SCALE SENSITIVITY AND RANK PRESERVATION IN A MULTIPLICATIVE VARIANT OF THE ANALYTIC HIERARCHY PROCESS (SUMMARY)

F.A. Lootsma

Faculty of Technical Mathematics and Informatics
Delft University of Technology
Mekelweg 4, 2628 CD Delft, The Netherlands

Abstract

In this paper we present a new version of Saaty's Analytic Hierarchy Process with several innovations. Because the fundamental scale introduces inconsistencies which are not present in the mind of the decision maker, we replace it by a number of geometric scales which are based on psychophysical arguments. These scales are only used at the first evaluation level, when we compare the alternatives under each of the criteria separately. The impact scores are estimated by logarithmic regression because it frequently happens that the pairwise comparisons are incomplete (missing data). The rank order of the impact scores appears to be invariant when geometric scales are employed. Next, the delicate aggregation step is improved by the geometric-mean aggregation rule and by a new definition of the relative importance of the criteria. At the second evaluation level there is only one geometric scale to quantify the relative importance of the criteria. The rank order of the final scores of the alternatives does therefore not depend on the geometric scale employed to calculate the impact scores at the first level.

1 Introduction

Saaty's Analytic Hierarchy Process (AHP), a pairwise-comparison method in the field of multi-criteria decision analysis, has primarily been designed to evaluate a finite number of decision alternatives $A_1,\ldots,A_n$ under a finite number of conflicting performance criteria $C_1,\ldots,C_m$, by a single decision maker or by a decision making body. In the basic experiment, where two stimuli are presented (two alternatives under a particular criterion), the decision maker is requested to express his indifference between the two, or his weak, strict, strong, or very strong preference for one of them. The AHP is clearly based upon the elicitation of preference intensities. Such a method can reasonably be used if the wider context of the decision problem is stable and if it is clearly described.

The work of Saaty on pairwise-comparison methods has been an incentive for many theoretical studies and practical applications. Several obsolete details must be replaced, however, such as the fundamental scale to quantify human judgement, the Perron-Frobenius eigenvector to estimate the impact scores of the alternatives, and the arithmetic-mean aggregation rule to calculate the final scores of the alternatives. The controversial issues, to be treated in the present paper, are not new. Only a few years ago, Zahedi (1986) signalized that the criticism on the AHP concentrated on the estimation of the impact scores, but that no major controversy existed concerning the aggregation step. Criticism on Saaty's fundamental scale was not mentioned in Zahedi's survey paper, but Belton (1986) brought forward several arguments against the scale and the aggregation rule.

We propose a multiplicative version of the AHP which operates as follows. In the basic experiment at the first evaluation level, where two alternatives A_j and A_k are compared under the criterion C_i, we collect the preference information, and we convert the verbal statement of decision maker d (the selected gradation of his comparative judgement) into a numerical value $r_{jkd}^{(i)}$ on a geometric scale. Next, we use logarithmic regression to calculate the single-criterion scores $\overline{v}_i(A_j), j = 1, \ldots, n$, also referred to as the impact scores, approximating the subjective values of the alternatives under criterion C_i. The impact scores are not unique. They have a multiplicative degree of freedom, so that they can be normalized in such a way that

$$\sum_{j=1}^{n} \overline{v}_i(A_j) = 1, i = 1, \ldots, m. \tag{1}$$

The basic experiment at the second evaluation level, where two criteria are mutually compared, is somewhat more complicated. We suggest the decision maker to consider two real or imaginary alternatives with the property that his preference for one of them under the first criterion equals his preference for the other alternative under the second criterion. Next, we could ask him to state whether he is indifferent between the two alternatives under the two criteria simultaneously, or whether one of the two criteria gives a decisive (weak, strict, strong, or very strong) preference for one of the two alternatives. This is the idea underlying a somewhat simplified procedure to estimate the relative importance of the criteria. Thereafter, the judgemental statements are converted into numerical values on a particular geometric scale. Logarithmic regression yields normalized weights $\overline{w}(C_i), i = 1, \ldots, m$, for the respective criteria. Finally, there is an aggregation step generating the final, multi-criteria scores $\overline{f}(A_j)$ via the geometric-mean aggregation rule

$$\overline{f}(A_j) = \alpha \prod_{i=1}^{m} (\overline{v}_i(A_j))^{\overline{c}_i}, \tag{2}$$

where $\overline{c}_i$ simply denotes the weight $\overline{w}(C_i)$, and α the normalization factor to guarantee that the final scores sum up to unity. By these quantities, the alternatives are unambiguously ranked in a subjective order of preference when we operate with geometric scales.

