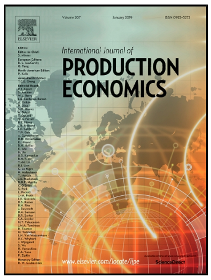
Accepted Manuscript



The Voting Analytic Hierarchy Process Revisited: A Revised Method with Application to Sustainable Supplier Selection

Grigory Pishchulov, Alexander Trautrims, Thomas Chesney, Stefan Gold, Leila Schwab

PII:	S0925-5273(19)30035-0
DOI:	10.1016/j.ijpe.2019.01.025
Reference:	PROECO 7276
To appear in:	*International Journal of Production Economics*
Received Date:	16 April 2018
Accepted Date:	20 January 2019

Please cite this article as: Grigory Pishchulov, Alexander Trautrims, Thomas Chesney, Stefan Gold, Leila Schwab, The Voting Analytic Hierarchy Process Revisited: A Revised Method with Application to Sustainable Supplier Selection, *International Journal of Production Economics* (2019), doi: 10.1016/j.ijpe.2019.01.025

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

The Voting Analytic Hierarchy Process Revisited: A Revised Method with Application to Sustainable Supplier Selection

Grigory Pishchulov, Alexander Trautrims, Thomas Chesney, Stefan Gold, Leila Schwab

Authors and contact information

Grigory Pishchulov (🖂 — Corresponding author)
Alliance Manchester Business School, The University of Manchester, Manchester M1 3BU, UK
St. Petersburg State University, 7/9 Universitetskaya nab., St. Petersburg, 199034 Russia
grigory.pishchulov@manchester.ac.uk

Alexander Trautrims
Nottingham University Business School, Nottingham NG8 1BB, UK
alexander.trautrims@nottingham.ac.uk

Thomas Chesney
Nottingham University Business School, Nottingham NG8 1BB, UK
thomas.chesney@nottingham.ac.uk

Stefan Gold
Faculty of Economics and Management, University of Kassel, Germany
gold@uni-kassel.de

Leila Schwab
Chair for Operations, Production and Logistics Management
University of Neuchâtel, Switzerland
leila.schwab@unine.ch

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

The Voting Analytic Hierarchy Process Revisited: A Revised Method with Application to Sustainable Supplier Selection

Abstract

Civil society increasingly holds focal companies accountable for ensuring socially and environmentally sustainable production standards among their supply base. These standards entail increased levels of complexity to be addressed by appropriately designed tools, such as the Voting Analytic Hierarchy Process (VAHP) proposed by Liu and Hai (2005). This method of multi-criteria group decision making structures decision criteria in a hierarchical fashion as per Saaty's Analytic Hierarchy Process (AHP) and employs data envelopment analysis (DEA) for deriving criteria weights from the ordinal preferences of the group members. Compared to AHP, the method permits a simpler application in a group decision context. However, its theoretical underpinnings have been questioned in the literature. This specifically concerns (i) the requirement of a strong convex order for the importance weights of ordinal rank gradations, and (ii) the choice of discrimination threshold for consecutive rank weights in the underlying DEA model. We propose a revised VAHP method that overcomes both issues (i) and (ii) by pursuing a game-theoretic approach to elicitation of criteria weights — so as to remove subjectivity from rank discrimination. We illustrate the application of the method on a real-world problem of sustainable supplier selection. We contribute to theory by proposing a more robust VAHP tool that helps supply chain and purchasing managers selecting suppliers based on a comprehensive set of criteria spanning all three sustainability dimensions (economic, environmental, and social), while coping with parsimonious input by group decision-makers.

Keywords: supplier selection, sustainability, multi-criteria decision making, group decisions, ordinal preferences, preference aggregation, game theory.

1. Introduction

Reduction of trade barriers, digitalisation and related communication technologies as well as efficient long-haul carriage has enabled companies to tap comparative cost advantages by increasing their ratio of international sourcing, in particular from emerging and developing economies (Lund-Thomsen and Lindgreen 2014). While international sourcing from other continents has become a strategic option even for small and medium-sized enterprises located in the industrialized world (Rodríguez and Nieto 2016), the collateral damage of this business strategy has received increasing societal attention through media coverage of unethical business conduct, labour and human rights violation, environmental pollution, and product safety issues (Wolf 2014). Often referring to the concept of sustainability as defined and popularized by the Brundtland report (WCED 1987), NGOs, trade unions and other pressure groups (forging alliances with consumer groups at times) have vigorously denounced those adverse side-effects of international business and have lifted them up the political agenda of developed countries (Lund-Thomsen and Lindgreen 2014).

Companies have responded to related threats to their brand image and their 'social license to operate' (Demuijnck and Fasterling 2016) mostly by ensuring minimum standards throughout their supply chains; however, even minimum standards are not easy to achieve given supply networks' complexity, buyer–supplier distance and lack of transparency (Gold and Heikkurinen 2018). While the extraordinary challenge of managing suppliers for sustainability is rather non-controversial, the significant role of supplier selection and evaluation as preconditions of effective supplier management is also beyond dispute. Scholars have taken up the challenge and developed tools of sustainable supplier selection, in particular for the concluding step of the four-step supplier selection process according to De Boer et al. (2001) — following on from problem definition, formulation of selection criteria and preselection of candidates. Zimmer et al. (2016) show in their review paper that these selection tools often do not consider all three dimensions of sustainability equally but neglect social aspects. Still, recently there has been a growing number of supplier selection approaches that simultaneously integrate criteria from the triple bottom line (Dyllick and Hockerts 2002) — comprising the social, environmental, and economic dimension (e.g., Awasthi et al. 2018, Bai and Sarkis 2010, Kannan et al. 2015). Although managers often prefer making decisions based on intuition or heuristics, Ishizaka and Siraj (2018) have used incentive-based experiments to demonstrate the — also subjectively perceived — usefulness of decision support by multi-criteria decision-making tools.

Given the challenges of sustainable supplier selection regarding data availability, credibility and uncertainty (Awasthi et al. 2018), committees of experts are well suited for this task as they hold complementary opinions and experience that reduce the risk of biased decisions. The members of those committees of experts are frequently recruited from across various firm-internal functions and may be complemented by external stakeholders (Watkins 1999). The advantages of group decision making in terms of reduced bias and increased organisational legitimacy of the final decision face its disadvantages in terms of the workload associated with the elicitation of the experts' preferences and the complexity of aggregating individual preferences. Workload and complexity augment strongly with the number of supplier selection criteria, which rise exponentially with each added dimension of sustainability.

The supplier selection problem has been tackled by a wide variety of methods proposed in the literature including multi-objective programming, analytical hierarchy process (AHP), analytical network process (ANP), data envelopment analysis (DEA) and artificial neural networks. Among those, the Voting Analytic Hierarchy Process (VAHP) has gained considerable attention, as it is a multi-criteria group decision-making method featuring reduced complexity in comparison to AHP while maintaining its systematic nature (Liu and Hai 2005). To this end, the VAHP method employs a DEA approach for deriving criteria weights from the ordinal preferences of the group members. However, its theoretical underpinnings have been questioned in the literature (Llamazares and Peña 2009, Wang et al. 2007). This specifically concerns (i) the requirements imposed on the weights of ordinal rank gradations in the underlying DEA model, and (ii) arbitrariness of discrimination between ordinal rank gradations. Based on a game-theoretic approach to deriving criteria weights (Tüselmann et al. 2015), we propose a revised VAHP method that overcomes both issues. We contribute to theory by proposing a more robust VAHP tool that copes with a comprehensive set of decision criteria as well as parsimonious input by group decision-makers, which describes the specific decision situation of sustainable supplier selection that we use as an application case.

The remainder of the paper is structured as follows. Section 2 offers an overview of related literature. In Section 3 we present the detailed rationale and procedure of the VAHP method. Section 4 critically discusses its shortcomings and proposes a revised method for overcoming them. Section 5 presents application of the revised method to sustainable supplier selection at a real-world company. Section 6 discusses results of the application and concludes with avenues for future research.

2. Literature Review

Our study is most closely related to three research areas: supplier selection criteria and methods, sustainable supplier selection, and preference aggregation using data envelopment analysis. In the stream of supplier selection literature, the seminal work by Dickson (1966) has offered an important insight into the criteria used by companies for evaluating and selecting suppliers. By conducting a survey among 170 procurement professionals, Dickson identified a list of 23 such criteria, which have further been rated by their average importance as perceived by the respondents (Yahya and Kingsman 1999). Dickson's results reveal that the price of goods has not been seen as the most important criterion of supplier selection; instead, factors such as quality, delivery and performance history have been found to be of a higher importance because they may have a stronger economic impact on the performance of the buying company than the price charged by the supplier. A literature review conducted by Weber et al. (1991) over two decades later has nevertheless revealed that the criterion most frequently used in academic research has still been the price, followed by delivery and quality. This seems to have changed over the next two decades, as the literature review by Ho et al. (2010) reveals the most popular criteria in the literature to be quality, delivery, and price, followed by manufacturing capability and service.

Evidence suggests that the criteria list by Dickson (1966) and their relative importance has been sustained for a long time in industry practice (Yahya and Kingsman 1999). Still, new criteria have entered the list over time, reflecting new developments in industry. Weber et al. (1991) highlight in this regard Just-in-Time (JIT) criteria. Apart from JIT, Cheraghi et al. (2004) further identified supplier flexibility and reliability as new criteria referred to by academic research, among others. A survey of literature by Thiruchelvam and Tookey (2011) shows that criteria of environmental and social responsibility have received some attention over the subsequent decade. As work on sustainable supply chain management has recently seen a significant growth (Rajeev et al. 2017), this has been reflected in the stream of literature on sustainable supplier selection. Zimmer et al. (2016) offer a recent comprehensive review of this stream of literature and identify 448 unique supplier selection criteria used across 143 research publications over the time span from 1997 to 2014. Of these, 52.5% of criteria refer to the economic dimension of sustainability, 38.1% to the environmental one, and only 9.4% to the social dimension. Indeed, 59% of the studies reviewed address economic and environmental dimensions only. In this regard, Zimmer et al. (2016) stress the insufficient attention that the social criteria have received in the literature to date, which can be explained by difficulties in measurement and quantification of social aspects, as well as different political, ethical and ideological attitudes of stakeholders (Zimmer et al. 2016).

Wetzstein et al. (2016) emphasise the ever-growing importance of supplier selection as supply chains became more international, a larger proportion of operations are outsourced, and more criteria making the selection increasingly complex. Despite the strategic commitment to

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

improve the sustainability of their supply chains, in everyday practice many companies struggle to incorporate sustainability aspects in their supplier selection decisions and the ultimate selection decision is still often dominated by economic considerations (Karjalainen and Salmi, 2013; Genovese et al. 2013) and a lower price may be achieved at the risk of lower environmental or ethical standards (Goebel et al. 2012). However, at a strategic sourcing level — which ought to precede the supplier selection decision — companies will direct the selection process by focussing for example on geographies and supply market structures, and categorise supplier on their capabilities, for example the ability to innovate or responsiveness (Li and Shao 2015; Trautrims et al. 2017).

