CONF-IRM 2012 Proceedings

International Conference on Information Resources Management (CONF-IRM)

5-2012

Pairwise comparison techniques for preference elicitation: using test-retest reliability as a quality indicator

Martina Maida
University of Economics and Business, martina.maida@wu.ac.at

Konradin Maier
University of Economics and Business, konradin.maier@wu.ac.at

Nikolaus Obwegeser
University of Economics and Business, nikolaus.obwegeser@gmail.com

Recommended Citation

Maida, Martina; Maier, Konradin; and Obwegeser, Nikolaus, "Pairwise comparison techniques for preference elicitation: using test-retest reliability as a quality indicator" (2012). *CONF-IRM 2012 Proceedings*. 65.
https://aisel.aisnet.org/confirm2012/65

Pairwise comparison techniques for preference elicitation: using test-retest reliability as a quality indicator

Martina Maida
University of Economics and Business
martina.maida@wu.ac.at

Konradin Maier
University of Economics and Business
konradin.maier@wu.ac.at

Nikolaus Obwegeser
University of Economics and Business
nikolaus.obwegeser@wu.ac.at

Abstract

This study aims at augmenting the profile distance method (PDM) with techniques that support the elicitation of a relative weight vector. Therefore, prominent methods from the field of preference elicitation and related research are investigated according to their task fitness. Preference elicitation is widely regarded as one integral part of decision support. It has found broad attention in multiple scientific fields, such as psychology of choice or behavioural decision making. The technique of pairwise comparison, often utilized in applied decision support, is examined according to practicability and validity for the estimation of a relative weight vector. The concrete task of weight vector assessment is tested in an experimental setting using widely accepted scales and techniques derived from the literature survey conducted. We distinctly identify two key figures for the measurement of outcome quality, accounting for both mathematical consistency and internal (or human) consistency. Preliminary experimental results from a web-based study with sixty two (62) distinct users provide for valuable insights in consistency ratio and test-retest reliability, indicating that unmodified pairwise comparisons are a suboptimal method for criteria preference elicitation. Additionally, we propose a number of improvements to practical preference elicitation, such as the use of a guided, process-based weight elicitation process.

Keywords

Pairwise Comparison, Test-Retest Reliability, Consistency Engineering

1. Introduction and Motivation

The profile distance method (PDM) (Bernroider & Stix 2006) is a multi-criteria decision making method that can be applied to numerous decision problems in the practical field, such as information systems (IS) vendor selection. It is designed to combine the merits of two prominent approaches to decision making, the utility ranking approach and the data envelopment analysis (DEA). While the PDM has been found to deliver valuable insights into multi-criteria decision problems, one of its major drawbacks is its dependence on a solid weight vector that is used as an input for the decision process. Like most decision support methods, the PDM is highly dependent on the quality of input data in its process. In order to alleviate this problem, we propose the augmentation of the PDM with elicitation support techniques that assure a certain level of data quality.

Therefore, following a broad literature review in the area of (preference) elicitation techniques, an experiment is conducted in order to ensure validity and applicability of the technique chosen for integration with the PDM. Members of a student mailing list were asked to complete a series of pairwise comparisons to define their preferences among a pre-defined set of criteria. Each participant had to complete two different experimental settings (in randomized order), accounting for possible errors in the data. While the experimental setting accounted for a great number of possible (negative) influences, our findings nevertheless reveal numerous problematic issues.

Applying pairwise comparisons as a technique for preference elicitation is widely used and utilized in some of the most prominent decision making methods (e.g. AHP). Previous research has pointed out several drawbacks of this technique, such as high levels of decisional conflict and cognitive effort (Aloysius et al. 2006), and the lack of one best method to derive the preference vector from the resulting pairwise comparison matrix (Barzilai, 1997). Moreover, from a psychological perspective, the issue of lacking retest-reliability applies to pairwise comparisons, especially for the measurement of preferences (in contrast to physical objects) (Thurstone, 1927). Both of this problems have been subject to intensive scientific research that delivered numerous valuable findings.