2 Numerical scales for verbal comparative judgement

In the basic experiment at the first evaluation level, two stimuli S_j and S_k (two alternatives A_j and A_k under a particular criterion) are presented to the decision maker whereafter he is requested to express his indifference between the two, or his weak, strict, strong, or very strong preference for one of them. It is tacitly assumed that the stimuli have unknown subjective values V_j and V_k for him. The purpose of the basic experiments and the subsequent analysis is to approximate these values. The verbal comparative judgement, given by the decision maker and converted into a numerical value r_{jk}, is taken to be an estimate of the ratio V_j/V_k. Hence, since we only have ratio information, we may take the subjective values to be normalized, in the sense that they sum up to 1 or to 100%. Saaty has never made it clear, however, how the subjective values might be understood. Hence, the numerical values assigned to verbal statements such as weak, strict, strong, or very strong preference for S_j with respect to S_k, are highly controversial.

In an earlier paper (Lootsma (1990)), we addressed this question as follows. First, we provided a heuristic introduction to illustrate the transition from car prices to the subjective judgements whereby cars are referred to as "cheap", "somewhat more expensive", "more expensive", or "much more expensive". In fact, we subdivided a given price range into a number of price categories (intervals) which are felt to be of the same order of magnitude, and we used the corresponding echelons (levels) to establish ratios of price increments expressing what we mean by "somewhat more", "more", and "much more". Next, we demonstrated that human judgement leads in many unrelated areas to the same categorization of intervals: there are roughly four major categories, the echelons constitute a sequence with geometric progression, and the progression factor is roughly 4. We used these results to propose a natural geometric scale for the quantification of verbal, comparative judgement: a scale with major as well as threshold echelons, and the progression factor 2. Sensitivity analysis with a short and a long geometric scale in the neighbourhood of the natural scale usually shows that the impact scores are rather stable. This is illustrated by the numerical example at the end of the paper.

The relative importance of the criteria can also be established via the pairwise comparison of two alternatives. By this new approach, we found only one geometric scale to quantify the relative importance: a scale with major and threshold echelons, and with progression factor $\sqrt{2}$. Pairwise comparisons at the first and the second evaluation level will accordingly be rather similar, despite the conceptual differences. Sensitivity analysis with a variety of scales does not make sense at the second level. These issues are further discussed in section 4.

In the basic experiment just mentioned, we shall be assuming that the decision maker judges the stimuli on the one-dimensional axis of desirability, under each of the respective criteria. In comparing two motor cars, for instance, he estimates their relative desirability under the consumer-price criterion, their relative desirability under the reliability criterion etc. This hypothesis is further discussed in section 4. We shall also assume that the desirability of the stimuli varies uniformly varies between a minimum level D_{min} and a maximum level D_{max}, at least during the decision process at hand. Moreover, we suppose that the interval (D_{min}, D_{max}) may be categorized in the same way as the intervals considered in Lootsma (1990). Taking $D_j = D_{max} - e_j$ and $D_k = D_{max} - e_k$ to denote the desirability of S_j and S_k respectively, we model the preference for S_j with respect to S_k as

$$\frac{V_j}{V_k} = \frac{e_k}{e_j} = \frac{D_{max} - D_k}{D_{max} - e_j}. \tag{3}$$

The echelons e_j and e_k may be found on a geometric scale. Hence, we also convert the gradations of the decision maker's comparative judgement into numerical values on a geometric scale which is conveniently characterized by a scale parameter γ. Thus, we set

$$r_{jk} = \exp(\gamma \delta_{jk}) \tag{4}$$

where δ_{jk} is an integer-valued index designating the gradation of the decision maker's judgement as follows:

-8 S_j vastly less desirable than S_k,

-6 S_j much less desirable than S_k,

-4 S_j (definitely) less desirable than S_k,

-2 S_j somewhat less desirable than S_k,

 0   S_j as desirable as S_k (equally desirable),

+2 S_j somewhat more desirable than S_k,

+4 S_j (definitely) more desirable than S_k,

+6 S_j much more desirable than S_k,

+8 S_j vastly more desirable than S_k.

