ARTICLE IN PRESS

Journal of Purchasing & Supply Management xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Purchasing and Supply Management

journal homepage: www.elsevier.com/locate/pursup

Supplier selection with rank reversal in public tenders

Fredo Schotanus [a,*], Gijsbert van den Engh [b], Yoran Nijenhuis [b], Jan Telgen [b]

[a] Department of Economics, Utrecht University, Utrecht University Centre for Public Procurement, the Netherlands

[b] Department of Technology Management & Supply, University of Twente, University of Twente Initiative for Purchasing Studies, the Netherlands

ARTICLE INFO

Keywords:
Supplier selection
Rank reversal
Relative scoring methods
Interdependent scoring methods
WSM
Multi-criteria decision making

ABSTRACT

For supplier selection in the public sector, the Weighted Sum Model is often used in combination with relative scoring methods that allow rank reversal. With rank reversal we refer to a changed order in the ranking of bids leading to a new winner, after removing or adding a non-optimal bid that does not win the original tender. In practice, an important reason indicated by practitioners for using methods that allow rank reversal is that it would rarely occur in practice. Based on an analysis of 303 Dutch public tenders, this research shows this is not true. In about 1 out of 5 the tenders, rank reversal occurs after adding non-optimal fictional bids to tenders that do not have quality thresholds. After removing bids, the rate is about 1 out of 40 if a curved relative scoring method is used. In addition, the research shows that rank reversal rates increase when (i) there is no quality threshold, (ii) the number of bids increases, (iii) bid price variance increases, and (iv) price weights are not very low or high. We argue that relative scoring methods that allow rank reversal should not be used in public procurement, or otherwise only in exceptional cases, as it conflicts with public procurement principles and leads to reduced overall bid value.

1. Introduction

The supplier selection decision is considered to be one of the most significant decisions in procurement (e.g. Luzzini et al., 2014; Mummalaneni et al., 1996; Wu and Barnes, 2011). Supplier selection is thus a much debated and studied topic in both academic literature and practice. In practice, both private and public buyers use many different models for selecting suppliers. In academic literature, normative decision theory prescribes, among other things, which method would be optimal to use in different circumstances (e.g., De Boer et al., 1998; Choi and Hartley, 1996; Munson and Rosenblatt, 1997; Weber and Current, 1993).

Despite this attention, the use of formal supplier selection methods is not without problems. One reason for this is that based on decision theory, many decision methods can be considered, though the effects of these methods in the real world are not always known. As a result, many organizations, especially those in the public sector, struggle with the pressure to make and explain sound supplier selection choices (De Boer et al., 2006). The struggle is larger in the public sector, as supplier selection models and decisions are regulated to some extent. The EU public procurement directives state for instance that to enhance transparency, equal treatment, objectivity, and non-discrimination, public buyers should publish calculation methods, award criteria and their relative importance in a request for a proposal.

A specific supplier selection issue for both public and private organizations that use formal selection models is the issue of rank reversal. With rank reversal we refer to a change in the ranking of bids from suppliers leading to a new winner, after adding or removing a non-optimal bid (based on De Farias Aires and Ferreira, 2018). In supplier selection, rank reversal can occur when buyers use multi-criteria selection methods such as Analytic Hierarchy Process (AHP), TOPSIS, or Weighted Sum Model (WSM) in combination with certain relative scoring methods. In this article, we focus on the occurrence of rank reversal in the latter combination as this is a well-known and widely used method all over the world (e.g. Aissaoui et al., 2007). In WSM, all bids of suppliers are awarded scores on all award criteria. These scores are multiplied with the respective weights of the award criteria. The bid with the highest total score wins the contract or proceeds to the next selection phase in case there are multiple selection rounds. The scores on individual criteria such as price can be calculated by the buyer through various scoring methods. These methods can be classified under absolute (i.e. independent) and relative (i.e. interdependent) scoring methods. The calculations under absolute methods are independent of the other bids and can therefore not lead to rank reversal. The

* Corresponding author.
E-mail address: f.schotanus@uu.nl (F. Schotanus).

https://doi.org/10.1016/j.pursup.2021.100744
Received 26 September 2021; Received in revised form 6 December 2021; Accepted 19 December 2021

calculations under relative scoring methods however depend on the best, worst or average bids. This interdependency can lead to rank reversal. For instance, if a non-optimal bid (i.e. an irrelevant bid that does not have the highest total score) with the best price would be removed or added to the pool of bids, this could lead to a new ranking, including a new winner.

Using a supplier selection method that allows rank reversal is problematic for three reasons. First, it conflicts with public procurement principles as transparency and equal treatment (Manunza, 2018). Principles that are also important for many private organizations. Second, it could lead to flawed decision making (e.g. Saaty and Vargas, 1984; Chen, 2008; De Boer et al., 2006; Smith, 2010; Stilger et al., 2017; Sykes, 2012; Mufazzal and Muzakkir, 2018; Wang and Luo, 2009), as a bid could be selected that is not the best match with the value functions of the buyer. Finally, it could lead to lower price-quality ratios compared to methods that do not allow rank reversal (Albano et al., 2008; Telgen and Schotanus, 2010). This lower ratio can be explained by buyers not indicating price preferences or value functions, suppliers making an educated or strategic guess what the lowest price will be, and possibly even bid rigging (Telgen and Schotanus, 2010).

Because of these problems, relative scoring methods are prohibited by law for public procurement in Portugal. In almost all countries however, relative scoring methods are allowed and are often used in both private and public procurement practice (e.g. Chen, 2008). In some countries such as South Africa, it is even mandatory for public buyers to use a relative scoring method for price when contracts are awarded based on multiple criteria. As a result, WSM in combination with a relative scoring method is used in thousands of supplier selection models.

Earlier research on rank reversal mostly focuses on theoretical aspects, such as the nature of rank reversal, and describes causes and conditions. To a smaller extent, solutions and simulations have been published (De Farias Aires and Ferreira, 2018). However, earlier research does not tackle common counterarguments used by buyers who use methods that allow rank reversal. In our experience, such buyers often argue that rank reversal is only a theoretical problem and is not likely to occur in the specific circumstances of the tender at hand (e.g. a low or high expected number of bidders). Such buyers keep using methods that allow rank reversal, despite that solutions exist that prevent rank reversal that have been described for both the private (e.g. De Farias Aires and Ferreira, 2019; Yu and Hou, 2016; Žižović et al., 2020) and public sector (e.g. Bergman and Lundberg, 2013; Kumar et al., 2019; Stilger et al., 2017).

This article therefore aims – by conducting a large-scale empirical analysis and simulations based on public tender data from the real world – to fill the research gap mentioned above by answering the following research question: To what extent does rank reversal occur in different supplier selection circumstances while using WSM in combination with relative scoring methods?

