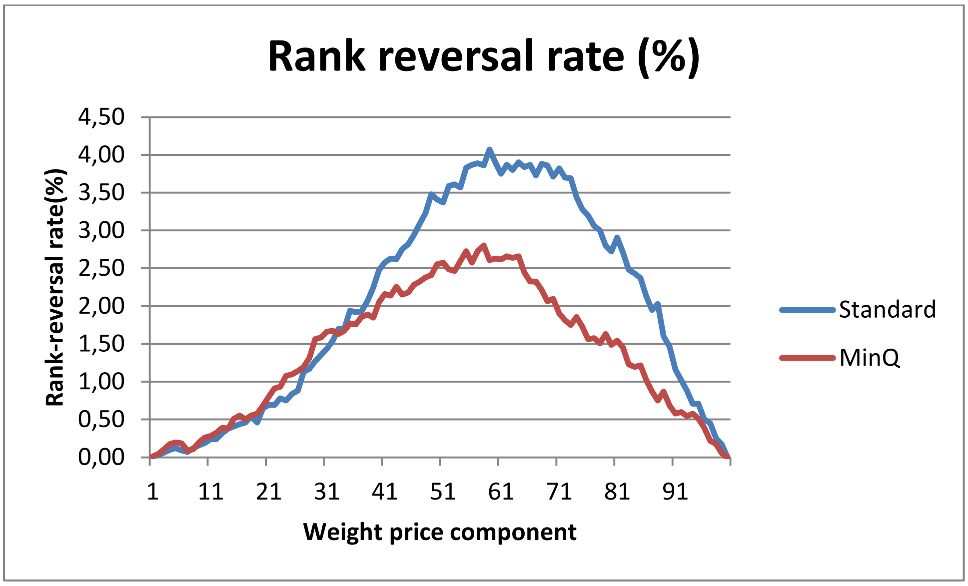

Figure 8: Rank-reversal rate as outcome from simulation

Figure 8 shows the main results of the simulation. In general the occurrence of rank-reversal follows a parabolic pattern as the rank-reversal rate is higher when quality and price have a comparable weight. The pattern is sometimes capricious due to the changing random numbers in each run of the simulation. After the appliance of the minimum quality threshold of sixty percent, the maximum rank-reversal rate drops 1.27 percentage points. Compared to the standard situation the pattern is less parabolic and straighter.

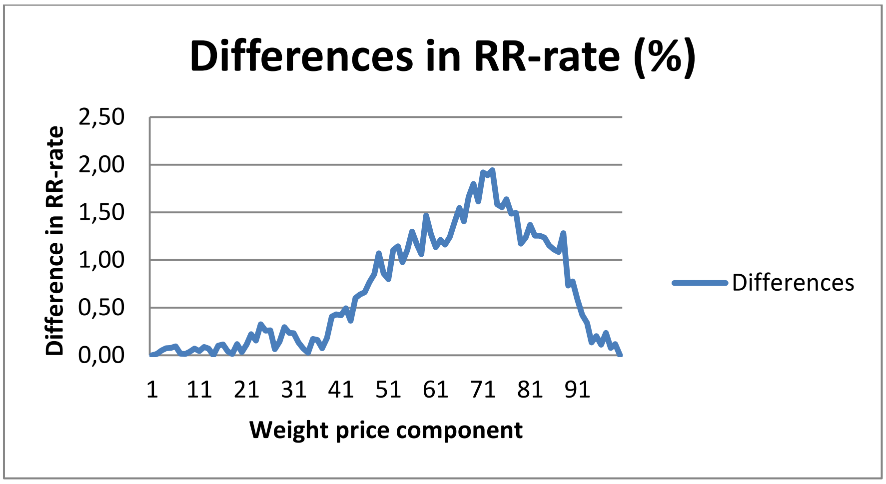

Figure 9: Differences in RR-rate for the analysed weight of price

Figure 9 shows the differences in rank-reversal between the standard situation and the applied minimum quality threshold of 60%. Though the difference in the maximum of the rank-reversal between the situations is 1.27 percentage points, the maximum difference between the two at a certain weight is almost two percentage points.