It is current practice to describe the gradations in terms of preference, so that the gradation index δ_{jk} can equivalently be used to designate the strength of preference in the following way:

-8   very strong preference for S_k versus S_j,

-6   strong preference for S_k versus S_j,

-4   strict (definite) preference for S_k versus S_j,

-2   weak (mild, moderate) preference for S_k versus S_j,

 0   indifference between S_j and S_k,

+2   weak (mild, moderate) preference for S_j versus S_k,

+4   strict (definite) preference for S_j versus S_k,

+6   strong preference for S_j versus S_k,

+8   very strong preference for S_j versus S_k.

Obviously, weak ("somewhat more") preference for S_j with respect to S_k is converted into $\exp(2\gamma)$, strict ("more") preference into $\exp(4\gamma)$, etc. When the progression factor $\exp(\gamma)$ is set to 2, we have precisely the ratios for comparative judgement announced at the beginning of this section. It is easy to understand why we set δ_{jk} to 1 if the decision maker hesitates between indifference and weak preference for S_j, etc. In summary, we use the even values of the gradation index δ_{jk} to designate the major echelons (the major gradations) of comparative judgement, and the odd values for the threshold echelons (the threshold gradations) to be used when the decision maker hesitates between two adjacent major gradations.

These results prompted us to propose a geometric scale with $\exp(\gamma) = 2$ and $\gamma = 0.7$ as a natural scale for the quantification of the gradations just mentioned. We do not see any reason to maintain the fundamental scale of Saaty (1980). In earlier experiments, we have used a short scale ($\gamma = 0.5$) and a long scale ($\gamma = 1$). Those scales are still recommended for a sensitivity analysis. In summary, the natural scale has the following values assigned to the major gradations:

$\frac{1}{16}$ strict preference for S_k versus S_j,

$\frac{1}{4}$ weak preference for S_k versus S_j,

1   indifference

4   weak preference for S_j versus S_k,

16   strict preference for S_j versus S_k, etc.

so that it is much longer than Saaty's fundamental scale with the respective values $\ldots, \frac{1}{5}, \frac{1}{3}, 1, 3, 5, \ldots$. It is easy to see now, why the last-named scale is controversial. Usually, ratios are expressed on a geometric scale with echelons such as

$$\ldots, 10^{-2}, 10^{-1}, 10^0, 10^1, 10^2, \ldots .$$

Alternatively, we do not directly demarcate the ratios themselves, but we mark off their orders of magnitude on an arithmetic scale with echelons such as

$$\ldots, -2, -1, 0, 1, 2, \ldots .$$

The fundamental scale, however, is neither geometric nor arithmetic. Saaty (1986, 1988, 1990) invokes Fechner's law to explain the choice of the echelons $3, 5, 7, \ldots$, although Stevens' power law (1957) is now generally accepted in psychophysics, and he chooses the echelons $\frac{1}{3}, \frac{1}{5}, \frac{1}{7}, \ldots$ in order to maintain reciprocity. No physical meaning is further assigned to these echelons. The reader may note that the natural scale to express major echelons has the form

$$\ldots, 4^{-2}, 4^{-1}, 4^0, 4^1, 4^2, \ldots . \tag{5}$$

The scales introduced so far seem to apply particularly when the performance of the alternative stimuli can be expressed in physical or monetary units. We propose a direct procedure when the performance can only be expressed in qualitative terms. In the evaluation of cars under the criterion of styling, for instance, we present any pair of cars and we ask the decision maker whether he is indifferent between the two or whether one of them is somewhat more/more/much more desirable than the other from the viewpoint of car-body design. Thereafter, we can immediately assign the proper scale values to the gradations of his comparative judgement.

3 Logarithmic regression

Saaty's proposal (1977) to approximate the vector $V = (\ldots, V_j, \ldots, V_k, \ldots)$ of subjective stimulus values by the Perron-Frobenius eigenvector of the matrix $R = \{r_{jk}\}$ has been criticized by various authors (Johnson, Beine, and Wang (1979), Cogger and Yu (1983), Takeda, Cogger, and Yu (1987)). Alternative proposals were also brought forward. Particularly logarithmic regression or the logarithmic least-squares method has been proposed, not only because it is an evident technique to be used in these circumstances (de Graan (1980), Crawford and Williams (1985), see Saaty and Vargas (1984)), but also on axiomatic grounds (Barzilai, Cook, and Golani (1987)). Following the suggestions of de Graan (Lootsma (1987)), we approximate the vector V of subjective stimulus values via logarithmic regression, that is, we approximate V by the normalized vector $\bar{v}$ which minimizes the expression

$$\Sigma_{j<k}(\ln r_{jk} - \ln v_j + \ln v_k)^2, \tag{6}$$

where the summation is further restricted to the pairs (j, k) judged by the decision maker. He does not really have to consider each pair of stimuli, an advantage which the eigenvector method of Saaty (1980) signally fails to offer. Minimization of (6) is carried out by solving the associated, linear system of normal equations, with variables $u_j = \ln v_j$. Obviously, the u_j have an additive

degree of freedom. The v_j will accordingly have a multiplicative degree of freedom, which is used to single out the normalized vector $\overline{v}$ with components summing up to unity.