The supplier selection process is usually conducted in four stages as described for example by Cousins et al. (2000): initial supplier qualification, agreement of measurement criteria, obtainment of relevant information, and the selection itself. One can however argue that the first step of initial supplier qualification is already pre-selecting suppliers on set criteria. At this stage compliance to labour standards and other sustainability considerations can be included. Suppliers who fail to satisfy the initial supplier qualification will be excluded from further stages in the process. This approach can be considered the procurement equivalent to Hill's (1985) concept of order-qualifiers and order-winners, in which suppliers who do not satisfy the qualification criteria will not proceed to the stage where the order winners come into play.

Given the multi-criteria nature of the subsequent supplier selection step, the literature offers a broad range of methods for decision-making support. Yahya and Kingsman (1999) offer an elaborate discussion of strengths and weaknesses of various methods from simple, like categorical and weighted-average methods, to more advanced, such as data envelopment analysis (DEA), which is instrumental in overcoming the issue of subjective criteria weighting. They further highlight the strengths of the Analytic Hierarchy Process (AHP) in structuring evaluation criteria in the form of a hierarchy and eliciting expert preferences on the relative importance of criteria. Indeed, Ho et al. (2010) found out that the most popular decision-making approaches used in the literature have been DEA and AHP, followed by goal programming, while combinations of AHP and DEA represented the most popular integrated approach. Findings by Zimmer et al. (2016) in the area of sustainable supplier selection support this, as their work reveals the dominance of AHP, Analytic Network Process, and DEA models.

As indicated in the Introduction, the Voting Analytic Hierarchy Process (VAHP), proposed originally by Liu and Hai (2005), represents an integrated method of supplier selection that proves useful in group decision-making, especially when a large number of criteria is involved. The method features a combination of the AHP approach for structuring the criteria in the form of a hierarchy, and DEA for deriving group preference with regard to criteria from the ordinal preferences of individual group members, who are required to submit their preferences in a ranked voting system. This approach significantly reduces the workload in comparison to AHP. To date a broad number of approaches to preference aggregation in ranked voting systems have been proposed in the literature, see Cook and Kress (1990) and Llamazares and Peña (2009) for a review. The principal paradigm adopted in the recent research work in this area is that of the DEA, which helps to overcome the decision-maker's subjective bias in multi-attribute performance evaluation and achieve a fair assessment (Charnes et al. 1978). Cook and Kress (1990) have been the first to apply this to aggregation of ordinal preferences; their approach has further been revised and extended by follow-up research (Green et al. 1996, Hashimoto 1997, Noguchi et al. 2002, Wang et al. 2007). Liu and Hai (2005) have specifically adopted the DEA approach of Noguchi et al. (2002) in implementing the VAHP method. However, as indicated in the Introduction, its theoretical underpinnings have been questioned in the literature (Llamazares and Peña 2009, Wang et al. 2007). Our work accordingly proposes a revised method following Tüselmann et al. (2015) that overcomes this criticism and offers a more robust supplier selection tool, which we test in a real-world application. In the next section we present the detailed rationale and procedure of the VAHP method.

3. The Voting Analytic Hierarchy Process

The Voting Analytic Hierarchy Process (VAHP) has been proposed by Liu and Hai (2005) as a method of multi-criteria decision making that combines essential elements of the Analytic Hierarchy Process (Saaty 1980) and data envelopment analysis (Charnes et al. 1978). It is specifically designed to suit decision making by committees of experts and has been framed by Liu and Hai as a supplier selection method in the following six steps:

Step 1: Determine criteria for supplier evaluation. In the initial step, the committee has to agree upon the list of specific criteria according to which the candidate suppliers should be evaluated. This list can be based on the existing research work as reviewed in Section 2, as

ACCEPTED MANUSCRIPT

well as involve criteria suggested by the committee members based on their professional experience. This step may involve two rounds: first, the experts are surveyed for the criteria that they consider to be relevant for supplier selection and second, all of the suggested criteria are discussed in a joint meeting to clarify their meaning and operationalisation, and settle on the ultimate criteria list (Yahya and Kingsman 1999).

Step 2: Structure the hierarchy of the criteria. In the second step, the Analytic Hierarchy Process is utilised for structuring the criteria in a hierarchical fashion. In this way, certain criteria may be clustered together and made subordinate to the others according to their definition, as agreed upon in Step 1. This yields a hierarchy of *criteria* and *sub-criteria*. An example of a criteria hierarchy, as used by Liu and Hai (2005) based on Yahya and Kingsman (1999), is shown in Figure 1. The purpose of producing such a hierarchy is, in the sense of AHP, to break down the complex decision problem into constituent parts and thus help the experts deliver reliable judgements by dealing with elements of the same order of magnitude in each level of the hierarchy (Saaty 2012).

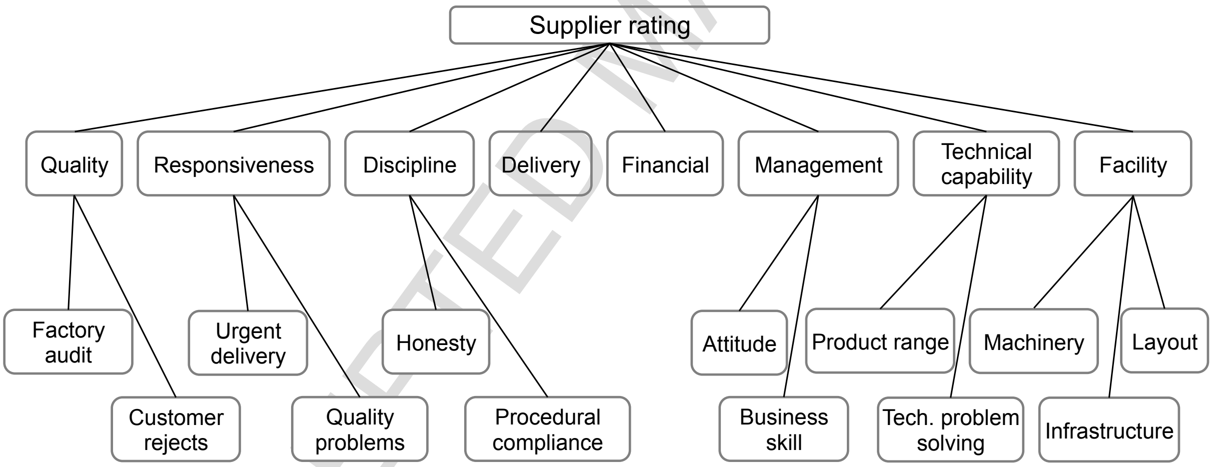


Figure 1. Hierarchy of criteria used by Liu and Hai (2005) and Yahya and Kingsman (1999).

Step 3: Vote on the importance of criteria and sub-criteria. In this step, the committee members are required to submit their individual preferences with respect to the relative importance of criteria and sub-criteria. The principal feature of the VAHP that distinguishes it from the AHP is that it foregoes the pairwise comparisons between the elements at each level of the hierarchy. Instead, the method requires the committee members to submit their preferences as ordinal rankings. Specifically, each member has to rank order the criteria according to their importance for achieving the superior goal of selecting the most suitable supplier, as perceived by him or her. Further down the criteria hierarchy, each member has to rank order the sub-criteria according to their importance for satisfying the parent criterion with which they are associated. Thus the child nodes of every branched node in the criteria hierarchy become ranked by each committee member according to the importance attached to the children in representing the parent. In this way, the committee members are voting on the importance of the respective criteria or sub-criteria, which represents a *ranked voting system* (Llamazares and Peña 2009). Assuming that the committee comprises n members, this step accordingly yields n preference rankings at each branched node of the criteria hierarchy.

Let further K represent the number of children of a specific branched node in the hierarchy. As devised by Liu and Hai (2005), the VAHP method offers the flexibility of restricting the length of rankings to any number L of places, where $L \leq K$, so that the committee members are required to rank order only their L most preferred criteria or sub-criteria out of K given.

Step 4. Derive the importance ratings of criteria and sub-criteria. In the fourth step, the individual preferences submitted by the committee members have to be aggregated to rate the importance of criteria and sub-criteria from the perspective of the committee as a whole. As indicated in Section 2, Liu and Hai (2005) have adopted for this purpose a DEA approach due to Noguchi et al. (2002), which we present below.

Consider a branched node in the criteria hierarchy. As before, let K denote the number of its children, and L the number of places in the preference rankings submitted by the committee members with respect to this node in Step 3. Let V denote the *voting matrix*, whose elements $v_{k\ell}$ represent the number of times a (sub-)criterion k has been placed at rank position ℓ by the committee members ($k = 1, \ldots, L$, $\ell = 1, \ldots, L$). The *importance rating* of (sub-)criterion k with respect to its parent node can then be expressed in terms of the weighted average score

$$\theta_k \quad = \quad w_1 v_{k1} \; + \; \ldots \; + \; w_L v_{kL} \, ,$$

where the weights w_ℓ associated with rank positions $\ell = 1, \ldots, L$ can be understood as the 'worth of being ranked in the ℓ th place' (Cook et al. 1997), and need to obey the *strong order* $w_1 > \ldots > w_L$ so as to respect the ordinal nature of rankings submitted by the committee members in Step 3. However, any specific choice of such weights by the decision maker is likely to introduce a subjective bias into the importance rating of the (sub-)criteria, favouring certain of them and disfavouring the others, which potentially leads to biased decisions and undesirable outcomes. This issue can, nevertheless, be resolved by adopting the DEA approach (Cook and Kress 1990), which foregoes pre-specified rank weights. Specifically, this approach treats the peer (sub-)criteria $k = 1, \ldots, K$ as candidates in a preferential election who are allowed, in the DEA spirit, to suggest their own rank weights w_ℓ, $\ell = 1, \ldots, L$. In so doing, each candidate cross-evaluates himself against his peers in terms of importance rating θ_k, and selects such rank weights that yield the maximum possible rating score for him. Therefore, all rank weights are determined endogenously — by letting the preference data 'speak for itself', thus avoiding the subjective bias. To this end, Liu and Hai (2005) employed the following DEA model due to Noguchi et al. (2002):

$$\theta_{kk} \; := \; \max_{\{w_\ell\}} \sum_{\ell=1}^{L} w_\ell v_{k\ell} \tag{1}$$

subject to:

$$\sum_{\ell=1}^{L} w_\ell v_{i\ell} \leq 1 \qquad\qquad i = 1, \ldots, K \tag{2}$$

$$w_1 \; \geq \; 2w_2 \; \geq \; 3w_3 \; \geq \; \ldots \; \geq \; Lw_L \tag{3}$$

$$w_L \geq \varepsilon \tag{4}$$

where

$$\varepsilon = \frac{2}{nL(L+1)}. \tag{5}$$

In (1), the importance rating score of the k th (sub-)criterion is to be maximized. Decision variables w_ℓ ($\ell = 1, \ldots, L$), representing the rank weights proposed by the k th (sub-)criterion, need to satisfy a number of side constraints. Constraint (2) is a standard DEA constraint that places an upper bound on the importance rating to be achieved by any (sub-)criterion, including the k th itself, under the proposed rank weights; this upper bound is normalised to unity (Cook and Kress 1990; Green et al. 1996). Constraint (3) is intended to ensure that the rank weights decrease from first to last, yet in a convex fashion — so that the difference between the first and second places in the ranking is perceived stronger than between the second and the third ones, and so on. Constraint (4) ensures that the last rank weight is a positive number, for which Noguchi et al. (2002) suggested a minimum threshold as per (5).