5.2.1 Win on price

Figure 10: Win on price as outcome from simulation

Figure 10 shows that even when the weight of the price component is close to zero, twenty-five to thirty percent of the winning suppliers also have the best offer on price. When the minimum quality threshold is applied more suppliers win by also having the lowest price. Since in these tenders in general fewer suppliers are evaluated on their prices offers due to the quality constraint, the chance a supplier has both the lowest price and wins the tenders is larger.

5.2.2 Number of bids

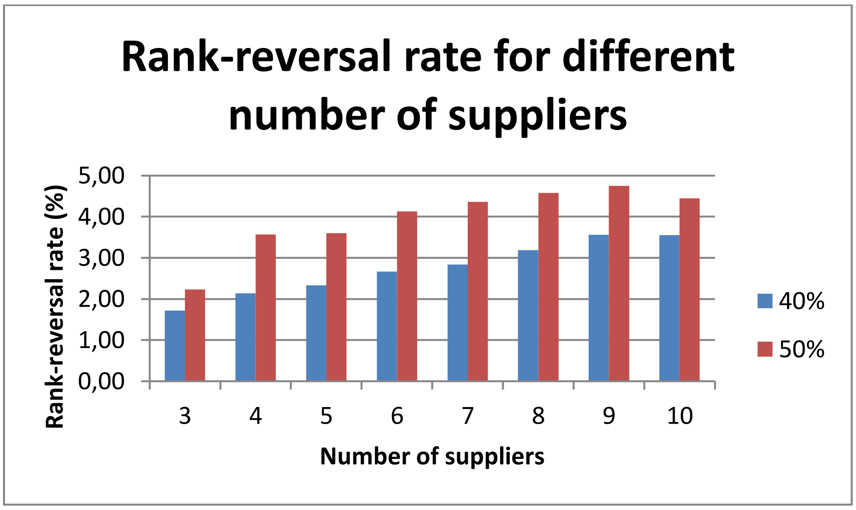

Figure 11: Rank-reversal rate for different numbers of participating suppliers

Figure 11 shows the rank-reversal rate for different number of participating suppliers in a tender. The number of bidding suppliers varies from three to ten as these were the minimum and maximum in the obtained dataset. The analysis was conducted at different weights for price, namely forty and fifty percent.

The rank-reversal rate increases when there are more suppliers participating in the tenders. As more suppliers participate the chances of one supplier submitting a differentiated price offer increases. After removal of this supplier with the lowest price the differences in scores between the other suppliers become larger and rank-reversal is more likely to happen. In addition to this, the pattern of Figure 8 is confirmed as the rank-reversal rate increases when changing the weight of price from forty to fifty percent.