By this procedure we calculate stimulus weights for an individual decision maker. We obtain a vector $\overline{v}$ of group weights, possibly a compromise, by minimizing

$$\Sigma_{j<k}\Sigma_{d\in D_{jk}}(ln\ r_{jkd} - ln\ v_j + ln\ v_k)^2, \tag{7}$$

where D_{jk} stands for the set of decision makers who judged the pair (j,k), and r_{jkd} for the estimate of V_j/V_k expressed by decision maker d. We are clearly assuming that the values V_j and V_k of the respective alternatives are the same for all decision makers, an assumption that will be further discussed in section 4. We solve the variables $u_j = ln\ v_j$ from the associated, linear system of normal equations, and we use the multiplicative degree of freedom in the v_j to obtain the normalized minimum solution of (7). It is interesting to note that the calculations remain unchanged when each term in (7) is multiplied by the factor p_d, the relative power of decision maker d, for instance the relative size of the state or the constituency which he represents (see Lootsma (1988)). Then we minimize the expression

$$\Sigma_{j<k}\Sigma_{d\in D_{jk}}(ln\ r_{jkd} - ln\ v_j + ln\ v_k)^2 p_d \quad . \tag{8}$$

by solving a linear system of normal equations. We have the impression that the power game in groups has hitherto been under-estimated in multi-criteria analysis. This might explain why the decision makers sometimes reject the proposed, formalized approach.

The rank order of the calculated stimulus weights does not depend on the scale parameter γ. The leading stimulus remains number one. This can easily be shown as follows. A solution $\overline{u}_j$ to the normal equations associated with (6), (7) or (8) has the form

$$\overline{u}_j = \gamma\ u_j^* + \eta.$$

where u_j^* denotes a solution to the normal equations for $\gamma = 1$, and η the above-named additive degree of freedom (see the author (1987, 1988)). The normalized solution $\overline{v}_j$ is accordingly given by

$$\overline{v}_j = \exp(\gamma\ u_j^*)/ \sum_{j=1}^{p} \exp(\gamma\ u_j^*).$$

For any pair of stimuli S_j and S_k, the ratio of the calculated stimulus weights is

$$\overline{v}_j/\overline{v}_k = \exp(\gamma(u_j^* - u_k^*)).$$

This quantity is greater (smaller) than 1 when $u_j^* - u_k^*$ is positive (negative), so that the positive parameter γ does not affect the rank order of the stimuli.

4   Aggregation, final scores, and rank preservation

. Aggregation is a delicate operation which may present unexpected results to the decision makers when the underlying assumptions are ignored. First, we have to find a common nominator. In the car-selection example (the author (1990)), we have initially been working on two different dimensions: consumer price and reliability. Judgemental statements like "somewhat more

expensive" and "somewhat more reliable" cannot be aggregated, however, unless we make a transition to a new, common dimension. That is the reason why we have taken the expression "somewhat more reliable" to stand for "somewhat more desirable" under the criterion of reliability. Similarly, we assume that the expression "somewhat more expensive" may stand for "somewhat less desirable" under the consumer-price criterion. We shall also assume that the desirability of the alternatives varies over the same interval (D_{min}, D_{max}) under each of the respective criteria. By this hypothesis, we avoid the difficulty that criterion weights may depend on the range of measurement in the original dimensions (Gbodossou (1989)). Working on the axis of desirability has the additional advantage that the decision makers do not have to agree explicitly on the range of performance measurement in the original dimensions. They are only supposed to consider the alternatives under each of the criteria, while the strength of their feelings (or their blood pressure) varies uniformly between the same limits.