ACCEPTED MANUSCRIPT

Problem (1)–(4) needs to be solved for each (sub-)criterion $k = 1, \ldots, K$, which respectively yields the *self-rating* θ_{kk} of the respective (sub-)criterion as the optimal objective value in (1). Furthermore, by evaluating the left-hand sides in (2) at the optimal solution, we obtain the *peer ratings* θ_{ik} for $i = 1, \ldots, K$, which represent the importance of the peer (sub-)criteria from the perspective of the k th one. When $i = k$, the peer rating coincides with the self-rating. The peer ratings across all $i = 1, \ldots, K$ and $k = 1, \ldots, K$ naturally comprise the *cross-evaluation matrix* Θ. By taking a geometric mean of the k th row in Θ, we obtain a *cross-rating* $\overline{\theta}_k$ of the k th (sub-)criterion, $k = 1, \ldots, K$ (Noguchi et al. 2002). By normalising the cross-ratings (Liu and Hai 2005), we obtain the *relative importance rating* of the k th (sub-)criterion with regard to its parent node in the criteria hierarchy as

$$\theta_k \quad := \quad \frac{\overline{\theta}_k}{\overline{\theta}_1 \ + \ \ldots \ + \ \overline{\theta}_L}. \tag{6}$$

After repeating this procedure for each branched node in the criteria hierarchy, we can finally express the *absolute importance rating* of each criterion and sub-criterion with respect to the superior goal of selecting the most suitable supplier. This is accomplished in the AHP fashion as follows: the absolute importance rating of a criterion is by definition equal to its relative importance rating, whereas the absolute importance rating of a sub-criterion is calculated by multiplying its relative importance rating by the absolute importance rating of its parent.

Step 5: Measure supplier performance. In the fifth step, the performance of the candidate suppliers is measured against those criteria and sub-criteria that are represented by the leaf nodes in the criteria hierarchy. Liu and Hai (2005) followed in this regard Yahya and Kingsman (1999) and measured the suppliers' performance against each of those leaf (sub-)criteria on a point scale from 0 to 10. Yahya and Kingsman (1999) provide a detailed account of how these measurements have been conducted in their case-study, which involved both factual data and qualitative judgements.

Step 6: Identify supplier priority. In the final step, an *overall rating score* is derived for each candidate supplier by calculating its weighted average performance across all leaf (sub-) criteria, with the weights being equal to the absolute importance rating scores of the latter. A supplier's overall rating score ultimately determines its priority for the company.

4. Critical Discussion and Revision of the VAHP Method

The VAHP method turns out to be instrumental in reducing the workload required from the committee members for submitting their preferences, as compared to the AHP. This represents an attractive feature, especially given a large number of supplier selection criteria and/or a large number of committee members. Furthermore it has the advantage of using DEA for deriving the relative importance of individual criteria from the group perspective, which is instrumental in avoiding subjective bias.

Despite these advantages, criticism can be levelled at the DEA approach used in Step 4 of the VAHP method. Below, we discuss the points of criticism in detail and we offer a revised DEA approach for use in the VAHP method that overcomes this criticism.

4.1 Critical Discussion

4.1.1 Restrictiveness of the Convex Weight Order. As indicated above, constraint (3) of the DEA model is intended to ensure a convex decreasing succession of weights w_ℓ associated with rank positions $\ell = 1, \ldots, L$. Specifically, Noguchi et al. (2002) require convexity to hold in the strict sense, i.e.:

$$w_1 - w_2 \ > \ w_2 - w_3 \ > \ \ldots \ > \ w_{\ell-1} - w_\ell \ > \ w_\ell - w_{\ell+1} \ > \ \ldots \ > \ w_{L-1} - w_L \ > \ 0 . \qquad (7)$$

While using convex orders has indeed been advocated in the literature on ranked voting systems (Stein et al. 1994; Hashimoto 1997), the above strict convexity excludes the popular Borda rule — which is questionable, as pointed out by Llamazares and Peña (2009).

4.1.2 Arbitrariness of the Convex Weight Order. Furthermore, Llamazares and Peña (2009) point to a certain arbitrariness of the strict convexity imposed by constraint (3). Indeed, (3) is derived by Noguchi et al. (2002) from condition (7) as follows. First, note that the inequality

$$w_\ell - \frac{\ell - 2}{\ell - 1} \cdot w_{\ell+1} \ > \ w_\ell - w_{\ell+1} \qquad (8)$$

holds true for any $w_\ell, w_{\ell+1} > 0$ and $\ell = 2, \ldots, L-1$. Thus, given that, the following inequality:

$$w_{\ell-1} - w_\ell \;\geq\; w_\ell - w_{\ell+1} \cdot \frac{\ell-2}{\ell-1} \tag{9}$$

implies the corresponding strict inequality $w_{\ell-1} - w_\ell > w_\ell - w_{\ell+1}$ in (7), where $\ell = 2, \ldots, L-1$. This and the last inequality in (7) implies: $w_\ell > w_{\ell+1}$ for $\ell = 1, \ldots, L-1$, so that (9) leads to:

$$w_{\ell-1} \;\geq\; 2w_\ell - w_\ell \cdot \frac{\ell-2}{\ell-1}, \tag{10}$$

where $\ell = 2, \ldots, L-1$. Rearranging the terms in (10) gives

$$(\ell-1)w_{\ell-1} \geq \ell w_\ell, \qquad \text{for } \ell = 2, \ldots, L-1. \tag{11}$$

From this, Noguchi et al. (2002) obtain constraint (3). However, this approach raises the following two concerns:

1) First, as Llamazares and Peña (2009) note, the choice of the ratio $(\ell-2)/(\ell-1)$ in (8) is rather arbitrary: instead, this ratio could have been replaced with any $\alpha_{\ell-1} \in [0,1)$, which would have resulted in a different set of inequalities in (3) as well.

2) Second, we note that (11) applies to $\ell = 2, \ldots, L-1$ only, which thus defines the following inequalities in (3): $w_1 \geq 2w_2 \geq 3w_3 \geq \ldots \geq (L-1)w_{L-1}$. However, the subsequent inequality in (3): $(L-1)w_{L-1} \geq Lw_L$ — does not follow from (11). This inequality is seemingly intended to imply the last inequality in (7): $w_{L-1} - w_L > 0$ — which it does, but, again, in a rather arbitrary way.

4.1.3 Possible Non-convexity of the Sequence of Weights. Surprisingly, constraint (3) does not actually guarantee convexity of the sequence of weights $w_1, \ldots, w_L$. To the best of our knowledge, this has not been noticed in the literature before. In fact, the above derivation of constraint (3) has a logical flaw. To see this, observe that the purpose of constraint (3) as per Noguchi et al. (2002), is to imply condition (7). However, the way in which (3) has been derived does not yield this implication. Indeed, let (4) hold. A closer look at the derivation in

ACCEPTED MANUSCRIPT

Section 4.1.2 reveals the implications shown in Figure 2 — which indicates that constraint (3) does not imply (7). Specifically, the first $L-2$ inequalities in (7) do not follow from (3).

$$(7)^{1,\ldots,L-2} \longleftarrow \left\{ \begin{array}{lcr} (9) & & \\ (7)^{L-1} & \longleftarrow & (3)^{L-1} \end{array} \right\} \longrightarrow (10) \longleftrightarrow (11) \longleftrightarrow (3)^{1,\ldots,L-2}$$

Note: Superscripts restrict the respective condition to the specified inequalities, by indicating their running numbers.

Figure 2. Logical implications between conditions (3), (7), (9)–(11).

4.1.4 Possible Arbitrariness of Cross-Evaluations. Noguchi et al. (2002) do not specify an approach to follow when there exist multiple optimal solutions to problem (1)–(4) (see also Llamazares and Peña 2009). The choice of a specific solution does not affect the self-rating θ_{kk} of the respective (sub-)criterion k, whereas it may affect peer ratings θ_{ik} for $i \in \{1,\ldots,K\}\setminus\{k\}$. Thus, leaving this choice up to the solver or the analyst solving (1)–(4) potentially leads to arbitrariness of cross-evaluations and hence, of cross-ratings $\bar{\theta}_k$, where $k \in \{1,\ldots,K\}$. Two approaches are conceivable in such a situation (Green et al. 1996):

AX: *Aggressive cross-evaluation*, where each (sub-)criterion $k \in \{1,\ldots,K\}$, acting like a candidate in a preferential election, chooses those optimal solutions to (1)–(4) that respectively minimize the peer ratings θ_{ik} of its individual opponents $i \in \{1,\ldots,K\}\setminus\{k\}$. This requires changing the objective function of problem (1)–(4) to

$$\theta_{ik} := \min_{\{w_\ell\}} \sum\nolimits_{\ell=1}^{L} w_\ell v_{i\ell}\,, \tag{1a}$$

introducing an additional constraint

$$\sum\nolimits_{\ell=1}^{L} w_\ell v_{k\ell} = \theta_{kk}\,, \tag{2a}$$

and solving problem (1a)–(4) for each $i \in \{1,\ldots,K\}\setminus\{k\}$, which yields peer ratings θ_{ik}.

BX: *Benevolent cross-evaluation*, where each (sub-)criterion $k \in \{1,\ldots,K\}$ maximizes the peer ratings of its opponents $i \in \{1,\ldots,K\}\setminus\{k\}$. This boils down to altering the direction of optimization in (1a) to 'maximization' and proceeding further as in AX.

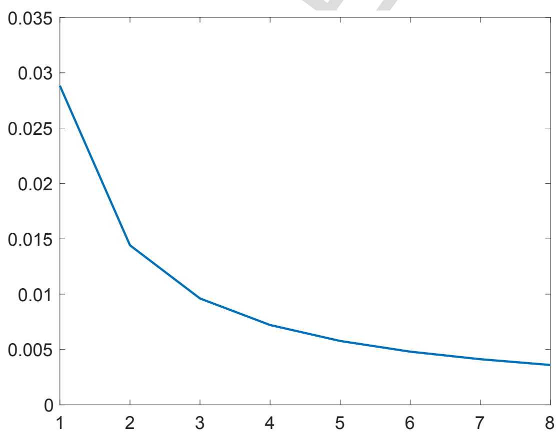
ACCEPTED MANUSCRIPT

Both aggressive and benevolent cross-evaluation can instead be conducted in an aggregate way, which replaces objective functions for individual opponents $i \in \{1, \ldots, K\} \setminus \{k\}$ in (1a) with their average:

$$\min_{\{w_\ell\}} \frac{1}{K-1} \sum_{i \neq k} \sum_{\ell=1}^{L} w_\ell v_{i\ell} \ . \tag{1aa}$$

This approach, originally due to Sexton et al. (1986), requires solving (1aa)–(4) just once and results in a set of weights that equally applies to all candidates. The respective approaches will be referred to as AAX and ABX.