5.2.3 Variability in offers

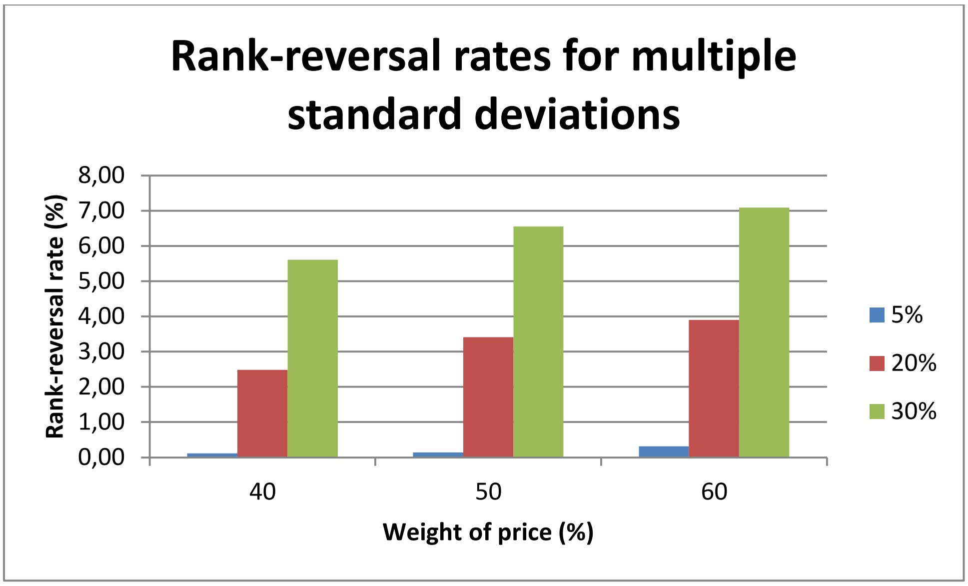

Figure 12: Rank-reversal rates for different standard deviations of the price distributions

Figure 12 shows the rank-reversal rates for different standard deviations and different price weights. A standard deviation of twenty percent, corresponding with the red graph, was used as default value in the previous simulation runs. When the standard deviation is set at five percent instead of the default of twenty the rank-reversal rate decreases drastically to values close to zero percent. The other way around the rank-reversal rate more than doubled after setting the standard deviation to thirty percent. Using different three different weights for price leads to different rank-reversal rates, but the shape of graph stays largely the same.

Chapter 6: Discussion of results

In this chapter the most interesting findings and results will be discussed. This section will refer to both the results of Chapters 4 and 5 and if possible to the analysed literature of Chapter 3.

6.1 Number of sub-criteria

Though at a high level a tender can be split in a price and quality component fulfilling the EMAT definition, these components are usually split up in multiple sub-criteria. For example in the historic HIS-data suppliers where on average evaluated on six different quality criteria. For the price quality mostly a lump sum offer was put in, but price was sometimes also split up in multiple sub-offers each being evaluated in a relative way.

Assessing the influence of multiple sub-criteria for the price component is difficult, since there are multiple factors that have an impact on the scores. First of all, like in all different situations the price scores are interdependent. Secondly, the offers to submit over the multiple sub-criteria can differ a lot. Bidding suppliers for example could be required to submit an hourly wage, a selling price of a product and the total service costs for the whole contract. As these offers can differ a lot, the offers in these three categories will most likely fall in different orders of magnitude.

Finally, the distribution of points over the multiple sub-criteria determines the influences the impact of each sub-criterion. If two of the three criteria only account of 10 percent each of the price component, suppliers could be tempted to 'ignore' these criteria and focus on the third criterion with a weight of 80%.

6.2 Relation between analyses

Though one might be tempted to compare the results of the what-if analysis and the simulation study as they both have the same performance indicators, a comparison is not as straightforward as it might look. First of all, the what-if analysis is a mixture of tenders with different characteristics whereas the simulation looks at each possible weight combination between price and quality.

Secondly, parts of the historic data serve as input for the simulation. For example, the quality scores and number of suppliers participating in a tender were used to derive input distributions. This means both analyses have a relative high correlation. When looking at the average weight of price of the historic dataset, 32 %, this is indeed the case. The what-if analysis gave a rank-reversal rate of 2.4 % compared to 1.5 % for the simulation. The average of the often used price weights of forty, fifty and sixty percent however was higher at 3.26 %.

Percentagewise there still is a large difference between both analyses. This is most likely caused by the earlier mentioned mixture of tenders and the difference in the number of analysed tenders. In addition to this the characteristics of the normal distribution can lead to outliers which normally would not occur in practice.

6.3 Weight factor and rank-reversal

As Figure 8 shows the rank-reversal rate gradually becomes higher when increasing the weight for the price component with peak occurring at a price weight of 59%. After this the rank-reversal starts to drop towards zero. The shape of graph is parabolic, especially in the range between thirty and ninety percent for the price component. For both sides of the spectrum a similar line of reasoning can be found to explain the pattern. Towards the edges of the graph either quality or price determines the outcome of the tender. When the weight of quality is high the influence of price becomes negligible. With a high weight for price it works the other way around; the obtained scores on quality do barely matter.