Preliminary experimental results indicate that unmodified pairwise comparisons are an inadequate technique for preference elicitation, being likely to result in inconsistent judgements and lacking retest-reliability. Therefore, the underlying paper proposes a new approach to preference elicitation, accounting for a human-centered perspective on preference elicitation. A process model is proposed that attempts to alleviate two major issues when using pairwise comparisons for preference elicitation, the lack of retest-reliability and the mathematical consistency of judgements (originating in the high level of cognitive effort).

The remainder of this article is structured as follows. Firstly, we present the state-of-the-art of the underlying research areas. Secondly, we describe our experiment and its results. Finally, we propose a sequential approach to overcome the identified drawbacks.

2. State-of-the-art

2.1 The profile distance method

Bernroider and Stix (2006) have introduced the PDM as an integrated method for multi-criteria decision making. It has been implemented as a software program in Bernroider et al. (2010) and tested according to usability and heuristic validity in Bernroider et al. (2011). The PDM is a combined approach that integrates the classic DEA approach to objectively generate weights for each alternative with the inputs of the decision maker (DM) in terms of utility values. The DM is allowed to fade between both paradigms and explore the decision space in between. In addition to many other multi-criteria decision support methods, the PDM accounts for the need for deeper insights into the structure of the decision problem and does not simply recommend one optimal, or a number of efficient alternatives. The PDM literally allows the DM to understand the underlying decision problem and evaluate it from different perspectives. As stated before, the PDM requires the DM to input a relative weight vector that represents the DMs target structure. To support the construction of such a desired weight profile (DWP), the underlying research investigates into the area of preference elicitation, especially for the application of pairwise comparisons as utilized in the analytic hierarchy process (AHP) (Saaty, 1990).

2.2 Preference elicitation techniques

Preference elicitation has been an active area of research for the last decades. A great number of contributions to this area has been derived from various scientific fields, such as cognitive and behavioral psychology, mathematics or sociology. It can undoubtedly be stated that preference elicitation is at the very core of most decision support tasks, representing the fundamental cognitive task of decision analysis. While a vast number of researches added to the enhancements of this field, a unified theory of preference elicitation still seems out of reach. A collection of different theories contributing to this area is provided e.g. by Lichtenstein and Slovic (2006). Focusing especially on biases and errors in the elicitation process and possible debiasing techniques, we found that the idea of preference construction, that is a DM is not aware of preferences by nature but constructs a preference ad-hoc when being asked to do so, is fairly widespread and accepted (Lichtenstein & Slovic 2006). This leads to the conclusion that the process of preference elicitation is dependent on context and time, resulting in a lack of procedure invariance and retest reliability. Especially when considering decisions in corporate context with high impact one has to pursuit a consistent preference declaration, regardless of method or context framing.

2.3 Pairwise comparisons in decision making

Pairwise comparison techniques are generally recognized as a powerful tool when trying to support a DM or a group of DMs in the process of expressing preferences. On the one hand, this highly critical process is the first step in most decision scenarios which actually defines the decision problem, and can therefore be described as the single most important step in MCDM (Triantaphyllou, 2000). On the other hand, due to the creative (and thus error-prone) nature of this task, Triantaphyllou regards it as the art aspect of MCDM (versus the science aspect).

The area of pairwise comparisons has especially found broader attention in the practical field when T. Saaty introduced the Analytic Hierarchy Process (AHP). The AHP has found broad attention in both academics and practical decision making and triggered off what can be seen as a whole stream of AHP-related research. Moreover, numerous researches investigated on the foundations of the AHP, that is pairwise comparison techniques in decision making. This has led to deep insights into what now is regarded a rather mature field of research. Moreover, due to the widespread use of the AHP, the vast number of empirical data that could be collected confirms the techniques proposed in the AHP.