Example 1. Consider the criteria hierarchy shown in Figure 1 and a voting matrix V that results from $n = 60$ committee members having voted on $K = 8$ criteria by ranking them on the ordinal scale of length $L = 8$ in Step 3 of the VAHP method, given in Liu and Hai (2005, Table 1). In Step 4 of the method, by solving problem (1)–(4) for $k = 3$, we obtain self-rating $\theta_{33} = 0.76155$ of the criterion 'Discipline'. However, multiple optimal solutions exist. If we let the solver select one (we use the default LP solver from Optimization Toolbox v. 7.6 of MATLAB R2017a) then the resulting solution vector $(w_1, \ldots, w_L)$ — comprising the weights attached by criterion k to ordinal scale gradations — is as depicted on the left in Figure 3. Using aggressive cross-evaluation AX against criterion $i = 1$, we obtain a different solution, depicted on the right in Figure 3. It obviously represents a non-convex sequence of weights. This example shows that cross-evaluations as defined in the VAHP method carry some arbitrariness, and that the resulting sequence of weights may turn out to be non-convex indeed.



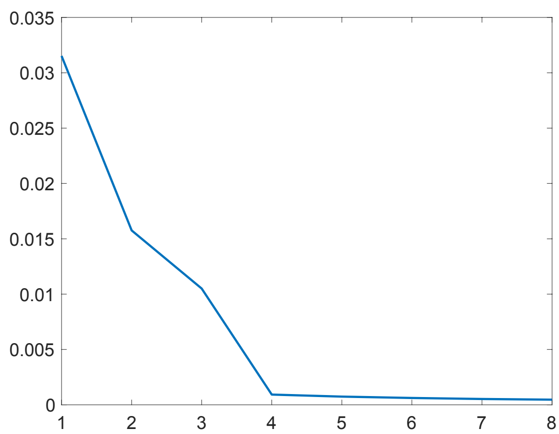


Figure 3. Rank weights in Example 1.

4.1.5 Arbitrariness of Rank Discrimination. Consider the threshold ε defined in (5) — that serves as the lower bound for the last rank weight w_L as per (4). By taking into account (3), we obtain that ε further determines a minimum difference between any two consecutive rank weights. Indeed, it is easy to derive from (3) that conditions

$$w_{\ell-1} - w_\ell \geq \frac{1}{\ell-1} w_\ell \qquad \text{and} \qquad w_\ell \geq \frac{L}{\ell} w_L$$

must hold for $\ell = 2, \ldots, L$. From this and (4) we thus have:

$$w_{\ell-1} - w_\ell \geq \frac{L}{(\ell-1)\ell} \cdot \varepsilon . \tag{12}$$

The right-hand side in (12), denoted by $d(\ell,\varepsilon) := g(\ell)\varepsilon$, represents a *discrimination intensity function* that specifies the minimum amount by which the consecutive rank weights must respectively differ (Cook and Kress 1990). The optimal solution vector is affected by (i) the specific choice of such a function, and (ii) the choice of the threshold ε. With regard to (i), several alternative functions have received attention in the literature — e.g., ε, ε/ℓ, and $\varepsilon/\ell!$ (Cook and Kress 1990, Green et al. 1996). Noguchi et al. (2002) neither derive the specific expression of $d(\ell,\varepsilon)$ as per (12) nor give arguments for why this function should be preferred over the others. Regarding (ii), the problem of choosing ε has long been acknowledged in the extant literature (Cook and Kress 1990, Green et al. 1996, Llamazares and Peña 2009, Park and Jeong 2011, Llamazares and Peña 2013). To address this issue, Noguchi et al. (2002) proposed to select ε as per (5). Their approach has, however, been criticized for its arbitrariness (Wang et al. 2007) — a criticism which we share.

4.2 A Revised Method

In the present work, we propose a revised VAHP method which is intended to remove the deficits of the DEA model employed in Step 4 of the original method (Section 4.1). Specifically, the proposed revision addresses the issues of (I) arbitrariness, restrictiveness and possible non-convexity of the order imposed on rank weights, and (II) arbitrariness of rank discrimination. These issues are accordingly addressed below.

4.2.1 Ensuring a Weakly Convex Weight Order. We address issue (I) by replacing constraint (3) with the following:

$$w_{\ell-1} - w_{\ell} \;\geq\; w_{\ell} - w_{\ell+1} \qquad \ell = 2, \ldots, L-1 \tag{13a}$$

$$w_{L-1} - w_{L} \;\geq\; \varepsilon \tag{13b}$$

In so doing, we follow the approach advocated by Stein et al. (1994) and Hashimoto (1997). This approach removes the deficits explained in Sections 4.1.1–3 because: 1) it permits the Borda rule by requiring a weakly convex sequence of weights in (13a); 2) it removes the arbitrariness of strict convexity imposed by the original method; and 3) it removes the possibility of obtaining a non-convex sequence of weights by enforcing convexity in (13a). Solving (1), (2), (13), (4) for each given $k \in \{1, \ldots, K\}$ accordingly yields self-ratings θ_{kk}.

In order to obtain peer ratings θ_{ik} for $i \in \{1, \ldots, K\} \setminus \{k\}$, we propose using the aggressive form of cross-evaluation AAX (Section 4.1.4) so as to remove arbitrariness of choice between multiple optimal solutions of problem (1), (2), (13), (4) and better discriminate between criteria (Green et al. 1996), while still applying the same set of weights equally to all candidates. This requires solving problem (1aa), (2), (2a), (13), (4) for each $k \in \{1, \ldots, K\}$. While Noguchi et al. (2002) employ the geometric mean of peer ratings θ_{ik}, $i = 1, \ldots, K$, to obtain cross-rating $\bar{\theta}_k$ of a kth (sub-)criterion (Section 3), we advocate using harmonic means for this purpose. The reason for this is that problem (1), (2), (13), (4) represents an input-oriented DEA model with L outputs and a single input, equal to unity (Lovell and Pastor 1999). It is well-known that the optimal objective value of an input-oriented model represents an input reduction factor for reaching the best-practice frontier by the decision-making unit and thus its *inverse* radial distance to that frontier (Balk et al. 2017; Cooper et al. 2011, Section 1.5). Therefore, averaging peer ratings in terms of their harmonic mean lends itself to a natural interpretation: averaging of radial distances rather than their reciprocals.

4.2.2 Selecting a Rank Discrimination Threshold. Issue (II) concerns arbitrariness of rank discrimination in the original VAHP method (Section 4.1.5). Note that weak convexity constraints (13a) naturally suggest using $d(\ell, \varepsilon) \equiv \varepsilon$ as the discrimination intensity function, which is reflected by constraint (13b). We are thus concerned solely with a proper selection of

ACCEPTED MANUSCRIPT

ε, called accordingly the *rank discrimination threshold*. As this threshold is likely to severely affect the results, literature presents varied attempts to recommend how it is chosen. Specifically, Cook and Kress (1990) proposed to use the maximum feasible value of ε. However, Green et al. (1996) found this approach infringes on the basic principle of DEA. They suggested using $\varepsilon = 0$ instead, which has been criticized by Noguchi et al. (2002) as contradicting the basic purpose of ranking — a criticism which we share. Further, the intermediate value (5) proposed in turn by Noguchi et al. is prone to arbitrariness (Section 4.1.5).

We have to conclude that whatever positive value be exogenously assigned to ε, it carries a subjective bias, favouring certain criteria $k \in \{1, \ldots, K\}$ while disfavouring the others, and would therefore contradict the basic principle of DEA. For this reason, we adopt the perspective that such a bias ought to be removed by letting the data, in the spirit of DEA, 'speak for itself' and determine rank discrimination endogenously. To this end, we employ a game-theoretic approach proposed in a different context by Tüselmann et al. (2015). This approach assumes criteria $k \in \{1, \ldots, K\}$ to act like candidates in a preferential election and jointly settle on the value of ε via bargaining (see Wu et al. 2009 for a related approach). We model this bargaining situation in terms of a K-person Nash bargaining problem, as follows (see also Pishchulov et al. 2014).

First, we determine the maximum amount $\varepsilon_{\max}$ of rank discrimination for which problem (1), (2), (13), (4) remains feasible (Mehrabian et al. 2000). This is accomplished by solving

$$\max \varepsilon \quad \text{s.t. } (2), (13), (4). \tag{14}$$

Next, we elicit utility functions $u_k(\varepsilon)$ that capture the utility extracted by candidate $k \in \{1, \ldots, K\}$ from a specific value $\varepsilon \in [0, \varepsilon_{\max}]$. Let $\bar{\theta}_k(\varepsilon)$ represent that candidate's cross-rating produced in Step 4 of the VAHP method (Section 3) using the DEA approach of Section 4.2.1 when given that value of ε. We accordingly define utility $u_k(\varepsilon)$ as the *standing* of the candidate k among its peers, calculated as follows:

$$u_k(\varepsilon) = \frac{\sum_{i=1}^{K} \left(\bar{\theta}_k(\varepsilon) - \bar{\theta}_i(\varepsilon) \right)}{\max_i \bar{\theta}_i(\varepsilon) - \min_i \bar{\theta}_i(\varepsilon)} \cdot \frac{1}{K}. \tag{15}$$

Thus, a candidate's standing indicates his position on the committee's preference scale relative to the 'average' candidate after normalization of the scale length (Pishchulov et al. 2014). Provided that the denominator in (15) is positive, it is easy to verify that the difference between the maximum and the minimum standing among the candidates is equal to unity, and that all standings total to zero. If the denominator in (15) happens to be zero then the standing of each candidate is defined to be zero as well.

Next, we determine the cross-ratings $\overline{\theta}_k(\varepsilon)$ for $k \in \{1, \ldots, K\}$ at each of N equispaced values of ε between 0 and $\varepsilon_{\max}$. For each k, this produces a series of N data points capturing the utility function $u_k(\varepsilon)$. We then interpolate this function on the range $[0, \varepsilon_{\max}]$ by fitting a polynomial $\hat{u}_k(\varepsilon)$ of degree m to N data points. In our experience, $N = 10$ is a sufficient number of points, for which $m = 4$ gives acceptable approximation $\hat{u}_k(\varepsilon)$ of $u_k(\varepsilon)$.[1]

Consider then a *K*-person bargaining problem with the *bargaining set U* and *disagreement point d:*

$$U = \left\{ \left(\hat{u}_1(\varepsilon), \ldots, \hat{u}_K(\varepsilon) \right) \,\middle|\, 0 \le \varepsilon \le \varepsilon_{\max} \right\}, \qquad d = \left(\min_{0 \le \varepsilon \le \varepsilon_{\max}} \hat{u}_1(\varepsilon), \quad \ldots, \quad \min_{0 \le \varepsilon \le \varepsilon_{\max}} \hat{u}_K(\varepsilon) \right).$$

The set *U* contains all *K*-dimensional utility vectors induced by the feasible values of ε, whereas *d* represents the vector of minimum utility levels of the respective candidates (Diskin and Felsenthal 2007). Note that *U* is by construction connected, closed and generally non-convex. Using its convex hull, as in classical Nash bargaining, would not be suitable because the given bargaining setup does not permit lotteries over *U* — that is, a randomized choice of ε — as an applicable bargaining outcome (Pishchulov et al. 2014). We therefore employ a generalization of Nash bargaining to non-convex problems (Zhou 1997) and determine the bargaining outcome by maximizing the *K*-person Nash product

$$\prod_{k=1}^{K} \left(\hat{u}_k(\varepsilon) - d_k \right) \tag{16}$$

[1] Alternatively, when polynomial interpolation results in a significant oscillation of the polynomial between data points, $\hat{u}_k(\varepsilon)$ can be interpolated with polynomial splines to ensure monotonic transition between the points.

on the interval $[0, \varepsilon_{\max}]$. Note that (16) is a polynomial of degree m^K whose global maximum can be determined exactly. The maximizing value of ε represents the bargaining outcome $\hat{\varepsilon}$, which is to be used as the rank discrimination threshold in the DEA approach of Section 4.2.1. The next section illustrates the application of the method.