Answering this research question broadens decision theory understanding about when and to what extent rank reversal occurs in supplier selection decisions. Among other things, we show that rank reversal occurs often after adding non-optimal fictional bids to tenders. We also show under what circumstances rank reversal rates increase. With these results we aim to close the gap between the mostly theoretical world of decision theory and the real world in which actual decisions and supporting policies are made (similar to Bruno et al., 2012). The data are collected from Dutch public organizations, but the findings are relevant to all organizations worldwide, both public and private, that use WSM in combination with a relative scoring method. Based on our contribution, we provide insights for procurement policy makers and practitioners seeking to develop policies about applying scoring methods. We argue in the discussion section that relative scoring methods that allow rank reversal should not be used in public procurement, or otherwise only in exceptional cases.

2. Literature review

In this section, we first position our research in the general literature about rank reversal and decision theory. In the second part of the section, we focus on rank reversal in the specific context of procurement.

2.1. *Rank reversal in decision-making literature*

Decision theory concerns rational decision making. It can be divided in a normative and a descriptive part. Normative decision theory analyzes how people or organizations should make decisions. Descriptive decision theory analyzes and predicts how people or organizations actually make decisions (Peterson, 2017). In normative decision theory related literature, the rank reversal problem is an intensively discussed topic. The potential of rank reversal is considered as a major criticism for decision-making methods such as AHP (Dyer, 1990; Leung and Cao, 2001) as it can lead to arbitrary decisions (Belton and Gear, 1983, 1985) and conflicts with the feature of rank perseveration (Saaty and Sagir, 2009). It is therefore studied in many articles. In decision-making literature, rank reversal related studies can be found that study:

1. Different methods that have the possibility of rank reversal (e.g. Wang and Luo, 2009),
2. Different types of rank reversal (e.g. Belton and Gear, 1983; Triantaphyllou and Mann, 1989),
3. General causes and properties of rank reversal (e.g. Belton and Gear, 1983; Zanakis et al., 1998),
4. Avoidance and (undesired) effects of rank reversal (e.g. De Farias Aires and Ferreira, 2019; Millet and Saaty, 2000; Saaty, 1994; Schenkerman, 1994; Wang and Elhag, 2006).

In the rest of Section 2.1, we subsequently discuss these topics where they are relevant to our study.

2.1.1. *Different methods that have the possibility of rank reversal*

De Farias Aires and Ferreira (2018) wrote an extensive literature review about rank reversal. They find that many studies have been conducted about methods and many variants such as AHP (e.g. Belton and Gear, 1983), TOPSIS (e.g. Ren et al., 2007), Electre (e.g. Wang and Triantaphyllou), PROMETHEE (Zhang et al., 2009) and their sensitivity to rank reversal. However, no studies were found that study large sets of supplier selection decisions – using WSM in combination with relative scoring methods – from practice, in order to better understand the occurrence of rank reversal. Rank reversal is mostly studied in general and not specifically for supplier selection methods.

2.1.2. *Different types of rank reversal*

De Farias Aires and Ferreira (2018) also propose a classification of different rank reversal types. Our article focuses on the most common type of rank reversal in which the ranking is changed after adding or removing a bid. Examples of other types of rank reversal are when a non-optimal bid would be replaced by a worse bid (based on Triantaphyllou and Mann, 1989) or when a non-discriminating criterion (i.e. a wash criterion; De Farias Aires and Ferreira, 2018) would be removed from the supplier selection model (based on Finan and Hurley, 2002).

2.1.3. *General causes and properties of rank reversal*

Regarding the causes and general properties of rank reversal, decision-making literature provides some general insights. The main cause for rank reversal for methods such as AHP (Belton and Gear, 1983) and WSM is the usage of relative scoring methods. Another general insight about rank reversal is that rank reversal occurs more often in decision problems with more alternative solutions (i.e. more bids in the procurement context) (Zanakis et al., 1998). Zanakis et al. (1998) also show that the number of alternatives has more influence on rank reversal than the number of criteria. Although these are useful insights, they do not indicate specifically how often rank reversal can occur in supplier selection problems. In addition, Zanakis et al. (1998) do not study WSM in combination with relative scoring methods and they analyze randomly generated scores using a uniform distribution instead of actual scores from practice.

2.1.4. Avoidance and (undesired) effects of rank reversal

Finally, there are many studies comparing or proposing supplier selection methods that prevent rank reversal from happening in order to prevent undesired effects (e.g. Kumar et al., 2019; Yu and Hou, 2016). Such alternatives are typically more accurate in decision making, but are also more complex (Masi et al., 2013; Smytka and Clemens, 1993). A disadvantage of more complex methods is that many procurement practitioners do not start using them as long as they believe that their current methods lead to optimal decisions and are not prone to rank reversal. Although most academic studies state that rank reversal is problematic in multi-criteria decision making, rank reversal is not considered to be very problematic if rank reversal occurs only occasionally (Saaty, 1994; Millet and Saaty, 2000). So, if WSM in combination with relative scoring methods would rarely create rank reversal, it is understandable that practitioners prefer simple selection methods. That is why we believe that research is required to the actual occurrence of rank reversal in simple and commonly used supplier selection methods.

2.2. Rank reversal and relative scoring methods in the context of supplier selection

As mentioned before, rank reversal in the context of supplier selection is defined as a changed order in the ranking of bids resulting in a different winner after removal of a non-optimal bid (i.e. removing a bid which did not win the original tender) or entrance of a new non-optimal bid (i.e. adding a new bid to the original tender which does not win the tender). The latter means that the decision of suppliers with non-optimal bids whether or not to participate in the tender can influence who wins the tender. In this case, there is not an economic reason for winning the tender, but the participation of another non-optimal bid determines who wins. Note that this does not mean that adding or removing bids occurs often in practice. That is not the issue. The issue is that there may be an impact from a specific (non-optimal) bid being submitted, or not.

Earlier work on rank reversal in the context of supplier selection focuses on AHP and TOPSIS variants (Ic, 2014; Lima Junior et al., 2014; Rodriguez et al., 2013), and to what extent rank reversal can be problematic using these methods. No quantitative data are available about the actual use of these methods in practice, but as far as we know, these methods are not used as commonly for supplier selection as WSM. Especially TOPSIS seems to be a method that is mostly discussed and researched in academic literature, but practical applications for supplier selection seem to be relatively uncommon. This is not surprising when considering the context of supplier selection. Supplier selection models need to be understood not only by the procurement officer, but also by managers (Brun and Pero, 2011), internal clients, and suppliers. In addition, procurement officers are typically not trained as decision-making experts. This could explain the popularity of WSM in supplier selection practice. Zanakis et al. (1995) also note that WSM's simplicity makes it popular to practitioners.