2.3.1 The multiplicative pairwise comparison matrix

In the process of constructing a preference vector among a set of criteria using pairwise comparisons, a nXn-square matrix A is constructed. A holds all information about the comparisons made and is subject to the rules of reciprocity and transivity, that is A is defined as positive, multiplicative, reciprocal square matrix. For i, j and k being alternatives of the matrix, transivity (1) and reciprocity (2) are given as follows:

$$a_{ij} = a_{ik} \cdot a_{kj} \tag{1}$$

$$a_{ij} = \frac{1}{a_{ji}} \tag{2}$$

Only if all comparisons strictly satisfy these rules, the matrix can be considered as fully consistent. While these rules clearly define the formal requirements to consistency, a fully consistent matrix (i) is often only very hard to achieve due to the rising complexity with the number of elements and (ii) does not necessarily reflect the preferences of the DM most accurately (Barzilai et al. 1987).

2.3.2 Deriving a vector from pairwise comparison matrices

The derivation of a preference vector from the established pairwise comparison matrix is regarded as one critical part of the decision process. This step (consolidating the pairwise comparisons to one preference vector) is necessary due to the initial decomposition of the decision problem in order to generate pairwise comparisons. Due to the fact that fully consistent matrices tend to be rather rare in the practical field, vector derivation methods have to account for this imprecision by somehow incorporating the inconsistencies into the derivation process. Many weight derivation methods account for this problem using one of the concepts of pertubation theory or distance minimization. Saaty proposes to apply the principal eigenvector (EV) as a method for weight derivation and allows for slight inconsistencies in matrices to be reflected by the weight vector (Saaty, 2003). Lootsma proposes a different approach, based on logarithmic regression, summarized in Lootsma (1993). Barzilai (1997) analyses different approaches (including the principal EV) according to criteria like rank reversals, scale inversion, uniqueness etc. and finds that calculating the geometric mean represents the only acceptable solution.

2.3.3 Measuring consistency

The original AHP introduces the consistency index as the degree of inconsistency of a pairwise comparison matrix, stating that for a completely consistent matrix, the largest principal eigenvalue equals the size of the matrix, such as $\lambda_{max} = n$. Following this definition, the consistency index (CI) is given as

$$CI = \frac{\lambda_{max} - n}{n - 1} \tag{3}$$

Furthermore, to calculate the consistency ratio (CR) for a specific matrix, Saaty proposed to compare the calculated CI with a random consistency index (RI), which is the average consistency index for random matrices of a given size. To account for human error in decision making, Saaty states that the CR (4) is acceptable as long as below 10%, otherwise the judgements ought to be revised.

$$CR = \frac{CI}{RI} \tag{4}$$

Although often criticized and replaced by other consistency measures proposed in research, the CR as suggested by Saaty is still very common due to the high diffusion of the AHP in both academics and practice. For instance, in an analysis of various consistency measures for pairwise comparison matrices, Barzilai found that the CI has a number of drawbacks, including the lack of comparability of inconsistent matrices and missing explanations for the randomization of the RI or the 10% CR threshold. Barzilai et al. (1987) and Barzilai (1998) propose the relative consistency and relative error as alternatives for consistency measurement.

2.3.4 Scales for quantification

In the AHP, a linear scale for pairwise comparisons is given within the interval $[9,\frac{1}{9}]$ with 1 depicting equality and 9 depicting extreme domination of one element. This scale has been proposed and tested empirically in various settings (it will be referred to as the *Saaty-scale* in the remainder of this article). As with other parts of the AHP, the issue of choosing the right scales has been subject to lively discussion and is a matter of ongoing research. Previous research has indicated drawbacks of this particular scale and came up with variations (Triantaphyllou et al. 1994). For example, Ma and Zheng (1991) criticized the *Saaty-scale* for its uneven distribution of the values and presented an alternative, evenly distributed scale. Lootsma (1988) and Lootsma (1990) proposed the use of exponential scales.