Criterion weights are not only related to the emotional or the social values of the criteria. The weights also depend on the manner in which they are used, that is, on the aggregation procedure. Thus, the elicitation of the criterion weights should be related to the aggregation rule, and the rule should properly incorporate the relative importance of the criteria, a concept that seems to be intuitively present in the mind of the decision makers. For the time being, assuming that the criteria C_i have the respective values $W(C_i)$ and that we have approximations $\overline{w}(C_i)$ to them, we first address the construction of an aggregation rule.

The above assumptions enable us to operate with preference ratios because they are all defined in terms of distances from the target D_{max} on the common interval (D_{min}, D_{max}). We consider two alternatives A_j and A_k with their calculated profiles, the vectors $\overline{v}_i(A_j), i = 1, \ldots, m$, and $\overline{v}_i(A_k), i = 1, \ldots, m$, respectively. For each i the partial preference ratio

$$\overline{v}_i(A_j)/\overline{v}_i(A_k), \qquad (9)$$

expressing the relative preference for A_j with respect to A_k under criterion C_i, is unique. Since we are dealing with ratios, it is natural to model the global preference for A_j with respect to A_k by the expression

$$\Pi_{i=1}^{m}(\overline{v}_i(A_j)/\overline{v}_i(A_k))^{\overline{c}_i}, \qquad (10)$$

where $\overline{c}_i$ simply denotes the calculated weight $\overline{w}(C_i)$ of the i-th criterion. In an attempt to express the global preferences for the respective alternatives by final scores $\overline{f}(A_j)$ and $\overline{f}(A_k)$, we set the ratio

$$\overline{f}(A_j)/\overline{f}(A_k) \qquad (11)$$

to (10), whence

$$\overline{f}(A_j) = \alpha\Pi_{i=1}^{m}(\overline{v}_i(A_j))^{\overline{c}_i}. \qquad (12)$$

The final scores have a multiplicative degree of freedom α. They can accordingly be normalized to sum up to unity. We tacitly hope, of course, that the final scores approximate the subjective global values $F(A_j), j = 1, \ldots, n$, of the alternatives (provided that they exist).

The global preference ratio (10) is based on partial preference ratios which do not depend on the units of measurement! Moreover, the ratio (10) suggests the phrases to be used in the elicitation of the relative importance of two criteria C_{i_1} and C_{i_2}. We ask the decision maker to consider two real or imaginary alternatives A_j and A_k such that his preference for A_j over A_k under C_{i_1} is roughly equal to his preference for A_k over A_j under C_{i_2}. If the decision maker is indifferent

between A_j and A_k under the two-criteria simultaneously, the corresponding criterion weights must be equal. Otherwise, the importance of C_{i_1} with respect to C_{i_2} increases or decreases accordingly as his preference for A_j over A_k under the two criteria varies. This type of elicitation is only valid under the assumption that the preferences just mentioned do not depend on the performance of A_j and A_k under the remaining criteria. Moreover, we assume that the relative importance of C_{i_1} versus C_{i_2} does not depend on the strength of preference for A_j versus A_k. Considering the procedure in detail, we estimate the decision maker's preference for A_j over A_k under C_{i_1} and under C_{i_2} by $\exp(\gamma\delta_{jk})$ and $\exp(-\gamma\delta_{jk})$ respectively. Furthermore, we let the preference for A_j over A_k under the two criteria simultaneously be given by $\exp(\gamma\theta_{jk})$, where the index θ_{jk} designates the gradation selected by the decision maker to express his comparative judgement according to the list in section 2.3. Taking ω to denote the unknown ratio of the weights of C_{i_1} and C_{i_2}, we may set

$$\exp\left(\gamma\frac{\omega}{\omega+1}\delta_{jk}\right)\cdot\exp\left(-\gamma\frac{1}{\omega+1}\delta_{jk}\right) = \exp\left(\gamma\theta_{jk}\right).$$

which eventually leads to

$$\omega = \frac{\delta_{jk}+\theta_{jk}}{\delta_{jk}-\theta_{jk}} \tag{13}$$

and

$$\frac{\theta_{jk}}{\delta_{jk}} = \frac{\omega-1}{\omega+1}. \tag{14}$$

The crucial observation is that the ratio ω, whereby we model the relative importance of C_{i_1} versus C_{i_2}, does not depend on the scale parameter γ which is used to quantify the comparative judgement of the alternatives under the respective criteria.