As mentioned in Section 2.1, rank reversal has been studied in several theoretical articles. However, as far as we know, no empirical studies nor simulations based on data from practice indicate how often rank reversal occurs in supplier selection under different circumstances. Earlier supplier selection work also did not study rank reversal effects as a result of adding bids. As far as we know, only Stilger et al. (2017) show that depending on the type of relative scoring method used, rank reversal occurs in 1.27 up to about 6.35 percent (excluding uncommon methods) of all cases studied. However, Stilger et al. only show averages and not specific circumstances (such as how the number of bids affects rank reversal rates) and they only study what happens when bids are removed and not when bids are added. Finally, they apply different scoring methods on the same dataset, while the scoring methods could have influenced bid strategies of suppliers participating in tenders. Strategic bid behavior in the context of selection methods that allow rank reversal is discussed in more detail in the discussion section.

In closing, in countries such as Sweden (Bergman and Lundberg, 2013), South Africa (based on Raga and Taylor, 2010) and the Netherlands (Chen, 2008) relative scoring methods that allow rank reversal are popular scoring rules. When such methods are used, buyers often indicate that rank reversal is not likely to occur. Or that it might occur in other tenders, but not in the particular circumstances for the tender at hand. Earlier court rulings about rank reversal cases (e.g. Court of Arnhem, 2012) do not seem to have an effect on such perceptions. As far as we know, current literature does not provide a sound answer to this criticism. In the rest of this article, we therefore show how often rank reversal occurs or can occur under different circumstances based on data from the real world of public tenders. This broadens decision theory understanding about when and to what extent rank reversal occurs in actual supplier selection decisions. In addition, we add to the literature an analysis of adding non-optimal bids to tenders and under what specific circumstances the possibility of rank reversal is higher or lower for tenders in which WSM and relative scoring methods are used.

3. Method

In this section, the data collection procedure and properties of the dataset are presented first. Next, two types of data analyses are explained. The first type focuses on adding bids to a tender and the effects on rank reversal rates. The second type focuses on removing bids.

3.1. Data collection and description

To be able to answer the research question, tender documents and purchasing data about price and quality scores of all bidders participating in multi-criteria tenders were required. Although some countries have open data available about prices in lowest price tenders, as far as we know, data about price and quality scores of all bidders are not publicly available. For our research, we aimed to find a broad selection of tenders with all kinds of different number of bidders and weights for price and quality. In the Netherlands, many of such tenders are potentially available as it is common practice to use supplier selection methods that take into account both price and quality. We therefore contacted several Dutch public organizations and service providers. First, 51 tenders where available from five organizations including data from two consultancy firms which advised public organizations, a tender platform, a university and another public organization. Next, 252 tenders were received from the public organization HIS. This is a large public buying office and mainly responsible for the procurement of six Dutch ministries for many different procurements. After receiving this data, the dataset contained a sufficient number of tenders (303) to conduct our analyses.

The tender dataset includes information about the number of bids, the scoring method used, the weight for both quality and price, the number of sub-criteria, the obtained quality scores, the bid prices and some information about the subject-matter of the contract. Most of the tenders involved services and/or supplies. Each tender has a contract value of at least € 134,000.

For 233 out of 303 tenders, the Common Procurement Vocabulary (CPV) code or codes were available, describing the product group(s) of the tender at hand. Table 1 describes the different product groups that have been tendered in more detail. The table shows there is no even distribution and there are also differences between the groups in numbers of bidders and weights of price and quality. For instance, financial and business service tenders typically have lower price weights than tenders for electricity or transportable goods.

Table 1
Data description.

No.	Product group	Tenders	Average no. of bidders	Average weight of price	Tenders with relative scoring methods	Tenders with quality threshold
0	Agriculture, forestry and fishery products	1%	4.5	30.0	100%	100%
1	Ores and minerals; electricity, gas and water	3%	2.4	52.3	100%	88%
2	Food products, beverages and tobacco; clothing	1%	6.0	20.8	100%	100%
3	Other transportable goods, except metal products, machinery and equipment	11%	3.4	54.8	83%	41%
4	Metal products, machinery and equipment	7%	4.9	48.1	100%	36%
5	Constructions and construction services	3%	3.6	47.9	100%	56%
6	Distributive trade services; accommodation, food and beverage services; transport services; and utilities distribution services	8%	3.2	58.8	74%	33%
7	Financial and related services; real estate services; and rental and leasing services	41%	4.2	28.3	94%	75%
8	Business and production services	9%	3.4	27.9	86%	68%
9	Community, social and personal services	15%	4.1	27.6	95%	67%

In this research, there is not a focus on differences on product group level. The study aims to create insights that are generally relevant for all product groups. Among other things, research topics are the effect of different price weights on rank reversal rates and the effect of different (expected) numbers of bidders. When studying rank reversal, such characteristics can be studied independent of – among other things – the product group category and the country from which the data are collected. This means that for the interpretation of the findings, we argue that it does not matter what the product group is or in which country the data was collected. In addition, as is discussed later in this article, the average rank reversal rate found in the study is similar to Stilger et al. (2017). This also indicates that the dataset is not biased for the research purpose.

In most tenders the most expensive bid is at most 50% more expensive than the lowest bid. The differences between bids are smaller when taking the lowest three prices into account. For example, in 60% of the analyzed tenders the difference between the best and second best price is at most 20%. In 59% of the tenders, the winning bid has the lowest price. In about half of the tenders, in which the winning bid has the lowest price, it also has the best quality rating.

In various tenders, a quality threshold is applied by the buyer. In these cases a two-step approach is used for evaluating each bid. First, bids are scored on quality components. If this score is under a predetermined threshold, the bid is disqualified. Second, if a bid satisfies the quality minimum, it will be evaluated on price and quality. The quality threshold is on average about 60% with small differences between the different product groups.

3.2. Data analyses

We conducted two main types of data analyses regarding rank reversal. The first analysis focuses on adding non-optimal bids and its effects on rank reversal rates. The second analysis focuses on removing non-optimal bids and studies in more detail the effects of different characteristics (e.g. weights, number of bids) on rank reversal rates.

3.2.1. Analysis 1: rank reversal rates as a result of adding non-optimal bids

The first analysis focuses on adding non-optimal bids to the dataset and analyzes the effects on rank reversal. To this end, we calculated what price, offered by a new fictional supplier with a non-optimal bid, was necessary to result in rank reversal.

For the analyses, only tenders with a relative scoring method were used. Among other things, this means we excluded the bids submitted to tenders with absolute scoring methods, as the bids for such tenders might have been influenced by the scoring method. A second condition was that we did not analyze tenders in which two or more price criteria were evaluated with multiple relative scoring methods, as the dataset did not include a sufficient number of such tenders for our analyses. Because of these restrictions, 92 tenders were excluded. We analyzed 211 (out of 303) tenders in the first data analysis.

In the dataset, two types of relative scoring methods were used:

- A linear method defined as $2 \times \textit{max points} - \frac{\textit{Price supplier i}}{\textit{Lowest price}} \times \textit{max points}$ (see also the left figure below). In total there are 49 tenders in the dataset that use this method;
- A curved method defined as $\frac{\textit{Lowest price}}{\textit{Price supplier i}} \times \textit{max points}$ (see also the right figure below). In total there are 162 tenders in the dataset using this method.