3. Dimensions of quality

The issue of measuring quality can be interpreted in different ways for the task of preference elicitation using pairwise comparisons. We limit the definition of quality to the decisional quality, or output quality, which is distinctly different from the process quality. Output quality tries to measure two dimensions of the decision, one being the logical, mathematical part of the decision, the other being the psychological, cognitive part. Thus, we propose a twofold approach to define quality in the respective area, consisting of the following two perspectives.

Firstly, the mathematical interpretation of quality can be found in the consistency of the pairwise comparisons, forming the positive reciprocal pairwise comparison matrix. As discussed before, in order to be fully consistent, this matrix has to follow the rules of transivity (in the multiplicative case) and reciprocity. As investigated by Saaty and many others, pairwise comparisons tend to become more and more inconsistent with the amount of elements to be compared. In the AHP, Saaty defined two ways to cope with this problem: On the one hand, the consistency ratio (CR) was introduced as a measurement for the degree of how much a given matrix deviates from a consistent matrix. Saaty defines a threshold of 0.1 on the CR as acceptable for the derivation of a valid vector from the matrix. Additionally, a limit of 6 elements in one single hierarchy group of the AHP was introduced, thus limiting the inconsistencies introduced with a higher number of elements.

Secondly, quality can be interpreted as the degree to which the DM is consistent with the given preferences, regardless of context, time or other influences (although the DM's preference can clearly be subject to changes). We refer to this type of quality as inner consistency (versus mathematical consistency). In other words, the inner consistency can be defined as a stability-factor for the preferences chosen, indicating the degree of change that occurs in comparison of test and retest. We consider the classical test-retest reliability, that is the correlation between test and retest results as a valid measurement for the inner consistency. If known other influences to preference construction, such as variance over time or framing effects like context, are eliminated, the retest-reliability appears to give distinct information on the inner quality of the preferences chosen.

4. Empirical evidence

4.1 Methodology

An experiment was conducted to provide first empirical proof for the lack in procedure invariance and test-retest reliability that comes with pairwise comparisons. Testing the fitness of the elicitation technique, participants were given the task of supporting the generation of a sound relative weight vector for use in the further decision process of the PDM. To allow for a greater number of responses a web-based experiment was designed. The experiment was conducted using the setup described in the following.

4.2 Experimental setup

The experiment was designed on a web-based platform, that is, the only physical requirement for participation was a computer connected to the world-wide-web. The experiment web-site basically consisted of two separate areas:

The first (upper) text area was used to describe the experimental setting to the participant. It gave a general description of the task of weight vector elicitation, followed by an introduction to the concrete set-up of the tool. Users were introduced to the input-form which depicted an nXn-matrix, with each criteria on the horizontal and vertical header. Additionally, participants were introduced to the scale applied. The scale was derived from the original AHP, comprising the values from 1 (identical) to 9 (extreme domination). Since the combined pairwise comparisons form a positive reciprocal matrix, participants were only asked to fill out the upper half of the matrix, while the reciprocal values were calculated afterwards. To allow for reverse rankings, the scale was designed as -9 to -2 (column criterion dominates row criterion) and 2 to 9 (row criterion dominates column criterion). The value 1, depicting identical importance, was pre-defined for the all same-criterion comparisons (primary diagonal of the matrix) and could not be changed. This explanatory section of the web-site remained visible and unchanged through the whole experiment (with the participant being informed about this situation).

The second (lower) area of the web-site comprised the actual input-form for the participants. A short description of a concrete problem was given, followed by a depiction of a matrix with the request to fill out the empty fields. To support the user during the comparison-process, a tooltip text instantly textualized the comparison at the moment the user changed a field in the form of "[Criterion A] is 2-times more important than [Criterion B]". The user was asked to completely fill out the form (incomplete forms could not be submitted) and then press a button to advance to the next step.