For example, let's take a situation in which the weight of price is close to 100 percent. In this case almost always the supplier with the lowest price wins the tender. After removal of the supplier with the lowest price it's very likely that the supplier with the second lowest price wins the tender. The latter will obtain the maximum number of points on price and barring large differences on quality and small differences in the offered price the original number two will become the winner. The chances of rank-reversal hence converge towards zero at high weights for either of the two components.

6.3.1 Most used combinations

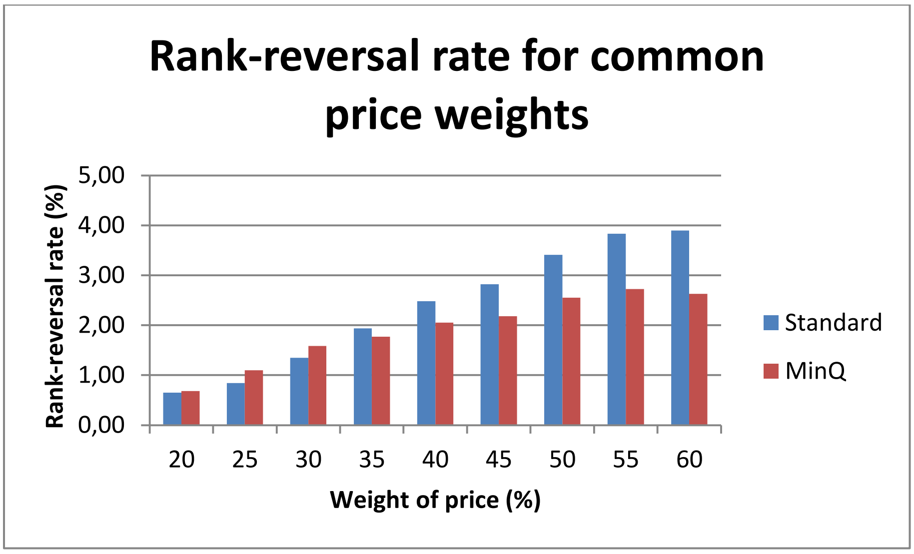

Figure 13: Rank-reversal rates from common price weights

Figure 13 shows the rank-reversal rate for both analysed situation with the price weights varying between twenty and sixty percent in intervals of five. These are the values applied most in the HIS and Negometrix tenders. For the standard situation with no quality threshold the rank-reversal rate lies above 1 percent from 30% and above 2% from 40% percent and onwards. Since Figure 13 includes the weights used most in practice it shows there is a real possibility of occurrence of rank-reversal in practice.

6.4 Minimum quality threshold

During tender procedures applying a minimum quality threshold is done to ensure the winning supplier fulfils all quality requirements. On each sub-criterion suppliers are scored on a scale from 0 to 10 corresponding with values from non-fulfilling to perfect. Since in most cases a score of sixty percent has to be earned for each sub-criterion, the threshold ensures the quality of the winner is judged satisfactory.

Using a quality threshold however can result in de facto changing the way a tender is evaluated. The *Economically Most Advantageous Tender* method can convert towards *lowest acceptable bid*. As each supplier scored on price satisfies the minimum requirements less difference can be made on the quality criteria. The difference between suppliers in case of a threshold of 60% can only be forty percentage points. When all quality scores are accepted, in theory a difference of hundred percentage points can be made.