The effects on the number of points that can be scored with different prices for a price award criterion is illustrated in Fig. 1 for both scoring methods.

3.2.2. Analysis 2: rank reversal rates as a result of removing non-optimal bids

In the second analysis, it is analyzed what happens if the bid with the lowest price from each tender of the original dataset is removed. In addition, it is analyzed under what conditions rank reversal occurs, by running simulations with tenders based on the properties of the original dataset.

For the simulations, the historic data on quality and the number of bids was used to derive data 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 bids 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 with a mean of 7.131 points and a standard deviation of 1.819 points on a 10-point scale. An empirical distribution was used as input for the number of bids per tender. By using a random number, each tender was assigned a certain number of bids with the selection being based on the histogram of the available dataset.

For the offered prices, all tested statistical probability density functions had to be rejected, because the error for the tested distributions was higher than the allowed error according to chi-squared tests. Because no probability density function could be derived from the dataset an assumption had to be made concerning the price scores. As the offered prices mostly lie close to each other, it was assumed that prices were normally distributed with a mean of € 10,000 and a default standard deviation of 20%. Other standard deviations (5 and 30%) were simulated as well to test the effects of the standard deviation on rank reversal rates.

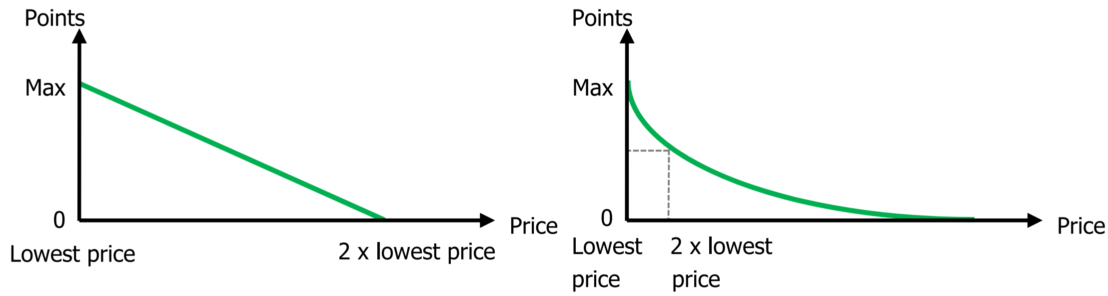

Fig. 1. Effects of linear and curved relative scoring methods on number of points to be scored depending on bid price.

4. Results

In this section, the results of data analyses 1 and 2 are described subsequently. Data analysis 1 focuses on adding bids to a tender and the effects on rank reversal rates. Data analysis 2 focuses on removing bids.

4.1. Analysis 1: rank reversal rates as a result of adding non-optimal bids

For data analysis 1, the occurrence of rank reversal after adding non-optimal bids to the dataset is analyzed. For each tender, using the propositions posed by Telgen and Timmer (2016), it is calculated what price of an added non-optimal bid would be required to create rank reversal. The prices of the added non-optimal bids are on average 42% lower than the original lowest price. About half of the non-optimal bids have a new lowest price that is not lower than 34% of the original lowest price.

Curved scoring methods are most prone to rank reversal (14.2%), after adding a fictional non-optimal bid to each tender. For linear scoring methods this number is 10.2% (see also Table 2). In Table 2, four price weight groups have been created as well. The data indicate that rank reversal occurs less often when the weight of price is very low or very high. All occurrences of rank reversal are found when the weight of price is between 10 and 60% and most occurrences of rank reversal are found between 50 and 60%. For weights higher than 60%, there is not sufficient data available for a sound analysis. Data analysis 2 provides more results about the full range of weights.

Table 2 also shows that rank reversal occurs less often in case there is a quality threshold. This can be explained by two reasons. First, the weight for price increases compared to the weight for quality when using a threshold. On quality, suppliers can make a difference with their competitors of only 4 points on a 10-point scale if the threshold is set at 60%, but on price they can make a difference of 10 points (i.e. the importance of price increases with a factor of $10 \div 4 = 25$ when a quality threshold of 60% is introduced, see also De Boer et al., 2006). Because of this, there are less possibilities to create a non-optimal bid that causes rank reversal. Second, as will be shown in Section 4.2.2, the rank reversal rate increases with an increasing number of bids. When there is a quality threshold, there are on average less acceptable bids as not all bids will pass the threshold.

Finally, for studying the number of bids, the data are separated in two groups in order to have a sufficient number of tenders for each group. Again, the groups have been divided in such a way that their group sizes are almost equal. Group 1 contains all tenders with two or three bids (105 tenders) and has a rank reversal rate of 9.6%. Group 2 contains all tenders with four or more bids (97 tenders) and has a rank reversal rate of 17.5%, indicating that rank reversal rates increase with a higher number of bids. Tenders with one bid (9 tenders) were not included in group 1 as rank reversal is not possible for such tenders. Again, analysis 2 provides more detailed results.

Table 2
Occurrence of rank reversal after adding non-optimal bids to supplier selection models using relative scoring methods.

Relative scoring method	*Number of tenders*[a]	Rank reversal rate
Effects of different scoring methods		
Linear	49	10.2%
Curved	162	14.2%
Effects of different weights of price		
Weight price <20%	67	4.4%
Weight price between 20 and 40%	104	20.0%
Weight price between 40 and 60%	25	19.2%
Weight price>60%	15	0%
Effects of quality thresholds[b]		
Quality threshold	111	8.1%
No quality threshold	100	19.0%
Effects of different numbers of bids		
Two or three bids	105	10.5%
Four or more bids	97	17.5%

[a] In 75 tenders (out of 211) the winning bid had both the best price and quality, in which case rank reversal is not possible.

[b] In most cases, the quality threshold is 60% and in other cases the quality threshold has another value between 40 and 70%.

4.2. Analysis 2: rank reversal rates as a result of removing non-optimal bids

For data analysis 2, the occurrence of rank reversal after removing non-optimal bids from the dataset is analyzed. As in practice the rank reversal problem is mostly considered problematic in case a bid is removed, we choose to conduct several additional more detailed analyses compared to data analysis 1.

We simulated several runs of 10,000 tenders with three bids or more based on the dataset. For one sub-analysis, we conducted an extra run of 50,000 tenders (see Section 4.2.2) to be able to gather sufficient data for that analysis. These numbers of tenders allow it to study in more detail under what circumstances rank reversal occurs. The number of 10,000 is the same as used in Figueira and Roy (2009) to analyze a different rank reversal problem.