5. Application to Sustainable Supplier Selection

As indicated in Section 1, the environmental and social impact of corporate supply chains has become a vital topic on companies' agendas which requires them to select suppliers in a way that respects all three sustainability dimensions: economic, environmental, and social. This significantly increases the number of evaluation criteria. Bearing in mind the multi-faceted nature of such an exercise, it should ideally involve a committee of experts so as to provide a balanced perspective on the importance of various criteria for the focal company. In this regard, the VAHP approach to supplier selection represents an attractive instrument requiring a parsimonious input from the committee members and thus letting them employ a comprehensive set of selection criteria. We have accordingly tested the feasibility of the revised VAHP method in a pilot application to sustainable supplier selection at a real-world company. The present section describes the procedure of this case-study and its results.

To conduct the study, we have partnered with a medium-sized company in the wood construction industry, based in Switzerland (the company name is not disclosed for confidentiality reasons). The company's product line comprises highly customised build-to-order and engineer-to-order dwelling and other building construction products involving lumber as its primary raw material. The company has seen a significant growth over the last few years and is currently expanding to international markets. Facing increasing sustainability concerns and its growing dependence on domestic as well as foreign suppliers, the company has indicated interest in conducting a pilot study on sustainable supplier selection so as to make its decisions in this area better informed and more systematic. It has commissioned for the study an expert committee comprising eight executive members who represent the company's key functional departments and managerial roles: director of production department, two technical directors, director of product engineering department, marketing director, HR director, assistant director, and the CEO.

The study was conducted on the company's premises in January 2018 and followed the six-step procedure by Liu and Hai (2005) as outlined in Section 3. In preparation for Step 1, the author team produced a catalogue of potential criteria for supplier evaluation, which was based on extant literature on supplier selection. Specifically our main source has been the recent study by Zimmer et al. (2016) who reviewed the literature on sustainable supplier selection and identified a comprehensive set of 448 selection criteria used in this stream of literature and spanning the economic, environmental and social dimensions of sustainability.[2] As Zimmer et al. (2016) intentionally exclude from consideration those studies that focus solely on economic criteria, we have additionally referred to the respective stream of literature to ensure that our criteria list is comprehensive. In so doing, we have employed the seminal criteria list by Dickson (1966), augmented with more recent criteria from the reviews by Cheraghi et al. (2004) and Ho et al. (2010) as well as additional criteria relating to technical capability (Yahya and Kingsman 1999), order volume (Kannan and Tan 2002), and quality assurance (Kuo and Lin 2012). As the criteria identified by Zimmer et al. (2016) exhibit a significant overlap both among themselves and with the latter stream of work, we matched criteria with each other in order to eliminate their duplication. This has allowed us to reduce the entire list down to 142 criteria, of which 58, 46 and 38 respectively represent economic, environmental and social dimensions of sustainability.

In order to ensure an easier perception and discussion of individual criteria by the committee members, the criteria have been structured in a hierarchical fashion following Yahya and Kingsman (1999), Govindan et al. (2013) and Zimmer et al. (2016), as shown schematically in Figure 4. The resulting hierarchy is four-level, with levels below the root node representing the *sustainability dimensions*, *criteria*, and *sub-criteria*, from top to bottom. Furthermore, to facilitate perception and discussion of criteria by the committee members, all elements of the criteria hierarchy have been translated from English to French by one of the authors as a native French speaker. Then, in Step 1 of the method, the committee members have accordingly been presented with the entire list of 142 sub-criteria and their grouping to criteria and dimensions. Each of the members has been asked to indicate sub-criteria relevant for supplier selection at the company from his or her individual perspective and add missing criteria, if any. This has resulted in elimination of 14 irrelevant sub-criteria (3 economic, 8 environmental, and 3 social). Five new sub-criteria were added: *Live tracking*, *Reliability*, *Modifiable / adaptable place of delivery*, *Defective product disposal due to environmental standards*, and *Compliance with (social) commitments undertaken*. Exclusions and additions

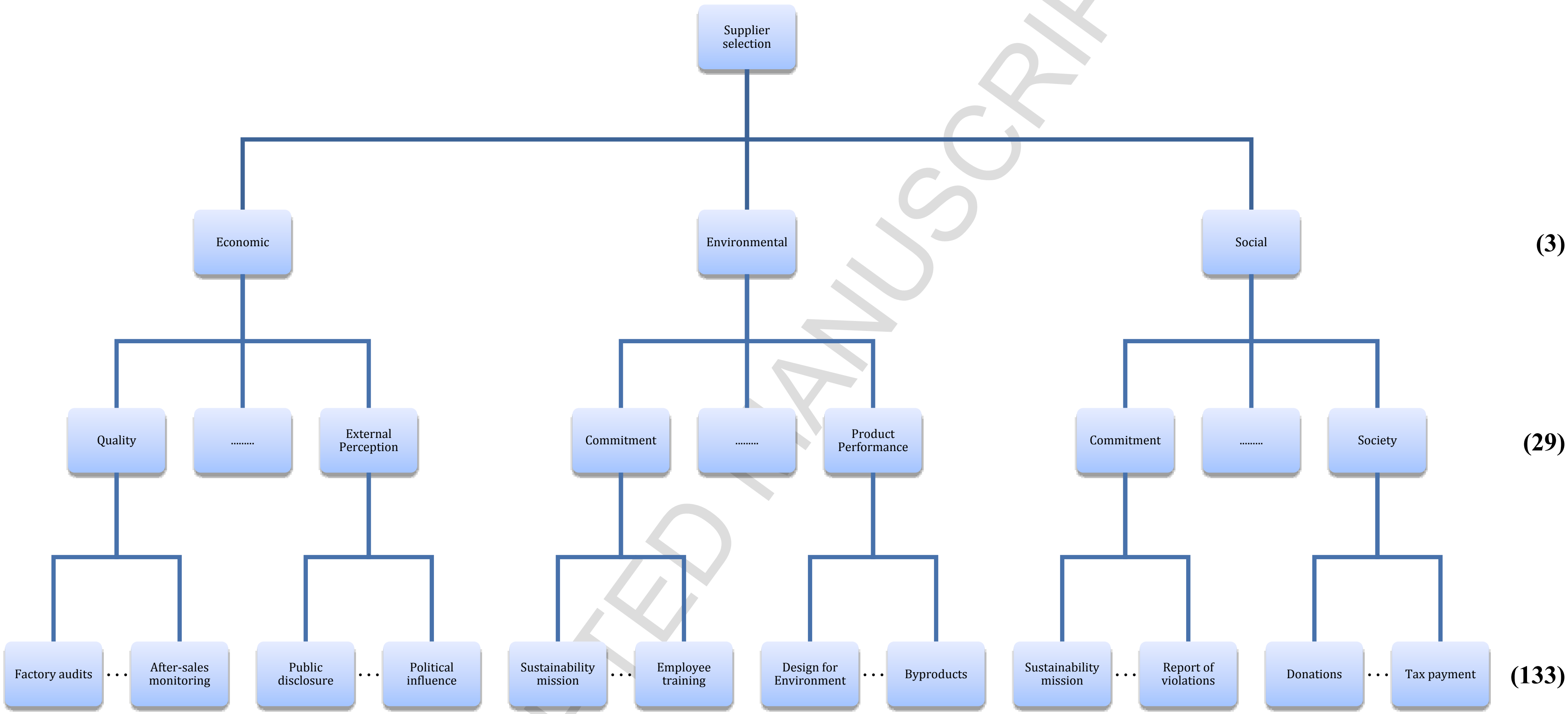
[2] We are thankful to Konrad Zimmer, Magnus Fröhling and Frank Schultmann of the Karlsruhe Institute of Technology, Germany, for providing us with the entire list of criteria identified in their work.

have been discussed within the committee. This has resulted in 133 sub-criteria on the list. In Step 2 of the method, it was agreed to follow the initial hierarchical grouping of (sub-)criteria to criteria and dimensions, as no re-grouping has been suggested. Figure 4 indicates the actual number of elements at each level of the resulting criteria hierarchy. The full list of all sub-criteria with their grouping to criteria and dimensions is presented in Appendix A.

In Step 3 of the method, the committee members have submitted their ordinal preferences with regard to the importance of elements at each level of the hierarchy relative to their respective parent node. This has been achieved by asking each committee member to rank order the sustainability dimensions with respect to the overall supplier selection goal; to rank order criteria with respect to their parent dimension; and to rank order sub-criteria with respect to their parent criteria. In this way, each committee member has submitted 1 + 3 + 29 = 33 ordinal rankings, which corresponds to the total number of nodes in the first, second and third levels of the criteria hierarchy (Figure 4). In each single ranking, the committee members have been asked to rank order all available elements; therefore, it holds that $L = K$ in the notation of Section 3. We note that the committee members have been able to submit their preferences anonymously and independently of each other, so as to exclude influence of the executives' opinions on one another and allow them to express an unbiased view.

Step 4 of the method is intended to aggregate the individual preferences of the committee members to the group preferences. This has been accomplished by the author team using the revised method presented in Section 4.2. Specifically, this method must be successively applied at the root node, at each dimension node, and each criterion node with two or more children in the criteria hierarchy. This has resulted in 31 executions of the method. As explained in Section 4.2, each execution requires:

(i) solving problem (14) for determining the feasible range $[0, \varepsilon_{\max}]$ of the rank discrimination threshold ε;
(ii) repeated solving of problems (1), (2), (13), (4) and (1aa), (2), (2a), (13), (4) for each $k \in \{1, \ldots, K\}$ at each of N points within the range $[0, \varepsilon_{\max}]$ for determining the bargaining outcome $\hat{\varepsilon}$;
(iii) solving problems (1), (2), (13), (4) and (1aa), (2), (2a), (13), (4) for each $k \in \{1, \ldots, K\}$ given $\varepsilon = \hat{\varepsilon}$.



Note: The levels in the hierarchy below the root node represent dimensions, criteria, and sub-criteria, from top to bottom. The figures on the right indicate the actual number of elements at the respective level in the hierarchy. See Appendix A for their complete list.