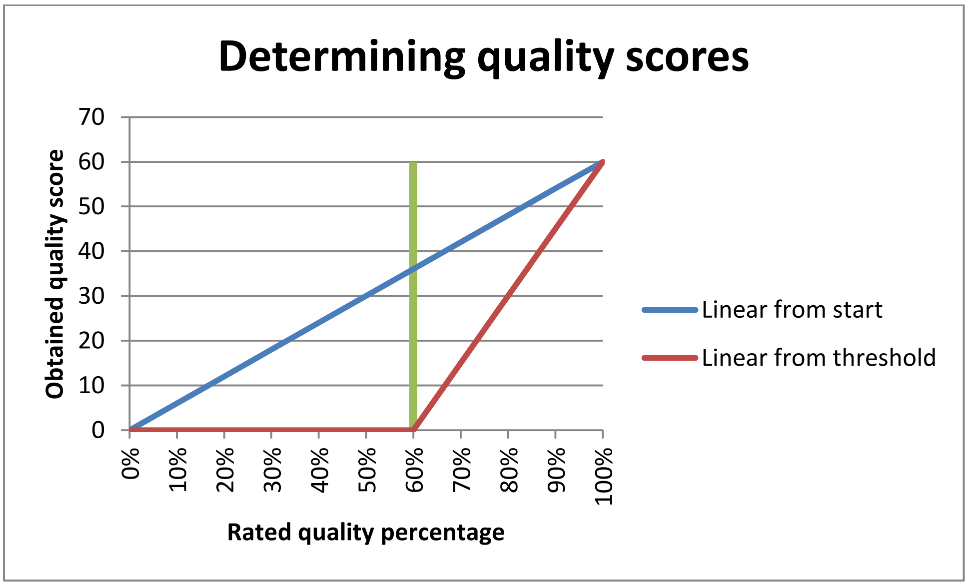

Figure 14: Ways of determining quality scores in case of a minimum threshold

Figure 14 shows different ways how quality can be scored in the case a quality threshold is applied. The blue line shows the method described above; the points are awarded linear to the obtained quality percentage. From the threshold on, which is the green vertical line, suppliers are evaluated on price. In this case the difference between suppliers can only be forty percent of the maximum number of obtainable points as mentioned earlier. To allow for more differentiation when a quality threshold is applied, the red line could be used instead. In this case a supplier which exactly reaches the threshold gets no points and a perfect quality score is awarded the maximum number of points. For all quality percentages in between the scores fall on the linear red line.

Though a quality score of zero is highly unlikely, this aggregated example shows the risk of applying the quality threshold. By less differentiating on quality, the relative weight of prices becomes higher. The intended weights for price and quality therefore can play out differently in practice. This effect can be countered by using different methods not applying a scoring system ranging from one to ten on each criterion.

6.4.1 Impact on results

As Figure 8 and 10 show the appliance of the minimum quality threshold clearly impacts the outcome of the simulation. Though towards axis of the rank-reversal graph no clear cut pattern can be deducted, the rank-reversal rate is significant lower in the middle of the graph. As mentioned this area correspondents with the weight factors used in the HIS and Negometrix tenders.

Since the quality threshold eliminates suppliers from the tender, there are in general less suppliers participating in the final step of the tender procedure. The influences of this can mostly be seen in the win on price graph. At all possible weights the win on price rate is higher than the normal situation. This could be caused by the fact that the evaluated suppliers have a higher chance of winning both the tender and having the lowest price. Though this will always depend on the submitted offers, the reduced number of evaluated suppliers increases the relative chance of winning the tender with also having the best price offer.

6.5 Ranges of submitted offers

As the possibility of rank-reversal is depended on the submitted offers, the price ranges between different suppliers could give an insight on the mathematical possibility of rank-reversal occurrence. Figure 15 and 16 show the price ranges between different suppliers. Figure X has data on the ratio between the highest and lowest offer and Figure Y includes the ratios between the second best and best price and third best and second best.