4.2.1. Occurrence of rank reversal with different price weights and quality threshold effects

Fig. 2 shows the main results of the simulations with different rates of rank reversal given a certain weight of price. Here for each weight point, a separate simulation has been conducted with 10,000 tenders. 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 (here the rank reversal rate is close to 4% if the weight of price is between 50 and 70% and if there is no threshold). After the appliance of a minimum quality threshold of 60%, the maximum rank reversal rate decreases with about 1.3%. The average rank reversal rate over all tenders in the historic dataset is 2.4%.

The probability of rank reversal converges towards zero at high weights for either price or quality. We explain this as follows. If the weight of price is close to 100%, then the bid with the lowest price almost always wins the tender. After removal of the bid with the lowest

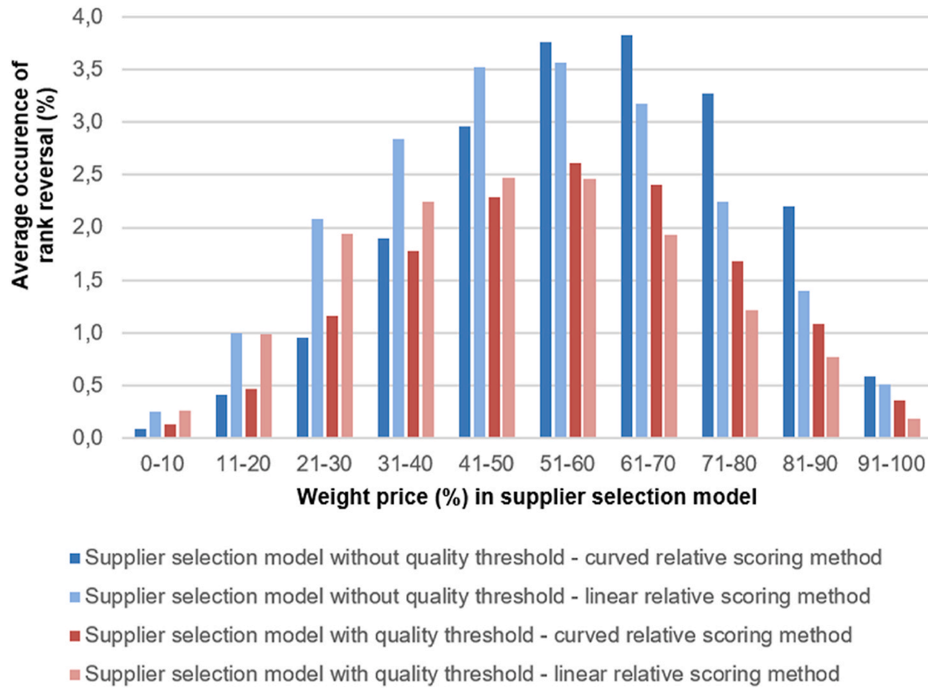

Fig. 2. Occurrence of rank reversal after removing non-optimal bids from supplier selection models - different weights of price and quality threshold effects.

price it is also likely that the bid with the second lowest price wins the tender. Hence, under such circumstances rank reversal does not occur or only in exceptional cases.

4.2.2. *Occurrence of rank reversal with different number of bids*

Fig. 3 shows rank reversal rates for different numbers of bids in a tender when there is no quality threshold while using a curved relative scoring method. Fig. 3 zooms in on this specific scoring method type as this is the most common type in our dataset (see also Table 2). The number of bids varies from three to nine as these were the minimum and maximum in the obtained dataset for which we had sufficient data. The analysis was conducted at three different weights for price, namely 40, 50 and 60%. We used these weights in our analyses because they are in the categories that are most prone to rank reversal (see also Fig. 2) and are used commonly in practice. For readability purposes other weights have not been included in the figure, but they show similar patterns.

The figure shows that rank reversal rates increase substantially when there are more bids: the average occurrence of rank reversal more than doubles from 2.1 (three bids) to 4.8% (eight bids). As there are more bids the probability of one bid with a differentiated price offer increases. After removal of this bid with the lowest price the differences in scores between the other bids become larger and rank reversal is more likely to happen. Note that most tenders in the model receive three up to six bids (about 90%) explaining the higher variability in rank reversal for tenders that receive seven bids or more.

4.2.3. *Occurrence of rank reversal with different variability of bid prices*

Fig. 4 shows the rank reversal rates for different standard deviations and different price weights when there is no quality threshold while using a curved relative scoring method. A standard deviation of 20%,

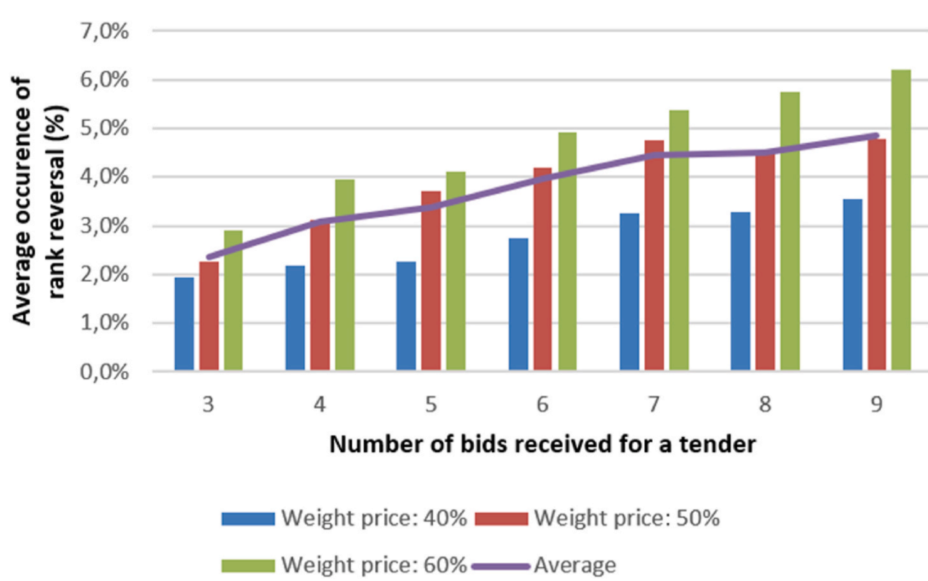

Fig. 3. Occurrence of rank reversal after removing non-optimal bids from supplier selection models using curved relative scoring methods - different numbers of bids.