Figure 4. Schematic representation of the criteria hierarchy produced in Steps 1 and 2 of the case-study.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

This accordingly requires solving $1+2\times K\times N+2\times K=2K(N+1)+1$ linear programs during each execution of the method. In (ii), we set $N=10$ and use quartic polynomials for fitting a utility function $\hat{u}_k(\varepsilon)$ to N data points represented by the standings $u_k(\varepsilon)$, for each $k\in\{1,\ldots,K\}$ (Section 4.2.2). This is illustrated for two economic criteria in Figure 5. The figure shows that the standing of a criterion can siginificantly change with ε and may vary in a non-monotonic way. The mean absolute error (MAE) of curve fitting has the maximum value of approx. 0.008 across all 163 criteria, sub-criteria and dimensions acting as players in the bargaining game with their peers, while average MAE amounts to approx. 0.001. Taking into account that the players' standings are located on the scale of a unit length (Section 4.2.2), this represents an acceptable accuracy.

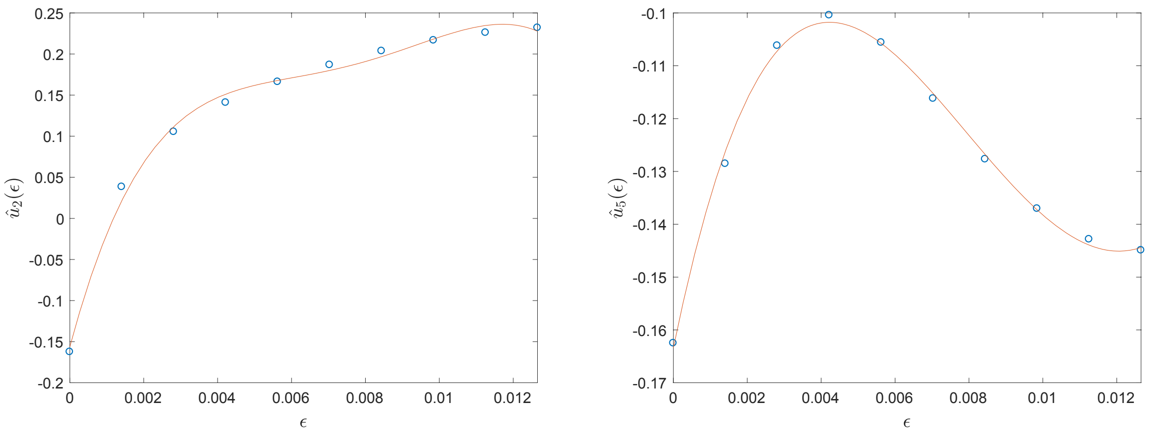


Figure 5. Fitting utility function $\hat{u}_k(\varepsilon)$ for criteria 'Delivery' (left) and 'Discipline' (right).

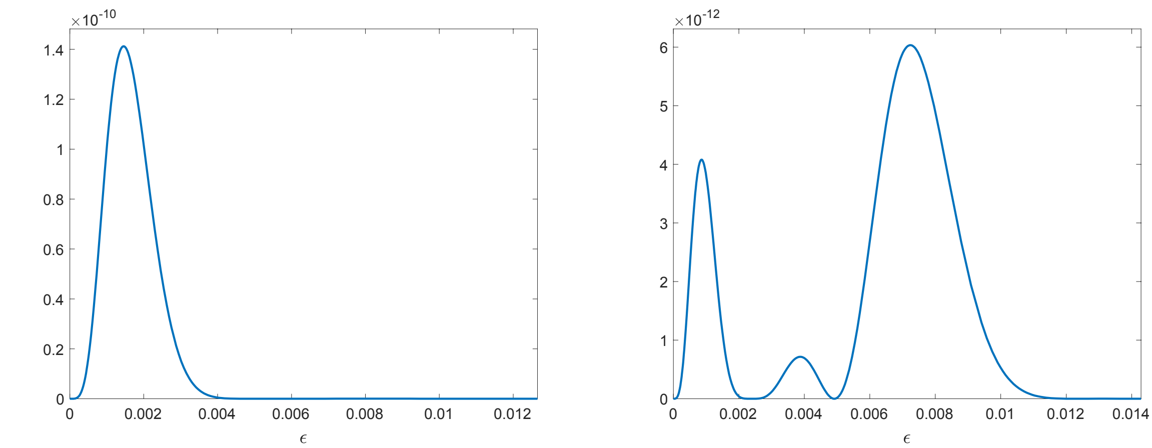


Figure 6. Nash product in barganing between economic criteria (left), and social ones (right).

Figure 6 (left) illustrates determination of the bargaining outcome $\hat{\varepsilon}$ between the economic criteria, as a maximizer of the Nash product (16). The panel on the right illustrates the same for the social criteria and reveals that the Nash product may possess multiple local minima on the feasible range $[0, \varepsilon_{\max}]$. The entire computational procedure has been implemented in MATLAB R2017a and required approx. 145 seconds to run on a MacBook Pro computer with a 2 GHz Intel Core i5 CPU and 8 GB of RAM. Below we present the main results.

Table 1 refers to the root node of the criteria hierarchy (Figure 4) and presents the corresponding voting matrix V (see Section 3, Step 4) along with the resulting importance ratings of the three sustainability dimensions. As we can see in the table, the economic dimension is by far the most important dimension from the group perspective. At the same time, the social dimension markedly surpasses the environmental one in importance. The latter observation is rather counterintuitive if one takes into account that the majority of the sustainable supply chain management literature predominantly refers to the environmental dimension and largely neglects the social one.

Rank Dimensions	1st	2nd	3rd	Rating
Economic	7	1	0	**0.51614**
Environmental	0	4	4	**0.19661**
Social	1	5	2	**0.28725**

Table 1. Voting matrix and the resulting importance ratings of sustainability dimensions.

We proceed to the next level in the criteria hierarchy and present in Table 2 the relative importance ratings of criteria in each of the three sustainability dimensions (the respective voting matrices are omitted for reasons of space). We can observe that the number of criteria in each dimension is quite similar and amounts to 10, 9, and 10 in the economic, environmental and social dimensions respectively. Comparison of the rating scores within each dimension shows their significant variation among the economic criteria: the ratio of the maximum to the minimum rating (*max/min ratio*) amounts to approx. 7.9 — meaning that the most important economic criterion is rated by the committee almost 8 times higher than the least important one, whereas the same indicator amounts to ca. 3.5 and 1.8 among environmental and social criteria, respectively. This suggests that the committee discriminates

ACCEPTED MANUSCRIPT

relatively strongly among economic criteria, less so among environmental ones, and least so among the social ones. Notably, the most important economic criterion has been found to be *Quality*, followed by *Financial*, *Responsiveness*, and *Delivery*, which aligns well with findings by Dickson (1966).

Economic criteria	Rating
Quality	0.224
Delivery	0.108
Responsiveness	0.111
Technical capability	0.087
Discipline	0.075
Management	0.058
Financial	0.161
Production & logistics	0.087
Facility	0.060
External perception	0.028

Environmental criteria	Rating
Env. commitment	0.197
Env. management	0.091
Env. capabilities	0.087
Material consumption	0.102
Energy consumption	0.094
Emissions	0.073
Water usage	0.057
Waste	0.105
Env. product performance	0.194

Social criteria	Rating
Social commitment	0.137
Social management	0.094
Child and forced labour	0.101
Occupational health & safety	0.116
Wages and working hours	0.094
Training of employees	0.108
Employment relationship	0.098
Discrimination and diversity	0.088
Stakeholder involvement	0.075
Society	0.090

Table 2. Relative importance ratings of criteria in each sustainability dimension.

Proceeding to the next level down in the criteria hierarchy, we similarly obtain relative importance ratings of sub-criteria. By multiplying these with the relative importance ratings of their ancestor nodes (Section 3), we obtain the absolute importance ratings of all 133 sub-criteria, which are presented in Appendix A. This concludes Step 4 of the revised VAHP method. Steps 5 and 6 ought to be conducted by the company in exactly the same way as in the original method by evaluating the candidate suppliers' performance against each sub-criterion (Step 5) and deriving supplier priority as a weighted average score (Step 6).

We conclude this section with a performance comparison between the original and revised VAHP methods. We refer first to the rank discrimination threshold used by either method. In Step 4, the original method uses the threshold defined by (5). Figure 7 (left) shows its distribution across 31 executions of the method in this step, relative to the maximum possible threshold value. As one can see, all threshold values happen to be in the lower third of the

ACCEPTED MANUSCRIPT

feasible range. This is largely driven by formula (5), which sets the threshold in dependence from the number of ranking positions L and — in our study setting — automatically from the number of peers K. The right panel in Figure 7 illustrates this. The revised method lets the preference data determine the rank discrimination threshold using a bargaining approach, which results in a more flexible choice of rank discrimination. Figure 8 (left) illustrates this. Its right panel shows that while there is a degree of association between the threshold and the number of peers, it is pronounced significantly less than in the original method.

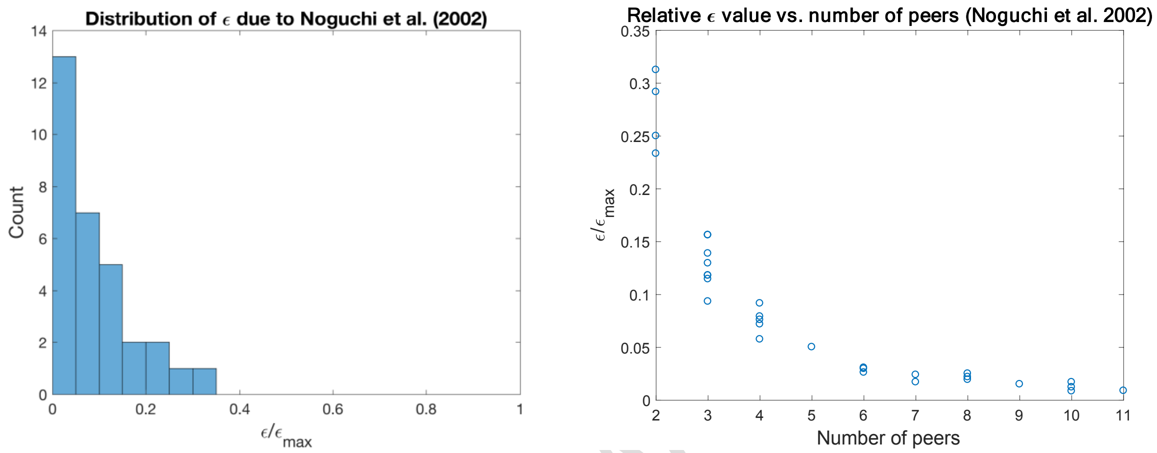


Figure 7. Distribution of the rank discrimination threshold in the original VAHP method, and its association with the number of peers across 31 executions of the method in Step 4.