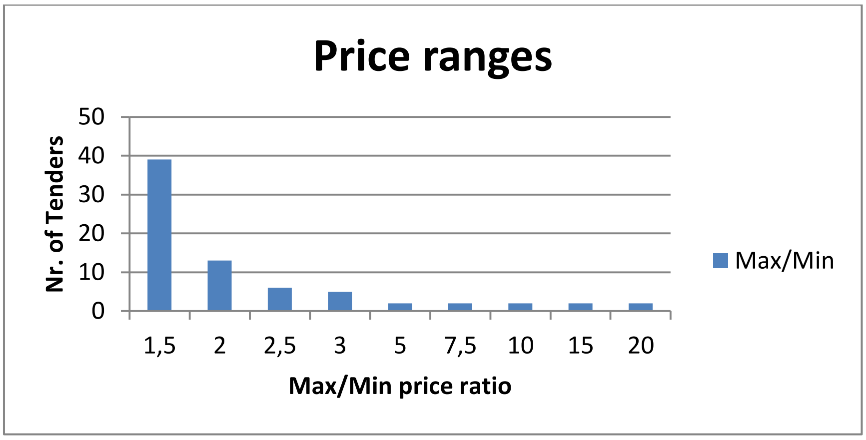

Figure 15: Price ratio between highest and lowest price

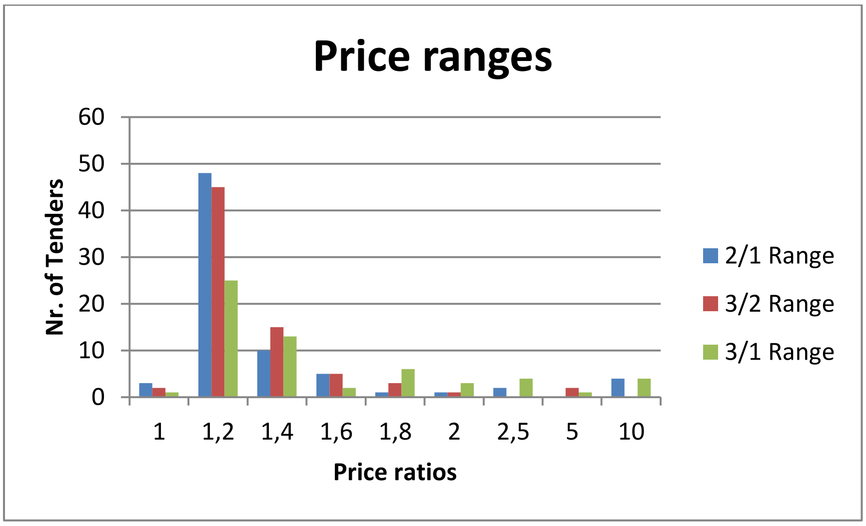

Figure 16: Price ratios between three best offers

Figure 15 shows that in the most tenders the highest offer is at most 50% higher than the lowest offer. However in ten of the analysed tenders the highest offer is at least five times as high as the best offer. Looking at ratios in Figure 16 the differences between offers are much smaller when taking the best three offers into account. For example, in the sixty percent of the analysed tenders the difference between the best and second best offer is at most twenty percent. This is also the case when looking at the ratio between the third best and best offer.

Most interesting when looking at the possibility between rank-reversal are large differences between the lowest and other prices. When a supplier for whatever reason submits a much lower offer compared to the other suppliers, the scores on price for the other suppliers will be close to each other. If the supplier with this much lower bid is removed, the scores of the remaining suppliers will differ more as the base price lies closer to the other bids. This increases the chance of rank-reversal, though this was not the case in the what-if analysis.

6.5.1 Price ranges in simulation

Due to the characteristics of the normal distribution applied in the simulation, it will show a different pattern compared to that in the historic tenders of Figure 16. Under the three sigma rule 68% of the values will fall within one standard deviation of the mean, 95% percent for two standard deviations to the mean and 99.7% within three standard deviations. Translating this to maximum divided by minimum ranges the corresponding values are 1.5, 2.33 and 4.0.

The shape of distribution almost completely eliminates the chance of the difference between minimum and maximum price being higher than four. In practice such values will not occur often as both an extreme price has to be generated for the best and worst offer.