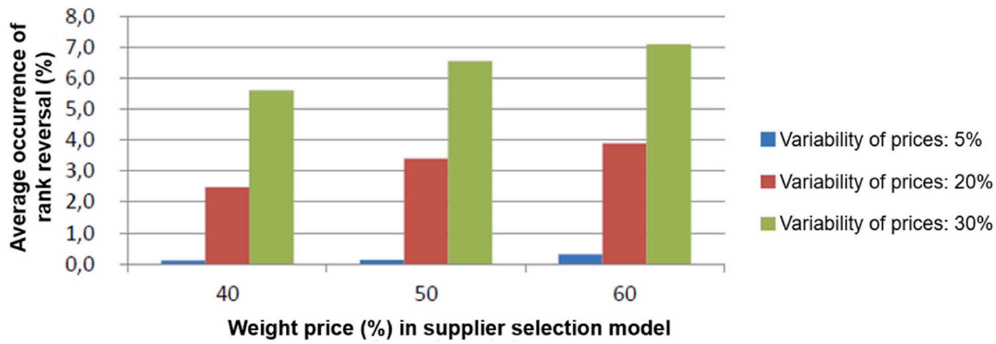

Fig. 4. Occurrence of rank reversal after removing non-optimal bids from supplier selection models using curved relative scoring methods - different variability of prices.

corresponding with the red graph, was used as default value in the previous simulation runs. When the standard deviation is set at 5% instead of the default of 20% the rank reversal rate decreases to values close to 0%. The rank reversal rate more than doubled on average after setting the standard deviation to 30%.

5. Discussion and conclusion

This final section starts with a discussion about the added value of this research for buyers. Second, the effects of relative scoring methods on strategic bid behavior by suppliers and overall bid value are described. Third, limitations and suggestions for future research are proposed. Finally, several managerial implications are presented.

5.1. Theoretical implications for buyers

Due to the nature of most relative scoring methods in combination with WSM, removing or adding a bid can change which bid wins a tender. Past research on rank reversal has focused on, among other things, disadvantages of rank reversal and methods to prevent rank reversal. In addition, rank reversal is mentioned often in decision theory related literature as a problematic property or a remarkable and mostly unwanted effect (e.g. Saaty and Vargas, 1984; De Boer et al., 2006; Chen, 2008; Manunza, 2018; Sciancalepore et al., 2011; Smith, 2010; Sykes, 2012; Mufazzal and Muzakkir, 2018).

Despite this earlier research and despite that several methods are known that can prevent rank reversal, relative scoring methods that allow rank reversal are popular in supplier selection practice in both the public and private sector. When such methods are used, it is our experience that it is often argued that rank reversal is a theoretical problem that rarely occurs in practice or not in the specific circumstances of the tender at hand. As far as we know, current literature does not provide a sound answer to this type of criticism. Earlier empirical research to rank reversal in the context of supplier selection has some methodological issues as discussed in Section 2.3 and only indicates an average percentage of rank reversal after bid removal (Stilger et al., 2017). However, it does not indicate how often rank reversal can occur after adding a new fictional bid and it does not indicate under what circumstances (such as different weights) rank reversal is more likely to occur. This study adds new insights to decision theory related literature about these topics using data from the real world, further improving our understanding of the effects of relative scoring methods in combination with WSM.

The research shows that rank reversal is not just a theoretical problem when buyers use WFS in combination with a relative scoring method. When removing bids from the dataset, rank reversal occurs in about 1 out of 40 of the studied tenders when a curved relative scoring method is used. This number confirms the results of earlier research conducted by Stilger et al. (2017). Our study also shows that after adding non-optimal fictional bids to the dataset, in about 1 out of 5 of the studied tenders rank reversal occurs if there is no quality threshold and when a curved or linear relative scoring method is used. These rank reversal rates imply that buyers who use relative scoring methods create some randomness in their tenders: which bid wins a tender can be (or could have been) influenced by a non-optimal bid. A bid can win or lose a tender not because of economic reasons, but only because another non-optimal bid is or is not submitted. In other words: a supplier submitting a bid that could win a tender can be lucky or not whether a non-optimal bid is submitted by another supplier.

Regarding the specific circumstances under which rank reversal can occur while removing bids, we add several new insights to decision theory for the specific context of supplier selection problems. First, the research shows that a curved relative scoring method is more prone to rank reversal than a linear one as defined in Section 3.2.1. Second, the research shows that rank reversal rates increase when (i) the number of bids increases and (ii) bid price variance increases, (iii) price weights increase from 0 to about 60% or decrease from 100 to about 60%, and (iv) there is no quality threshold. Regarding the number of bids, our research supports earlier theoretical findings of Zanakis et al. (1998) who studied rank reversal for other decision methods than studied in this article, using randomly generated scores from a uniform distribution. Zanakis et al. (1998) also found that rank reversal occurs more often in decision problems with more alternative solutions. As far as we know, the other indicators for increased or decreased possibilities of rank reversal have not been shown before in decision theory related literature.

The empirical evidence we present in this article about rank reversal in commonly used supplier selection models closes an important gap in decision theory related literature. Normative decision theory proposes several methods to make optimal supplier selection decisions and mostly acknowledges that rank reversal can lead to flawed decision making. However, this is apparently not sufficient to change actual buying behavior substantially, given the popularity of relative scoring methods in combination with WSM. An important missing piece in the puzzle was research to the actual occurrence of rank reversal. If rank reversal would not occur or only occasionally, using relative scoring methods would not be very problematic (based on Saaty, 1994; Millet and Saaty, 2000). This study adds to decision theory related literature that rank reversal does occur in the supplier selection problems studied. It also shows when rank reversal is likely to occur, depending on the number of bids, the weights of price, et cetera. We believe these insights to be important prerequisites for changing buying behavior of procurement officers. This is also illustrated by the case organization HIS. The findings of this research were used for the initiation of a change program resulting in a drastically reduced usage of relative scoring methods. Nowadays, HIS uses relative scoring methods only in exceptional cases. Although another theoretical article about a method solution to rank reversal or about theoretical conditions for rank reversal would have created even more insight into the issue, we expect that this would not have initiated such a change. We therefore believe to add important and relevant insights to decision theory in the context of supplier selection.

5.2. Theoretical implications for suppliers

For the supply side, our research also has implications regarding bid strategies of suppliers. Earlier research indicated theoretically that suppliers must divide their performance over all award criteria – while taking weights into account – instead of focusing on scoring very high on one or a few criteria (Mummalaneni et al., 1996). However, we argue that this might not be the best strategy when a buyer uses a relative scoring method. Depending on the expected bids of competitors and the rest of the supplier selection model, it can be a better strategy to aim for a very high score or settle for a low score for a specific award criterion. For instance, when multiple linear relative scoring methods are used by a buyer, a supplier could offer a very low price for one criterion to maximize score differences with competitors, and offer a very high price for another criterion to secure a positive business case. This could result in more points for the supplier compared to dividing performances, while the buyer receives less value for money.

Thus, relative supplier selection models create a risk that suppliers will bid strategically, according to the expected bids of other suppliers, instead of optimizing their bid according to their business model. Note that this means that there is not only the risk of flawed decision making, an often mentioned topic in decision theory (e.g. Saaty and Vargas, 1984; Chen, 2008; De Boer et al., 2006; Smith, 2010; Stilger et al., 2017; Sykes, 2012; Mufazzal and Muzakkir, 2018; Wang and Luo, 2009), but also that the pool of bids (i.e. the pool of alternatives) could be of lower total value as a result of the supplier selection model. As far as we know, this issue has not yet been addressed in decision theory related literature about supplier selection. It does provide an explanation for earlier findings of Albano et al. (2008) and Telgen and Schotanus (2010) who found that relative scoring methods can lead to worse price-quality ratios compared to absolute scoring methods. Note that this does not only apply to WSM in combination with relative scoring methods, but also to other relative selection methods.