The experiment consisted of two simulated cases, both representing classic selection processes that appear in a great number in our society. One case simulated the decision of choosing a new flat to rent with 6 criteria that were to be ranked. We will refer to this task as *Task1*. The other case was the selection of a new car, also including 6 criteria. We will refer to this task as *Task2*. This resulted in 15 pairwise comparisons for both cases (n*(n-1)/2). For both cases (referred to A and B), the order of *Task1* or *Task2* was randomized, that is which task is presented to the participant first. The order of the criteria in the input-form was also randomized to account for random errors. We traced every input given by the participants, that is, we were able to maintain a complete experiment log including timestamps and values for each action that was taken.

To test the retest-reliability, we interlinked the two cases to form one experimental process. In the retest, the participant was asked to raise the quality of the decision by comparing certain criteria again. The explanatory part of the web-page remained the same. The input-form was almost the same as for the initial questioning, except for the criteria order, which were randomized again. Additionally, only 7 of 15 fields were asked to be compared again to not cognitive frustrate the participants (the other fields were shown without values and blackened out).

In other words, after the randomization process, the participants where first presented with case A, then case B, then with the retest for case A followed by the retest for case B. By following this process we were able to ensure that the cognitive focus of the participants had switched to a different task before being asked again in the retest. This allows for interpretation of the comparison between retest and initial test as more than just memorization.

As a technical detail, we disabled the auto-complete functionality of html that is embedded in most web browsers, so the participants were not able to see the values for the comparison they had entered in the initial assessment.

4.3 Experimental results

After an initial screening and exclusion of erroneous data, the experiment was able to gain a set of 116 test-retest pairs, conducted by 62 distinct users. As with most web-based experiments, the number of participants starting the experiment process varied greatly from the number of actual finishers, a behavior which can also be attributed to the length of the experiment and the fact that no incentives were given to participants.

As described in Section 4.2, we designed the experiment with two tasks to be completed by each participant. The decision of which task is presented first was randomized, which led to the situation that 30 persons started with *Task1* and 32 persons started with *Task2*. The total average CR denoted 16 per cent, with the average CR for users that started with *Task1* was slightly lower (14 per cent) than for users that started with *Task2*. Looking at the situation from a different angle, the average CR over all *Task1* comparisons is 16 per cent, the same as the average CR over all *Task2*. Moreover, it has to be mentioned that of 116 comparisons, only 39 per cent (46/116) would pass the 10 per cent cut-off proposed by Saaty for the AHP. As Saaty proposed, in such a case the pairwise comparison process would have to start all over again until a valid CR (passing the 10 per cent cut-off) is established.

To determine the inner consistency, we measured the test-retest reliability. As stated in Section 4.2, we introduced methods to avoid memorization effects or to disable auto-complete functionalities in order to reduce bias. The result for test-retest reliability showed poor inner consistency. The responses given by the participants in the retest did only match the exact responses of the first test in few cases. Instead of raising the quality of the decisions given, their accuracy appeared to be questionable.

A further result of our experiment that accounts for poor consistency are the high levels of cognitive effort observed for performing pairwise comparison tasks. Data from the input log revealed that time effort decreased during the course of the experiment, which could be attributed to the fact that the participants learned how to use the technology (web-site) more efficiently. Nevertheless, decreasing time effort in combination with reduced consistency of the responses clearly points to frustration of the participants with the level of cognitive effort needed. Thus, when participants perceive a task as too complex or costly, their level of concentration fades out and their willingness to perform declines. This behavior obviously leads to inconsistencies in the pairwise comparison matrix.

5. Findings

Following from the literature review and the experiment conducted, we conclude that although research in the fields affected has led to substantial advancements, no theory of preference elicitation with pairwise comparisons could be established. This is mainly due to the enormous complexity of the field, representing the core of most decision problems with numerous different influencing variables.

Many research efforts have been conducted under the auspices of the AHP, especially in the area of pairwise comparisons and related fields such as scales for quantification, matrix vector derivation techniques or consistency calculation and optimization. Thus, a vast number of variations have been proposed with each having its own benefits and drawbacks, leading to a rather unstructured research field.