Because such an elicitation procedure is still rather complicated, and because the decision makers seem to be able to judge the relative importance of the criteria without regard to the performance of the alternatives, we propose a simplified procedure. It will be clear that $|\theta_{jk}| < \delta_{jk}$ since the preference for A_j over A_k under the two criteria simultaneously cannot be greater than the preference for A_j over A_k under one of the criteria individually, given the conflict between the alternatives. Table 1 shows the feasible combinations of the gradation indices δ_{jk} and θ_{jk}. The reader will see that the ratio ω varies over the range between 1 and 16 (between $\frac{1}{16}$ and 1) when C_{i_1} is somewhat more/more/ ... important (somewhat less/less/ ... important) than C_{i_2}. Choosing a limited number of major gradations and a geometric sequence of echelons for the conversion of these gradations we obtain the following geometric scale for the pairwise comparison of the criteria C_{i_1} and C_{i_2}:

$\frac{1}{16}$	C_{i_1}	vastly less important than C_{i_2},
$\frac{1}{8}$	C_{i_1}	much less important than C_{i_2}.
$\frac{1}{4}$	C_{i_1}	less important than C_{i_2},
$\frac{1}{2}$	C_{i_1}	somewhat less important than C_{i_2},
1	C_{i_1}	as important as C_{i_2},
2	C_{i_1}	somewhat more important than C_{i_2},
4	C_{i_1}	more important than C_{i_2},
8	C_{i_1}	much more important than C_{i_2},
16	C_{i_1}	vastly more important than C_{i_2}.

Using threshold echelons to quantify hesitations between adjacent gradations, we clearly have a scale with progression factor $\sqrt{2}$ and scale parameter $\gamma = 0.35$.

To illustrate the model, we imagine that the decision maker has a strong preference for A_j over A_k under $C_{i_1}(\delta_{jk} = 6)$ and an equally strong preference for A_k over A_j under C_{i_2}. If he declares C_{i_1} to be somewhat more important than $C_{i_2}(\omega = 2)$, than $\theta_{jk} = 2$ according to (14). This value of the gradation index θ_{jk} implies that the decision maker would have a weak preference for A_j over A_k under the two criteria simultaneously. It is difficult to prove that the model works correctly, we can only say that it seems to be plausible.

The procedure for the pairwise comparison of the criteria will now be clear. First, we have to explain the concept of their relative importance to the decision maker. Thereafter, the criteria may be compared in pairs, practically in the same way as the alternatives under the respective criteria. However, there is only one particular geometric scale ($\gamma = 0.35$), not a variety of scales ($\gamma = 0.5, 0.7$, and 1.0), to quantify the gradations of comparative judgement. Finally, logarithmic regression can also be used to calculate normalized criterion weights. In practice, the difference between pairwise comparisons at the first and the second evaluation level is small (it is ignored in the original version of the AHP). The gradations have the typical form of a linguistic variable: a primary term such as "desirable" at the first level or "important" at the second level, preceded by a quantifier such as "somewhat more", "more", or "much more". There are conceptual differences between the evaluation levels, and this implies that the numerical values of the quantifiers are level-dependent.

We now derive the important result that the rank order of the final scores of the alternatives does not depend on the scale parameter γ. The partial preference ratio of A_j over A_k under criterion C_i is

$$\frac{\overline{v}_i(A_j)}{\overline{v}_i(A_k)} = \exp\left(\gamma\left(u_{ij}^* - u_{ik}^*\right)\right),$$

where the vector with components u_{ij}^* and u_{ik}^* represents a solution to a system of normal equations (see section 3). The global preference ratio (10) for A_j over A_k takes the form

$$\exp\left(\gamma \ \sum_i \overline{c}_i \left(u_{ij}^* - u_{ik}^*\right)\right). \tag{15}$$

Because scale variations do not make sense for the criterion weights, the $\overline{c}_i$ do not depend on γ, and the sign of the exponent in (15) remains unchanged when γ varies over a range of positive values. Thus, the global preference ratio remains greater or smaller than 1 under scale variations within the class of geometric scales.