6.5.2 Change in standard deviation

As can be seen in Figure 13 in section 5.2.3 an increase in the standard deviation and therefore more variability in possible bids leads to an increase of the rank-reversal rate. For example at a price weight of forty percent the rank-reversal rate becomes 5.61% compared to the 2.48% for a standard deviation of twenty percent. The hypothesis that rank-reversal occurs more often when the variance in bids is higher is therefore confirmed by the simulation results.

6.6 Win on price

Both the what-if analysis on the historic dataset and the simulation showed a large number of suppliers who win the tenders also have the best price offer. This does not imply the suppliers win by virtue of the lowest price, since quality is judged as well but it is certainly a boost in the right direction.

In the simulation the win on price rate continuously increased when the weight of price became more important. In these cases the tender procedure converges to a lowest bid tender as less and less difference can be made on price. Hence, for price weights over 90% percent more than 95% of the suppliers win the tender by also having the lowest price.

Though the win-on-price rate converges to 100% when high importance is given to the submitted offers, the rates are also fairly high for more standard price-quality weights. For the most used price weight of forty percent, half of the winning suppliers have the lowest price. This is comparable to the outcome of the what-if analysis. This can be caused by the shape of the scoring graph as demonstrated in Figure 1. A fairly small deviation to the base price can lead to decisive gaps between suppliers.

6.7 Link with literature review

The mentioned advantage that relative scoring is easy when it comes calculating scores was seen in both analyses. To calculate the scores only the lowest offer had to be stored before a loop over all participating suppliers in a tender could take place. The standard templates made data gathering more convenient, but the main advantage of these templates lies at the authority or organisation initiating a tender.

The found disadvantages of relative scoring in Chapter 3 were not found in the analyses. Unrealistic or impossible offers were found in the simulation when the standard deviation was set to fifty percent, resulting in a higher rank-reversal rate. However since these offers were caused by the nature of the applied normal distribution the outcome of this specific simulation run was omitted. The other disadvantages were more general and cannot be traced in the data or simulation.

6.8 Further steps

Both the what-if analysis and the simulation showed rank-reversal occurs in practice. In the introduction it was stated that rank-reversal was considered undesirable. It is up to contracting authorities to judge whether they share this statement. Literature and real world conversations about the subject gave different opinions on acceptable rank-reversal rates. Some see every tender in which rank-reversal occurs as one too many, where others are willing to accept a small rank-reversal rate. Should relative scoring be judged as an undesirable method, a cooperative approach could be used to choose a different method. By working together to evaluate tender procedures and rules, both governmental as corporate parties can stress their problems and propose improvements to make sure the most deserving offer wins the contract.

6.8.1 Further research

There are a few steps can be taken to gain a better view on which factors influence the rank-reversal rate .First of all, more research can done on finding appropriate distributions for offers. The applied normal distribution may not yield enough variance in certain sectors which allow for multiple price-to-quality positions or in highly competitive markets. The other way around, in some tenders the variance in bids may have been too high. Not always can be differentiated on price due to the nature of the required product of service and the kind of offer which needs to be submitted to the tendering authority.

Secondly, the influence of entrance has not been analysed in this thesis. In the simulation results would have been similar as offers and quality scores would have drawn from the same distribution as all other bids. However, in the second running of a real world tender suppliers have gained additional knowledge by participating in the first running. Different and tailored offers could therefore be submitted in the second procedure, possibly leading to a new outcome.

Thirdly, research can be conducted to see whether other relative scoring methods have comparable results and patterns as in this report only one relative scoring method was analysed.

Chapter 7: Conclusion

In a large number of Dutch public tenders relative scoring methods are used. Due to the nature of relative scoring methods, removal or entrance of a supplier can change the outcome of a tender. Though past research has focussed on the circumstances in which rank-reversal can and will occur, not much was known about the real-world occurrence of rank-reversal. Therefore, this report tried to answer the following main research question:

How often does rank-reversal occur in practice and what are the characteristics and situations under which it occurs?