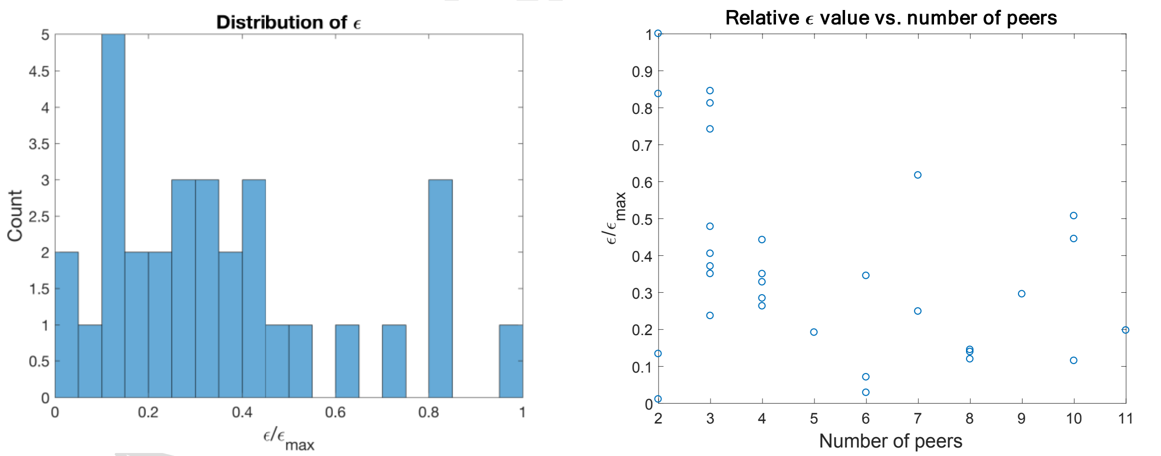


Figure 8. Distribution of the rank discrimination threshold in the revised VAHP method, and its association with the number of peers across 31 executions of the method in Step 4.

While the original method tends to choose a relatively low rank discrimination threshold, it compensates this by discriminating the peer criteria by means of the strong convex order (3). The numerical results confirm this: for example, the max/min ratio of the relative importance weights between the economic criteria amounts in the original method to ca. 95 when using geometric means for obtaining cross-ratings, as originally devised by Noguchi et al. (2002). The analysis in Section 4.1.5 reveals the likely reason for this: discrimination between high ranking positions, in comparison to low ones, becomes particularly strong when L is large, see equation (12). Thus, while (3) and (5) compensate for each other's effect, they are likely to represent too rigid a combination of model elements.

6. Conclusion

The VAHP method is useful in group decision-making on supplier selection, especially where there are a large number of selection criteria and/or a large number of group members. We proposed a revision of the method that overcomes criticism of its DEA-based preference aggregation procedure by (i) eliminating arbitrariness, restrictiveness and possible non-convexity of the order imposed on rank weights by the DEA model, and (ii) removing arbitrariness of rank discrimination by means of a bargaining approach, thus letting rank discrimination be data-driven in the DEA spirit.

We tested the feasibility of Steps 1 to 4 in the revised VAHP method using a pilot study of sustainable supplier selection at a real-world company. The comprehensive list of sustainable supplier selection criteria was based on work by Zimmer et al. (2016) and Dickson (1966). Our study contributes to business practice by providing a more robust VAHP tool that helps supply chain and purchasing managers select suppliers based on criteria spanning all three sustainability dimensions (economic, environmental, and social) while allowing parsimonious input by group decision-makers. Input required from managers is limited despite the fact that the criteria list is extensive and the VAHP approach allows for eliciting group members' individual preferences in a timely manner.

Expert review from the pilot company provided overall positive feedback regarding the practicability and usefulness of the VAHP tool. The managers claimed that the exercise involving a broad list of supplier evaluation criteria activated a learning process regarding the scope of criteria that may play a role in selecting suppliers. According to their evaluation, the VAHP tool helps to conduct supplier selection more systematically as well as to communicate the company's decisions to the suppliers with regard to supplier development, supplier switching, or termination of relationship. Furthermore, in-depth knowledge of a

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

comprehensive range of evaluation criteria prepares the company for negotiations with their own customers.

In our pilot application, social sustainability criteria were perceived overall as more important than environmental sustainability criteria. Further research is required to examine if this may be due to the specific case of the pilot company being situated in a context of high environmental expectations, standards, and regulation, therefore potentially alleviating the necessity of further scrutinizing supplier performance for environment sustainability. As companies are already bound by rather strict environmental government regulations in Switzerland, there might be less need to ensure such standards through supplier selection. Although Swiss law forces companies to comply with social sustainability criteria as well, in particular regarding labour rights and minimum standards, altogether company discretion on the social side could be considered higher than on the environmental side. Furthermore, attention to the social dimension at the case company could become increased during its recent period of growth, as giving due respect to social aspects has been considered by the management essential for sustaining the growth.

The revised method also lends itself to devise and include additional (sub-)criteria that are of special interest to the buying organisation. This enables a supplier selection process that more directly reflects the commercial and other interests of the buying company. This is even more so in a situation where a company operates or produces on behalf of a customer where this method would allow the provider or manufacturer to replicate its customers supplier selection criteria in its own procurement. However, in practice a large number of criteria used in supplier evaluation may require an unfeasible or disproportionate level of effort. In response to this, a possible approach could be culling those sub-criteria from the hierarchy that exhibit an importance weight below a certain threshold, either by applying a uniform threshold to all sub-criteria or using a threshold that is specific to each particular criterion. This would of course limit the benefits of the approach; therefore, the comprehensiveness of supplier assessment and its resource input need to be balanced, which represents an avenue for follow-up research. Future research can also be targeted towards empirically testing VAHP against other, established multi-criteria decision-making methods in the area of supplier selection.

ACCEPTED MANUSCRIPT

Acknowledgements

We are thankful to Konrad Zimmer, Magnus Fröhling and Frank Schultmann of the Karlsruhe Institute of Technology, Germany, for providing us with the list of supplier selection criteria identified in their research, and to Nikolay Kazantsev of the University of Manchester, UK, for his help with criteria matching. The first author gratefully acknowledges support by the European Union's Horizon 2020 research and innovation programme (grant agreement n° 723336).

References

Awasthi, A., Govindan, K., Gold, S., 2018. Multi-tier sustainable global supplier selection using a fuzzy AHP-VIKOR based approach. *International Journal of Production Economics*, 195, 106–17.

Bai, C., Sarkis, J., 2010. Integrating sustainability into supplier selection with grey system and rough set methodologies. *International Journal of Production Economics*, 124 (1), 252–64.

Balk, B.M., de Koster, M.B.M., Kaps, C., Zofío, J., 2017. An Evaluation of Cross-Efficiency Methods, Applied to Measuring Warehouse Performance. ERIM Report Series Research in Management ERS-2017-015-LIS, Erasmus University Rotterdam.

Charnes, A., Cooper, W.W., Rhodes, E., 1978. Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2 (6), 429–44.

Cheraghi, S.H., Dadashzadeh, M., Subramanian, M., 2004. Critical Success Factors for Supplier Selection: An Update. *Journal of Applied Business Research*, 20 (2), 91–107.

Cook, W.D., Doyle, J., Green, R., Kress, M., 1997. Multiple criteria modelling and ordinal data: Evaluation in terms of subsets of criteria. *European Journal of Operational Research*, 98 (3), 602–9.

Cook, W.D., Kress, M., 1990. A Data Envelopment Model for Aggregating Preference Rankings. *Management Science*, 36 (11), 1302–10.

Cooper, W.W., Seiford, L.M., Zhu J., 2011. Data Envelopment Analysis: History, Models, and Interpretations. In: Cooper W., Seiford L., Zhu J. (Eds.) *Handbook on Data Envelopment Analysis*. Boston, MA: Springer, 1–39.

Cousins, P., Lamming, R., Lawson, B., Squire, B., 2008. *Strategic Supply Management*. Harlow, UK: Pearson.

Demuijnck, G., Fasterling, B., 2016. The Social License to Operate. *Journal of Business Ethics*, 136 (4), 675–85.

De Boer, L., Labro, E., Morlacchi, P., 2001. A review of methods supporting supplier selection. *European Journal of Purchasing and Supply Management*, 7 (2), 75–89.

ACCEPTED MANUSCRIPT

Dickson, G.W., 1966. An Analysis Of Vendor Selection Systems And Decisions. *Journal of Purchasing*, 2, 5–17.

Dyllick, T., Hockerts, K., 2002. Beyond the business case for corporate sustainability. *Business Strategy and the Environment*, 11 (2), 130–41.

Diskin, A., Felsenthal, D.S., 2007. Individual rationality and bargaining. *Public Choice*, 133 (1–2), 25–9.

Genovese, A., Koh, S.C.L., Bruno, G., Esposito, E., 2013. Greener supplier selection: state of the art and some empirical evidence. *International Journal of Production Research*, 51 (10), 2868–86.

Goebel, P., Reuter, C., Pibernik, R., Sichtmann, C., 2012. The influence of ethical culture on supplier selection in the context of sustainable sourcing. *International Journal of Production Economics*, 140 (1), 7–17.

Gold, S., Heikkurinen, P., 2018. Transparency Fallacy: Unintended Consequences of Stakeholder Claims on Responsibility in Supply Chains. *Accounting, Auditing & Accountability Journal*, in press, doi: 10.1108/AAAJ-06-2015-2088.

Govindan, K., Khodaverdic, R., Jafarian, A., 2013. A fuzzy multi criteria approach for measuring sustainability performance of a supplier based on triple bottom line approach. *Journal of Cleaner Production*, 47, 345–354.

Green, R.H., Doyle, J.R., Cook, W.D., 1996. Preference voting and project ranking using DEA and cross-evaluation. *European Journal of Operational Research*, 90 (3), 461–72.

Hashimoto, A., 1997. A ranked voting system using a DEA/AR exclusion model: A note. *European Journal of Operational Research*, 97 (3), 600–4.

Hill, T., 1985. *Manufacturing Strategy*. London: Palgrave.

Ho, W., Xu, X., Dey, P.K., 2010. Multi-criteria decision making approaches for supplier evaluation and selection: A literature review. *European Journal of Operational Research*, 202 (1), 16–24.

Ishizaka, A., Siraj, S., 2018. Are multi-criteria decision-making tools useful? An experimental comparative study of three methods. *European Journal of Operational Research*, 264 (2), 462–71.

Kannan, D., Govindan, K., Rajendran, S., 2015. Fuzzy axiomatic design approach based green supplier selection: A case study from Singapore. *Journal of Cleaner Production*, 96, 194–208.

Kannan, V.R., Tan, K.C., 2002. Supplier Selection and Assessment: Their Impact on Business Performance. Journal of Supply Chain Management, 38, 11–21.

Karjalainen, K., Salmi, A., 2013. Continental differences in purchasing strategies and tools. *International Business Review*, 22 (1), 112–25.

ACCEPTED MANUSCRIPT

Kuo, R.J., Lin, Y.J., 2012. Supplier selection using analytic network process and data envelopment analysis. *International Journal of Production Research*, 50 (11), 2852–63.

Li, Z., Shao, L., 2015. Time-based procurement. *International Journal of Production Economics*, 169, 350–61.

Liu, F.-H.F., Hai, H.L., 2005. The voting analytic hierarchy process method for selecting supplier. *International Journal of Production Economics*, 97 (3), 308–17.

Llamazares, B., Peña, T., 2009. Preference aggregation and DEA: An analysis of the methods proposed to discriminate efficient candidates. *European Journal of Operational Research*, 197 (2), 714–21.

Llamazares, B., Peña, T., 2013. Aggregating preferences rankings with variable weights. *European Journal of Operational Research*, 230 (2), 348-55.

Lovell, C.A.K., Pastor, J.T., 1999. Radial DEA models without inputs or without outputs. *European Journal of Operational Research*, 118 (1), 46–51.