5   Example and Conclusions

We illustrate our multiplicative variant of the AHP via the example published by the author (1980) to show the potential of the AHP in a nomination procedure. Three candidates for a senior professorship in Operations Research were evaluated under seven criteria. Table 2 exhibits the indices designating the gradations that were unanimously selected by the nomination committee to express the relative desirability of the candidates A, B and C under the respective criteria (first evaluation level). Table 3 displays the gradation indices expressing the relative importance of the

		Gradation index θ_{jk}														
		-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7
	1								1							
	2							$\frac{1}{3}$	1	3						
	3						$\frac{1}{5}$	$\frac{1}{2}$	1	2	5					
Gradation	4					$\frac{1}{7}$	$\frac{1}{3}$	·	1	·	3	7				
index δ_{jk}	5				$\frac{1}{9}$	$\frac{1}{4}$	·	·	1	·	·	4	9			
	6			$\frac{1}{11}$	$\frac{1}{5}$	·	·	·	1	·	·	·	5	11		
	7		$\frac{1}{13}$	$\frac{1}{6}$	·	·	·	·	1	·	·	·	·	6	13	
	8	$\frac{1}{15}$	$\frac{1}{7}$	·	·	·	·	·	1	·	·	·	·	·	7	15

Table 1. Relative importance of two criteria when alternative A_j is preferred over A_k, under the first criterion with the intensity designated by δ_{jk}, under the second criterion with the inverse intensity designated by $-\delta_{jk}$, and under the two criteria simultaneously with the intensity designated by θ_{jk}.

Criteria	C_1	C_2	C_3	C_4	C_5	C_6	C_7	Weights
C_1 Didactic Capability	0	-2	-2	1	0	0	-4	9.3
C_2 Mathematical Creativity	2	0	0	4	2	0	-2	17.7
C_3 Creativity in Applications	2	0	0	1	2	2	0	18.6
C_4 Administrative Capability	-1	-4	-1	0	-1	0	-4	7.6
C_5 Knowledge of OR Methods	0	-2	-2	1	0	-2	-2	9.3
C_6 Experience in Practice	0	0	-2	0	2	0	-1	12.5
C_7 Human Maturity	4	2	0	4	2	1	0	25.1

Table 3. Pairwise comparison of criteria to evaluate three candidates A,B and C. The table exhibits the indices designating the gradations of comparative judgement, and the weights of the criteria on the geometric scale for their relative importance ($\gamma = 0.35$).

Criteria + weights		Gradation indices			Impact scores		
		A	B	C	$\gamma=0.5$	$\gamma=0.7$	$\gamma=1.0$
Didactic	A	0	-4	-2	9.0	4.7	1.6
Capability	B	4	0	2	66.5	76.5	86.7
9.3	C	2	-2	0	24.5	18.9	11.7
Mathematical	A	0	0	-1	27.4	24.9	21.2
Creativity	B	0	0	-1	27.4	24.9	21.2
17.7	C	1	1	0	45.2	50.2	57.6
Creativity in	A	0	2	0	42.2	44.5	46.8
Applications	B	-2	0	-2	15.5	11.0	6.3
18.6	C	0	2	0	42.2	44.5	46.8
Administrative	A	0	2	2	57.6	67.0	78.7
Capability	B	-2	0	0	21.2	16.5	10.7
7.6	C	-2	0	0	21.2	16.5	10.7
Knowledge of	A	0	-1	1	30.7	28.5	24.5
OR Methods	B	1	0	2	50.6	57.4	66.5
9.3	C	-1	-2	0	18.6	14.1	9.0
Experience in	A	0	2	2	57.6	67.0	78.7
Practice	B	-2	0	0	21.2	16.5	10.7
12.5	C	-2	0	0	21.2	16.5	10.7
Human	A	0	0	1	38.4	40.1	42.2
Maturity	B	0	0	1	38.4	40.1	42.2
25.1	C	-1	-1	0	23.3	19.9	15.5
	A				37.3	39.0	41.4
Final Scores	B				31.9	31.3	30.4
	C				30.8	29.7	28.3

Table 2. Evaluation of three candidates A,B and C under seven performance criteria. The table shows the indices designating the selected gradations of comparative judgement, the impact scores of the alternatives on the short scale ($\gamma = 0.5$), the natural scale ($\gamma = 0.7$), and the long scale ($\gamma = 1.0$), as well as their final scores. The criterion weights are taken from Table 3.

criteria (second evaluation level). We recall that in both tables the gradation indices represent the quantifiers of comparative judgement as follows:

```
.....................
-2   somewhat less,
 0   equally,
+2   somewhat more,
+4   more,
.....................
```

Table 2 also shows the impact scores of the alternatives, when the gradations of comparative judgement are encoded on the short scale ($\gamma = 0.5$), the natural scale ($\gamma = 0.7$), and the long scale ($\gamma = 1.0$). In Table 3 the reader may find the criterion weights, when the gradations are converted into numerical values on the typical scale for the criteria ($\gamma = 0.35$). With these data, Table 2 also gives the final scores of the alternatives. Candidate A is leading, the candidates B and C are second and third, regardless of the geometric scale. Note that EXPERT CHOICE, the program based on the original version of the AHP, gives the final scores 36.7, 31.9, and 31.3 respectively. In general, we do not expect large discrepancies between the weights and scores calculated via the original and the multiplicative variant, because the scale sensitivity of the AHP is surprisingly low, whereas the arithmetic-mean aggregation rule is a first-order approximation to the geometric-mean rule. The original version, however, does not always give the most plausible rank order of the final scores.