5.3. Limitations and future research

The analyses in this study depend on the initial sample of tenders. Although we analyzed a large number of different tenders, there is a risk that the data are not fully representative. A more stringent test would require a larger dataset covering tenders from different countries and more sectors. More data would also allow to test the effects of different weights of price to a larger extent and to test the effects of other price functions in the simulations. Nevertheless, as we can explain the results and as the findings of analysis 1 and analysis 2 support each other, we do not expect that new research would lead to other findings regarding the characteristics studied (e.g. the effects on rank reversal rates when the number of bids increases). We expect the rank reversal percentages to be most prone to a possible bias. Nevertheless, the overall rank reversal rate found in the study is in line with earlier research of Stilger et al. (2017).

Apart from the future research suggestions related to the limitations, we have a few additional recommendations. First of all, we believe it is relevant to study whether other relative scoring methods give similar results and patterns as found in this study. Other types of rank reversal as discussed in Section 2.1.2 could also be studied. Second, we suggest to research to what extent and under what circumstances absolute scoring methods in combination with WSM or other supplier selection methods that do not allow rank reversal lead to better price-quality ratios than when using relative scoring methods in combination with WSM. Third, it would be useful to study to what extent different supplier selection methods are used by public buyers in different countries. This would also allow more academic research aimed at specific needs of public buyers.

5.4. Managerial implications

This research has several managerial implications. First we discuss the implications for the private sector and then the implications for the public sector.

Although this research focuses on the public sector, reducing an important research gap regarding this sector (De Boer et al., 2001; Wu and Barnes, 2011), the findings also apply to the private sector and are relevant to suppliers as well. We advise private organizations that use formal supplier selection models to use supplier selection models that do not allow rank reversal, such as absolute scoring methods in combination with WSM or otherwise other alternatives proposed in the literature. As we argue in this article, this will lead to better bids and to better price-quality ratios. For suppliers, it is important to realize that offering their best bid possible might not be in their best interest (see also Section 5.3). When buyers use selection models that allow rank reversal, strategic bidding could increase win rates.

Regarding the public sector, our results – in combination with earlier studies of Albano et al. (2008), Manunza (2018) and Telgen and Schotanus (2010) – have important implications for legislators, policy makers, and public buyers. We argue that award models that allow rank reversal should not be allowed to use in the context of public procurement, or otherwise only in exceptional cases (e.g. when the weight of price if very low or high and when the buyer has limited information about price ranges in the market at hand). This would also be in line with legislation already present in Portugal (Mateus et al., 2010). Before such legislation is in place, policy makers and contracting authorities could implement rules, guidelines or strongly recommend to use alternative award models that do not allow rank reversal. It is also possible to only include award models without rank reversal possibilities in procurement templates. To be able to use such methods, top management support and proper training of procurement practitioners is required and guidance is necessary for project teams responsible for procurement that do not include procurement professionals.

If our advice to – in principle – prohibit relative scoring methods that allow rank reversal is implemented in public procurement laws worldwide, it will have a large effect on public procurement practice and supplier bid behavior. It will end current supplier selection 'randomness' in which the winners of tenders are not only based on which supplier offers the best economical bid, but also on whether or not a certain non-optimal bid is submitted. In addition, it will reduce strategic bid behavior as discussed in Section 5.3, it will prevent issues with equal treatment and transparency, and it ought to lead to bids that better fit with the needs of buyers. In the end, this should lead to increased overall bid value and more value for money for public buyers.

Author statement

This academic article is the result of an equal share of work and responsibility of the authors.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

References

Albano, G.L., Dini, F., Zampino, R., August . Suppliers' Behavior in Competitive Tendering: Evidence from the Italian Ministry of Economy and Finance's Acquisitions of IT Services. Proceedings of the 3rd IPPC Conference, Amsterdam (The Netherlands).

Aissaoui, N., Haouari, M., Hassini, E., 2007. Supplier selection and order lot sizing modeling: a review. Comput. Oper. Res. 34 (12), 3516–3540.

Belton, V., Gear, T., 1983. On a short-coming of Saaty's method of analytic hierarchies. Omega 11 (3), 228–230.