Motives for the usage of relative scoring are the easiness of calculating scores mainly due to the fact no predefined scoring table is required as scores are interdepend. Found problems in literature are the (mathematical) impact of unrealistic offers, the absence of a price indication for suppliers and the fact suppliers together can accommodate their bids to maximise their chances of winning.

A what-if analysis on a historic dataset of 252 governmental and 51 other tenders showed rank-reversal would have occurred after removal of the supplier with the lowest price in 2.40% of the cases. The rank-reversal rate for governmental tenders alone was slightly lower at 2.13% of the analysed tenders.

In the simulation rank-reversal rates varied from null to 4.07% depending on the weights of price and quality. When the weights for either price or quality are close to zero the rank-reversal rate converges to null. In the price range in which most tenders fall, with a weight between 40 to 60 percent, the rank-reversal rate is always higher than 2.5%.

Applying a minimum quality threshold of sixty percent gives comparable results to the standard situation towards lower weighs for either price or quality, but reduces the rank-reversal rate when weight and price are from comparable importance. The highest rank-reversal rate still lies at 2.80%.

Rank-reversal rates increase when more suppliers participate in a tender, but impact limits itself when there already more than six bidders. In addition, rank reversal occurs more frequently when the variance in submitted bids is increased.

Concluding it can be said that in practice rank-reversal occurs in over 2% of analysed real-world tenders. The simulation showed a parabolic pattern with rank-reversal rates converging towards zero when weights for either price or quality are low. The rank-reversal rate peaked at a weight for price at 59%. The appliance of a minimum quality threshold of 60 percent led to a reduced rank-reversal rate for commonly used weights. The peak however still lied at 2.80%.

To get an even better view on relative scoring and rank-reversal occurrence more research can be done on the finding better price distributions, the impact of entrance of suppliers and whether other relative scoring methods give similar results and patterns.

It is now up to contracting authorities to determine whether they want to change the rules and methods currently in place to evaluate suppliers in a tender. As there are multiple strong opinions on relative scoring, this can become a challenging process.

8. Citation

Aanbestedingswet 2012. (2016, July 1). Netherlands.

Cheng, T. (2009, April). Relatieve scores bevoordelen de zittende leverancier. *Tender Nieuwsbrief: actuele feiten en achtergronden over Europese aanbestedingen*(3), 4-5.

CPV-codes. (n.d.). Retrieved June 22, 2017, from PIANOo: https://www.pianoo.nl/regelgeving/cpv-codes

Dimitri, N. (2013). Best Value for Money in Procurement.

ECLI:NL:GHARL:2013:BZ8213, 200.120.024 (Gerechtshof Arnhem-Leeuwarden April 2013).

Merckel, W. B. (2015). *Een andere winnaar bij een EMVI aanbesteding?*

Ongeldige en bijzondere inschrijvingen. (n.d.). Retrieved June 27, 2017, from PIANOo: https://www.pianoo.nl/inkoopproces/fase-2-doorlopen-aanbestedingsprocedure/gunnen/ongeldige-bijzondere-inschrijvingen

Rijksoverheid. (2015). *Gedragscode Integriteit Rijk.*

Robinson, S. (2014). *Simulaton: The Practice of Model Development and Use* (2nd ed.). Palgrave Macmillan.

Stilger, S., Siderius, J., & Raaij, v. E. (2015). *A Comparative Study of Formulas for Choosing the Economically Most Advantageous Tender.*

Telgen, J., & Timmer, J. (2016). *Conditions for Rank Reversal in Supplier Selection.* Enschede.

The European Parliament and the Council of the European Union. (2014, March 28). Directive 2014/24/EU of the European Parliament and of the Council.

UBR|HIS. (2017, June 6). *UBR|HIS Overzicht*. Retrieved from Uitvoeringsorganisatie Bedrijfsvoering Rijk: https://www.ubrijk.nl/organisatie/ubr-haagse-inkoop-samenwerking