Lund-Thomsen, P., Lindgreen, A., 2014. Corporate social responsibility in global value chains: Where are we now and where are we going? *Journal of Business Ethics*, 123 (1), 11–22.

Mehrabian, S., Jahanshahloo, G.R., Alirezaee, M.R., Amin, G.R. 2000. An assurance interval for the non-Archimedean epsilon in DEA models. *Operations Research*, 48 (2), 344–7.

Noguchi, H., Ogawa, M., Ishii, H., 2002. The appropriate total ranking method using DEA for multiple categorized purposes. *Journal of Computational and Applied Mathematics*, 146 (1), 155–66.

Park, K.S., Jeong, I., 2011. How to treat strict preference information in multicriteria decision analysis. *Journal of the Operational Research Society*, 62 (10), 1771–83.

Pishchulov, G., Tüselmann, H., Sinkovics, R.R., 2014. Towards a consolidation of worldwide journal rankings — A classification using random forests and aggregate rating via data envelopment analysis. Working paper, http://ssrn.com/abstract=2492472.

Rajeev, A., Pati, R.K., Padhi, S.S., Govindan K., 2017. Evolution of sustainability in supply chain management: A literature review. *Journal of Cleaner Production*, 162, 299–314.

Rodríguez, A., Nieto, M.J., 2016. Does R&D offshoring lead to SME growth? Different governance modes and the mediating role of innovation. *Strategic Management Journal*, 37 (8), 1734–53.

Saaty, T.L., 1980. *The Analytic Hierarchy Process*. New York: McGraw-Hill.

Saaty, T.L., 2012. *Decision Making for Leaders*. Pittsburgh: RWS Publications, 3rd ed.

Sexton, T.R., Silkman, R.H., Hogan, A.J., 1986. Data envelopment analysis: critique and extensions. In: Silkman, R.H. (Ed.). *Measuring Efficiency: An Assessment of Data Envelopment Analysis*. San Francisco: Jossey-Bass, 73–105.

ACCEPTED MANUSCRIPT

Stein, W.E., Mizzi, P.J., Pfaffenberger, R.C., 1994. A stochastic dominance analysis of ranked voting systems with scoring. *European Journal of Operational Research*, 74 (1), 78–85.

Thiruchelvam, S., Tookey, J.E., 2011. Evolving trends of supplier selection criteria and methods. *International Journal of Automotive and Mechanical Engineering*, 4, 437–54.

Trautrims, A., MacCarthy, B.L., Okade, C., 2017. Building an Innovation-Based Supplier Portfolio: The Use of Patent Analysis in Strategic Supplier Selection in the Automotive Sector. *International Journal of Production Economics,* 194, 228–36.

Tüselmann, H., Sinkovics, R.R., Pishchulov, G., 2015. Towards a consolidation of worldwide journal rankings – A classification using random forests and aggregate rating via data envelopment analysis. *Omega*, 51, 11–23.

Wang, Y.-M., Chin, K.-S., Yang, J.B., 2007. Three new models for preference voting and aggregation. *Journal of the Operational Research Society*, 58 (10), 1389–93.

Watkins, T.A., 1999. Dual-Use Supplier Management and Strategic International Sourcing in Aircraft Manufacturing. In: Wessner, C.W. (Ed.). *Trends and Challenges in Aerospace Offsets*. Washington, D.C.: National Academy Press, 167–96.

WCED, 1987. *Our Common Future*. World Commission on Environment and Development, Oxford University Press, Oxford.

Weber, C.A., Current, J.R., Benton, W.C., 1991. Vendor selection criteria and methods. *European Journal of Operational Research*, 50 (1), 2–18.

Wetzstein, A., Hartmann, E., Benton, W.C., Hohenstein, N.O., 2016. A systematic assessment of supplier selection literature – State-of-the-art and future scope. *International Journal of Production Economics,* 182, 304–23.

Wolf, J., 2014. The relationship between sustainable supply chain management, stakeholder pressure and corporate sustainability performance. *Journal of Business Ethics*, 119 (3), 317–28.

Wu, J., Liang, L., Yang, F., Yan, H., 2009. Bargaining game model in the evaluation of decision making units. *Expert Systems with Applications*, 36 (3), 4357–62.

Yahya, S., Kingsman, B., 1999. Vendor rating for an entrepreneur development programme: a case study using the analytic hierarchy process method. *Journal of the Operational Research Society*, 50 (9), 916–30.

Zhou, L., 1997. The Nash bargaining theory with non-convex problems. *Econometrica*, 65 (3), 681–5.

Zimmer, K., Fröhling, M., Schultmann, F., 2016. Sustainable supplier management – a review of models supporting sustainable supplier selection, monitoring and development. *International Journal of Production Research*, 54 (5), 1412–42.

Appendix A. Grouping of sub-criteria to criteria and dimensions, and their absolute importance rating.

I. Economic criteria	Sub-criteria	Rating
Quality	Number of product rejects due to quality	0.0418
	Supplier's service, warranty and claim policies	0.0434
	After-sales product performance monitoring	0.0305
Delivery	Punctuality	0.0213
	Live tracking	0.0107
	Reliability	0.0127
	Modifiable / adaptable place of delivery	0.0111
Responsiveness	Lead time	0.0258
	With respect to urgent delivery requests	0.0202
	Time taken to solve quality problems	0.0114
Technical Capability	Range of products (type, variety)	0.0127
	Traceability of technical problems	0.0059
	Ability to solve technical problems	0.0164
	R&D and innovation capability	0.0097
Discipline	Honesty on transactions	0.0209
	Compliance to policies and guidelines	0.0179
Management	Attitude towards improvements, cooperation	0.0040
	Business skill (customer service, employees)	0.0038
	Partnership formation time	0.0022
	Strategic goals	0.0020
	Risk sharing capability	0.0038
	Openness in exchange of information	0.0026
	Access to markets, market analysis ability	0.0035
	Sub-tier supplier management	0.0019
	Operational controls	0.0020
	Labour relations	0.0026
	Organisational structure	0.0017
Financial	Costs or price	0.0234
	Quantity discounts	0.0111
	Flexible cash terms	0.0160
	Amount of past business (duration & quality) with the supplier	0.0092
	Share of order volume in the supplier's sales	0.0085
	Strength and stability of supplier's financial position	0.0069
	Supplier's growth	0.0047
	Timely, meaningful and reliable financial reporting	0.0035

ACCEPTED MANUSCRIPT

Production and Logistics	Capacity	0.0093
	Adaptability	0.0080
	Inventory management	0.0071
	Packaging	0.0022
	Transport	0.0087
	Just-in-time capability	0.0086
	Reverse logistics	0.0008
Facility	Infrastructure (accessibility, transportation modes)	0.0070
	Local availability of skilled labour	0.0038
	Local culture (e.g. corruption)	0.0027
	Local legal system	0.0048
	Local political stability	0.0019
	Information and communication technology	0.0051
	Machinery (level of automation, maintenance)	0.0032
	Layout	0.0025
External Perception	Public disclosure	0.0023
	Acquired certificates	0.0030
	Supplier's market share	0.0011
	Reputation within the branch of interest	0.0032
	Reputation outside of the branch (overall reputation)	0.0015
	Effect of contract with this supplier on other contracts	0.0015
	Reputation of sub-tier suppliers	0.0012
	Political influence	0.0009
II. Environmental criteria	**Sub-criteria**	**Rating**
Environmental Commitment	Adoption of a sustainability mission (goals, target)	0.0048
	Adoption of environmental code of conduct	0.0043
	Commitment to green production	0.0041
	Commitment to green logistics	0.0036
	Commitment to green supply chain management	0.0041
	Commitment to green marketing, eco-labelling	0.0034
	Commitment to green investments	0.0031
	Degree of R&D and investment into green technology	0.0033
	Cooperation w.r.t. environmental improvements	0.0044
	Employee training	0.0037
Environmental Management	Adoption of environmental programs and plans	0.0071
	Use of environmental management system and auditing	0.0065
	Sub-supplier evaluation and training	0.0043

ACCEPTED MANUSCRIPT

Environmental Capabilities	Environmental competencies	0.0046
	Sustainable product and process design capabilities	0.0054
	Sustainable land use	0.0008
	Environmental performance of upstream suppliers	0.0005
	Environmental certification	0.0054
	Defective product disposal due to environmental standards	0.0005
Material Consumption	Consumption of materials	0.0089
	Collection and remanufacturing of used products	0.0058
	Collection and recycling of used products	0.0054
Energy Consumption	Energy consumption or efficiency	0.0119
	Use of energy from renewable energy sources	0.0066
Emissions	Volume of carbon dioxide (CO_2) emissions	0.0091
	Volume of ozone depleting emissions	0.0053
Water Usage	Volume of waste water produced	0.0056
	Degree of water pollution	0.0056
Waste	Volume of solid waste	0.0048
	Volume of liquid waste	0.0050
	Volume of hazardous waste	0.0064
	Utilization or sale of waste or equipment	0.0044
Environmental Product Performance	Design for the environment (energy, pollution)	0.0073
	Remanufacturability	0.0055
	Recyclability	0.0066
	Biodegradability	0.0058
	Sustainable packaging	0.0047
	Content of hazardous materials	0.0044
	Undesirable byproducts	0.0037
III. Social criteria	**Sub-criteria**	**Rating**
Social Commitment	Adoption of a sustainability mission (goals, target)	0.0118
	Adoption of a social code of conduct	0.0072
	Management commitment and participation	0.0070
	Participation of employees (motivation and attitude)	0.0012
	Certification	0.0118
	Report of violations	0.0002
Social Management	Observing and living the guiding policies	0.0269
Child and Forced Labour	Use of child labour	0.0111
	Use of forced labour	0.0072
	Engagement in child trafficking	0.0108

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Occupational Health and Safety	Use of a health and safety management system	0.0112
	Frequency of work-related accidents or illness	0.0101
	Emergency preparedness/capability	0.0051
	Use of safety training and auditing	0.0037
	Health costs per employee	0.0032
Wages and Working Hours	Average working time	0.0059
	Offering a flexible working time	0.0031
	Use of restrictions on overtime	0.0041
	No work required on public holidays	0.0039
	Fair wage level	0.0065
	Overtime compensation	0.0033
Training of Employees	Average hours of training per year per employee	0.0312
Employment Relationship	Type of employment contract	0.0097
	Protection for pregnant workers	0.0059
	Career opportunities	0.0050
	Annual employee turnover	0.0076
Discrimination and Diversity	Diversity of workforce (age, gender, origin, minorities, disabilities, religion)	0.0074
	Gender balance of workforce	0.0039
	Equal treatment of workforce, tolerance	0.0082
	Number of incidents of harassment and violence recorded	0.0057
Stakeholder Involvement	Compliance with commitments undertaken	0.0111
	Effectiveness of supplier training in social issues	0.0043
	Fair sourcing by the supplier	0.0062
Society	Generation of employment	0.0097
	Consumers education	0.0078
	Payment of taxes	0.0083

Highlights

- A revised Voting Analytic Hierarchy Process is proposed.
- A novel procedure for aggregation of ordinal preferences.
- Featuring a data-driven bargaining approach.
- Application to sustainable supplier selection using a comprehensive set of criteria.

ACCEPTED MANUSCRIPT