We selected the most common and prominent configuration for pairwise comparison preference elicitation as utilized in the AHP, using the *Saaty-scale* and also Saaty's CR for consistency measurement. Experimental results show that, although the respective techniques are widely spread in theoretical and practical usage, the quality of the preference elicitation process are rather suboptimal. Moreover, due to the high level of cognitive effort combined with this approach, the better part of participants did not finish the experiment accordingly. Of those who finished the experiment, only 39 per cent qualify for further processing after passing Saaty's 10% cut-off rule. Following our two-fold definition of quality, pairwise comparisons in its original form neither fulfill the mathematical requirements for consistency nor do they successfully account for context and time framing, as can be seen in the results of our empirical investigation. Especially from a practical point of view, this procedure cannot be considered as a best-practice for decision makers but has to be adopted or augmented to overcome its drawbacks.

6. Proposing a sequential process for step-by-step preference elicitation

The strive for consistent judgements is a recurring issue when dealing with pairwise comparison matrices. Research has produced different approaches to handle this problem, ranging from inconsistency-thresholds to take-the-closest-consistent-solution methods. Additionally, it is still subject to discussion whether full mathematical consistency is a goal to pursue or if a certain degree of inconsistency simply is necessary when dealing with human judgement (see Section 3 for a detailed discussion). To account for higher test-retest reliability in the elicitation of a weight vector, we propose to integrate the findings from both literature and the experiment conducted to form a new, sequential approach to preference elicitation. In other words, while we believe that the underlying research field has rapidly evolved in terms of formal methods, a lack in the advancements of practical tools for preference elicitation seems obvious. Tversky and Kahneman (1981) listed four major goals of a decision making process: maximize accuracy, minimize cognitive effort, minimize negative emotion, maximize ease of justifying. Derived from our experimental findings, we propose to add another four supporting goals: the possibility to change the data within the process, to raise awareness to potential errors and biases (e.g. framing, anchoring effects), to break down each decision problem to the smallest structure possible (and reasonable), to ensure comparability between decision outcomes.

At the core of the proposed sequential process, we propose an optimization-based algorithm that iteratively uses the DMs input to navigate within the space of consistent matrices. Other research has come up with alternative approaches, such as to replace the completed (inconsistent) matrix with the closest fully consistent matrix. We argue that, while this procedure clearly achieves the goal of having a consistent matrix for vector derivation, it effectively changes the DMs preferences, at least in some borderline cases.

Therefore, we propose to interactively navigate the DM towards a consistent matrix by providing feedback and suggesting directions after each iteration. Thereby, it can be assured that the DM is in full control of the preference elicitation process but still implicitly directed towards rational judgement. The algorithm should possess the following characteristics that lead to optimization-based iterative consistency. Firstly, the decision problem should be structured into small, manageable pieces that can be reproduced in the same manner. Structuring could be implemented by rough separation according to comparability, e.g. degree of importance. Secondly, ranking should be conducted among groups of comparable items using an assisted pairwise comparison process. This assistance could be e.g. in form of explanation facilities or special visualization techniques. In the last step, the structured pieces are consolidated and refined. Special focus is on items that were ranked as very important. The DM is asked to reconsider the ranking or to confirm it.

7. Conclusion

A solid preference elicitation technique is a prerequisite of a successful decision making method. Pairwise comparisons are widely used techniques that are integrated e.g. in the AHP or PDM. However, previous research indicates that pairwise comparisons also possess several drawbacks, like high levels of decisional conflicts or cognitive effort. This raises the question of how to achieve adequate decisional quality that is reflected by consistency and test-retest reliability.