Belton, V., Gear, T., 1985. The legitimacy of rank reversal – a comment. Omega 13 (3), 143–144.
Bergman, M.A., Lundberg, S., 2013. Tender evaluation and supplier selection methods in public procurement. J. Purch. Supply Manag. 19 (2), 73–83.
Brun, A., Pero, M., 2011. Assessing suppliers for strategic integration: a portfolio approach. Int. J. Bus. Excel. 4 (3), 346–370.
Bruno, G., Esposito, E., Genovese, A., Passaro, R., 2012. AHP-based approaches for supplier evaluation: problems and perspectives. J. Purch. Supply Manag. 18 (3), 159–172.
Chen, T., 2008. An economic approach to public procurement. J. Public Procure. 8 (3), 407–430.
Choi, T.Y., Hartley, J.L., 1996. An exploration of supplier selection practices across the supply chain. J. Oper. Manag. 14 (4), 333–343 (11).
Court of Arnhem, 18 December 2012. ECLI:NL:GHARN:2012:BX9806, 200.096.019.
De Boer, L., Labro, E., Morlacchi, P., 2001. A review of methods supporting supplier selection. Eur. J. Purch. Supply Manag. 7 (2), 75–89 (15).
De Boer, L., M. Linthorst, F. Schotanus, J. Telgen, An Analysis of Some Mistakes, Miracles, and Myths in Supplier Selection, Working Paper, Proceedings of the 15th IPSERA Conference, San Diego (United States), 6-8 April 2006.
De Boer, L., van der Wegen, L., Telgen, J., 1998. Outranking methods in support of supplier selection. Eur. J. Purch. Supply Manag. 4 (2), 109–118 (10).
De Farias Aires, R.F., Ferreira, L., 2018. The rank reversal problem in multi-criteria decision making: a literature review. Pesqui. Oper. 38 (2), 331–362.
De Farias Aires, R.F., Ferreira, L., 2019. A new approach to avoid rank reversal cases in the TOPSIS method. Comput. Ind. Eng. 132, 84–97.
Dyer, S.S., 1990. Remarks on the analytic hierarchy process. Manag. Sci. 36 (3), 249–258, 1990.
Figueira, J.R., Roy, B., 2009. A note on the paper, "Ranking irregularities when evaluating alternatives by using some ELECTRE methods". by Wang and Triantaphyllou. Omega 37 (3), 731–733.
Finan, J.S., Hurley, W.J., 2002. The analytic Hierarchy process: can wash criteria be ignored? Comput. Oper. Res. 29 (8), 1025–1030.
Kumar, R., Padhi, S.S., Sarkar, A., 2019. Supplier selection of an Indian heavy locomotive manufacturer: an integrated approach using Taguchi loss function, TOPSIS, and AHP. IIMB Management Review 31 (1), 78–90.
Leung, L.C., Cao, D., 2001. On the efficacy of modeling multi-attribute decision problems using AHP and Sinarchy. Eur. J. Oper. Res. 132 (1), 39–40.
Lima Junior, F.R.L., Osiro, L., Carpinetti, L.C.R., 2014. A comparison between Fuzzy AHP and Fuzzy TOPSIS methods to supplier selection. Appl. Soft Comput. 21, 194–209.
Luzzini, D., Caniato, F., Spina, G., 2014. Designing vendor evaluation systems: an empirical analysis. J. Purch. Supply Manag. 20 (2), 113–129.
Manunza, E., 2018. Achieving a sustainable and just society through public procurement?: on the limits of relative scoring and the principles of equal treatment and transparency. In: Manunza, E., Schotanus, F. (Eds.), The Art of Public Procurement. Liber Amicorum in Honour of Prof. Dr. Jan Telgen', Enschede.
Masi, D., Micheli, G.J., Cagno, E., 2013. A meta-model for choosing a supplier selection technique within an EPC company. J. Purch. Supply Manag. 19 (1), 5–15.
Mateus, R., Ferreira, J., Carreira, J., 2010. Full disclosure of tender evaluation models: background and application in Portuguese public procurement. J. Purch. Supply Manag. 16, 208.
Millet, I., Saaty, T.L., 2000. On the relativity of relative measures–accommodating both rank preservation and rank reversals in the AHP. Eur. J. Oper. Res. 121 (1), 205–212.
Mufazzal, S., Muzakkir, S., 2018. A new multi-criterion decision-making (MCDM) method based on proximity indexed value for minimizing rank reversals. Comput. Ind. Eng. 119, 427–438.
Mummalaneni, V., Dubas, K.M., Chao, C.N., 1996. Chinese purchasing managers' preferences and trade-offs in supplier selection and performance evaluation. Ind. Market. Manag. 25 (2), 115–124.
Munson, C.L., Rosenblatt, M.J., 1997. The impact of local content rules on global sourcing decisions. Prod. Oper. Manag. 6 (3), 277–290 (14).
Peterson, M., 2017. An Introduction to Decision Theory. Cambridge University Press.
Raga, K., Taylor, J.D., 2010. Legislative and Administrative Directives Governing Procurement Procedures: A Case Study for South Africa.
Ren, L., Zhang, Y., Wang, Y., Sun, Z., 2007. Comparative Analysis of a Novel M-TOPSIS Method and TOPSIS. Applied Mathematics Research eXpress.
Robinson, S., 2014. Simulation: the Practice of Model Development and Use, second ed. Palgrave Macmillan.
Rodriguez, A., Ortega, F., Concepción, R., 2013. A method for the selection of customized equipment suppliers. Expert Syst. Appl. 40 (4), 1170–1176.
Saaty, T.L., 1994. How to make a decision: the analytic hierarchy process. Interfaces 24 (6), 19–43.
Saaty, T.L., Vargas, L.G., 1984. The legitimacy of rank reversal. Omega 12 (5), 513–516.
Saaty, T.L., Sagir, M., 2009. An essay on rank preservation and reversal. Math. Comput. Model. 49 (5–6), 1230–1243.
Schenkerman, S., 1994. Avoiding rank reversal in AHP decision-support models. Eur. J. Oper. Res. 74 (3), 407–419.
Sciancalepore, F., Falagario, M., Costantino, N., Pietroforte, R., 2011. Multiple-Criteria Bid Evaluation of Public Projects. In: Paper Presented at Conference Management and Innovation for a Sustainable Built Environment, Amsterdam, the Netherlands.
Smith, P., 2010. Evaluating bids and tenders - the scoring trap. Danish Purchasing and Logistics Forum 47, 34–39.
Smytka, D.L., Clemens, M.W., 1993. Total cost supplier selection model: a case study. Int. J. Purch. Mater. Manag. 29 (1), 42–49.
Stilger, P.S., Siderius, J., Van Raaij, E.M., 2017. A comparative study of formulas for choosing the economically most advantageous tender. J. Public Procure. 17 (1), 89–125.
Sykes, M., 2012. Can the use of price versus non-price factors in the evaluation of tenders undermine the pursuit of quality and value?. In: Proceedings of the 21st IPSERA Annual Conference, Naples, Italy.
Telgen, J., Schotanus, F., 2010. Supplier Selection Requires Full Transparency. In: Paper Presented at 19th Annual IPSERA Conference, Lapeenranta, Finland.
Telgen, J., Timmer, J.B., 2016. Conditions for rank reversal in supplier selection. Memoranda 2055.
Triantaphyllou, E., Mann, S.H., 1989. An examination of the effectiveness of multi-dimensional decision-making methods: a decision-making paradox. International Journal of Decision Support Systems 5 (3).
Wang, Y.M., Elhag, T.M., 2006. An approach to avoiding rank reversal in AHP. Decis. Support Syst. 42 (3), 1474–1480.
Wang, Y.M., Luo, Y., 2009. On rank reversal in decision analysis. Math. Comput. Model. 49 (5–6), 1221–1229.
Wang, X., Triantaphyllou, E., 2008. Ranking irregularities when evaluating alternatives by using some ELECTRE methods. Omega 36, 45–63.
Weber, C.A., Current, J.R., 1993. A multiobjective approach to vendor selection. Eur. J. Oper. Res. 68, 173–184.
Wu, C., Barnes, D., 2011. A literature review of decision-making models and approaches for partner selection in agile supply chains. J. Purch. Supply Manag. 17 (4), 256–274.
Yu, Q., Hou, F., 2016. An approach for green supplier selection in the automobile manufacturing industry. Kybernetes 45 (4), 571–588.
Zanakis, S., Mandakovic, T., Gupta, S., Sahay, S., Hong, S., 1995. A review of program evaluation and fund allocation methods within the service and government sectors. Soc. Econ. Plann. Sci. 29, 59–79.
Zanakis, S.H., Solomon, A., Wishart, N., Dublish, S., 1998. Multi-attribute decision making: a simulation comparison of select methods. Eur. J. Oper. Res. 107 (3), 507–529.
Zhang, K., Kluck, C., Achari, G., 2009. A comparative approach for ranking contaminated sites based on the risk assessment paradigm using fuzzy PROMETHEE. Environ. Manag. 44 (5), 952–967.
Žižović, Mališa, Pamučar, Dragan, Albijanić, Miloljub, Chatterjee, Prasenjit, Pribićević, Ivan, 2020. Eliminating rank reversal problem using a new multi-attribute model—the rafsi method. Mathematics 8 (6), 1015. https://doi.org/10.3390/math8061015.