The AHP was intended to structure a decision process by the introduction of a hierarchy of evaluation levels, much higher than the two-level model considered in the present paper. Given the difficulties encountered in the aggregation step, a hierarchical structure with more than two levels should be thoroughly studied before it is launched in a practical environment. In the present paper, we only formalized the concept of the relative importance of the criteria, via a model which is based on the pairwise comparison of alternatives. In a hierarchy of evaluation levels, we would run up against the relative importance of sub-criteria, sub-sub-criteria etc.. concepts which are still undefined. The original version of the AHP disregards these questions. and constructs multi-level hierarchies as audaciously as it carries out the subsequent analysis. Such a top-down approach is not yet sufficiently well established to be used in practice.

References

[1]   J. Barzilai. W.D. Cook, and B. Golani. Consistent Weights for Judgement Matrices of the Relative Importance of Alternatives. Operations Research Letters 6, 131-134, 1987.

[2]   V. Belton. A Comparison of the Analytic Hierarchy Process and a simple Multi-Attribute Value Function. European Journal of Operational Research 26. 7-21, 1986.

[3]   K.O. Cogger and P.L. Yu, Eigenweight Vectors and Least-Distance Aprroximation for Revealed Preference in Pairwise Weight Ratios. Journal of Optimization Theory and Applications 46, 483-491, 1985.

[4]   G. Crawford and C. Williams. A Note on the Analysis of Subjective Judgement Matrices. Journal of Mathematical Psychology 29, 387-405, 1985.

[5]   A. Gbodossou, La Notion de Poids de Critères/Attributs en Analyse Multicritère: une Etude Empirique. Thèse, Université Laval, Quebec, Canada, 1989.

[6]   J.G. de Graan, Extensions of the Multiple Criteria Analysis Method of Saaty. Paper presented at the EURO IV conference, Cambridge, UK, 1980.

[7]   C.R. Johnson, W.B. Beine, and T.J. Wang, Right-Left Asymmetry in an Eigenvector Ranking Procedure. Journal of Mathematical Psychology 19, 61-64, 1979.

[8]   F.A. Lootsma, Saaty's Priority Theory and the Nomination of a Senior Professor in Operations Research. European Journal of Operational Research 4, 380-388, 1980.

[9]   F.A. Lootsma, Modélisation du Jugement Humain dans l'Analyse Multicritère au moyen de Comparaisons par Paires. R.A.I.R.O./Recherche Opérationnelle 21, 241-257, 1987.

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

[11]   F.A. Lootsma, The French and the American School in Multi-Criteria Decision Analysis. R.A.I.R.O./Recherche Opérationnelle 24, 263-285, 1990.

[12]   Th. L. Saaty, A Scaling Method for Priorities in Hierarchical Structures. Journal of Mathematical Psychology 15, 234-281, 1977.

[13]   Th.L. Saaty, The Analytic Hierarchy Process, Planning, Priority Setting, Resource Allocation. McGraw-Hill, New York, 1980.

[14]   Th.L. Saaty, What is the Analytic Hierarchy Process? In G. Mitra (ed.), Mathematical Models for Decision Support. Springer, Berlin, 1988, pp. 109-122.

[15]   Th. L. Saaty, Physics as a Decision Theory. European Journal of Operational Research 48, 98-104, 1990.

[16]   Th. L. Saaty and L. Vargas, Inconsistency and Rank Preservation. Journal of Mathematical Psychology 28, 205-214, 1984.

[17]   S.S. Stevens, On the Psychophysical Law. Psychological Review 64, 153-181, 1957.

[18]   E. Takeda, K.O. Cogger, and P.L. Yu, Estimating Criterion Weights using Eigenvectors: a Comparative Study. European Journal of Operational Research 29, 360-369, 1987.

[19]   F. Zahedi, The Analytic Hierarchy Process - a Survey of the Method and its Applications. Interfaces 16, Nr. 4, 96-108, 1986.