Preliminary results of our experiment confirmed the problem of inconsistent judgements and lack of test-retest reliability due to increased cognitive effort. We therefore propose a modified technique of pairwise comparisons for preference elicitation that is sequential and consists of an optimization-based iterative algorithm. Thus, the problem is presented in a structured and refined way that is easier to comprehend by the DM. Furthermore, the DM should be assisted during preference elicitation process in order to reduce inconsistencies.

The findings of our experiment and the proposed modified technique will be the basis for further research, which includes a large empirical investigation to validate our previous findings. Additionally, we will advance with the mathematical development of the sequential pairwise comparison algorithm and its implementation for decision support systems.

Acknowledgement

This research has been funded by the Austrian Science Fund (FWF): project number TRP 111-G11.

References

Aloysius, J., F. Davis, D. Wilson, A. Ross Taylor, J. Kottemann (2006) "User acceptance of multi-criteria decision support systems: The impact of preference elicitation techniques", *European Journal of Operational Research*, (169)1, pp. 273-285.

Barzilai, J. (1997) "Deriving weights from pairwise comparison matrices", *Journal of the Operational Research Society*, (48)12, pp. 1226-1232.

Barzilai, J. (1998) "Consistency measures for pairwise comparison matrices", *Journal of Multi-Criteria Decision Analysis*, (7)3, pp. 123-132.

Barzilai, J., W. Cook, B. Golany (1987) "Consistent weights for judgements matrices of the relative importance of alternatives", *Operation Research Letters*, (6)3, pp. 131-134.

Bernroider, E., N. Obwegeser, V. Stix (2010) "Introducing complex decision models to the decision maker with computer software - the profile distance method", *Journal of Systemics, Cybernetics and Informatics*, (8)3, pp. 24-28.

Bernroider, E., N. Obwegeser, V. Stix (2011) "Analysis of heuristic validity, efficiency and applicability of the profile distance method for implementation in decision support systems", *Computers & Operations Research*, (38)5, pp. 816 - 823.

Bernroider, E. and V. Stix (2006) "Profile distance method: a multi-attribute decision making approach for information system investments", *Decision Support Systems*, (42)2, pp. 988-998.

Lichtenstein, S. and P. Slovic (2006) *The construction of preference*, Cambridge: UK, Cambridge University Press.

Lootsma, F. (1988) "Numerical Scaling of Human Judgement in Pairwise Comparison Methods for Fuzzy Multi-Criteria Decision Analysis", In: Mitra, G. (ed.) *Mathematical Models for Decision Support*, Berlin: Springer, pp. 57-88.

Lootsma, F. (1990) "The French and the American school in multi-criteria decision analysis" *Recherche opérationnelle*, (24)3, pp. 263-285.

Lootsma, F. (1993) "Scale sensitivity in the multiplicative AHP and SMART", *Journal of Multi-Criteria Decision Analysis*, (2)2, pp. 87-110.

Ma, D. and X. Zheng (1991) "9/9-9/1 scale method of AHP", In: *Proceedings of the 2nd international symposium on the AHP*, Pittsburgh, USA, pp. 197-202.

Saaty, T. (1990) "How to make a decision: the analytic hierarchy process", European Journal of Operational Research, (48)1, pp. 9-26.

Saaty, T. (2003) "Decision-making with the AHP: Why is the principal eigenvector necessary", *European Journal of Operational Research*, (145)1, pp. 85-91.

Thurstone, L. (1927) "A law of comparative judgment", *Psychological Review*, (34)4, pp. 273-286.

Triantaphyllou, E. (2000) *Multi-criteria decision making methods: a comparative study,* Netherlands: Kluwer Academic Publishers.

Triantaphyllou, E., F. Lootsma, P. Pardalos, S. Mann (1994) "On the evaluation and application of different scales for quantifying pairwise comparisons in fuzzy sets", *Journal of Multi-Criteria Decision Analysis*, (3)3, pp. 133-155.

Tversky, A. and D. Kahneman (1981) "The framing of decisions and the psychology of choice", *Science*, (211)4481, pp. 453-458.