ORIGINAL PAPER

Weighting methods in multi-attribute assessment of transport projects

Paolo Delle Site · Francesco Filippi

Abstract The paper considers the application to transport project assessment of multi-attribute value theory (MAVT). The aim is to compare the results of three weighting methods which are used in MAVT: ratio with swings, Saaty scale with swings, and trade-off. An experiment is set up for a decision problem relating to a public transport project. A sample of individuals is asked to provide judgments according to the three methods. The weights obtained by ratio with swings and by Saaty scale with swings show a high correlation, while weights obtained by each of these methods show a low correlation with trade-off. The weight attached to the cost attribute is higher in trade-off, which is implemented as pricing out, than in the other methods; two explanations for this result are proposed: the compatibility principle and loss aversion. The paper also provides insight on how the attributes of transport project alternatives should be described by indicators so that individuals can attach a value to them.

Keywords Assessment · Multi-criteria analysis · Multi-attribute value theory · Weights

1 Introduction

The assessment of projects, meant here as capital investments that create transport infrastructure, supports the activity of decision makers. In the case of public decision makers the assessment is used as a tool to assist the process of planning transport infrastructure. The assessment is concerned with achieving social (as opposed to private) objectives, including improvement of economic efficiency, reduction of the damage on the environment, improvement of safety.

A review of transport project assessment methodologies in use in European Union (EU) countries conducted within the EU-funded project EUNET has shown that, although cost-benefit analysis is predominant, recent developments are in the direction of greater use of multi-criteria analysis [7]. An example of this move towards multi-criteria analysis is the methodology for multi-modal studies adopted in the UK [8, 23]. One reason is the simplicity of multi-criteria analysis in taking into account non-marketable effects and qualitative criteria.

Another issue of concern in project assessment for public decisions is the inclusion of the variety of points of view of the stakeholders involved in the planning process. This calls for extending the traditional cost-benefit analysis approach [14]. A line of research explores how participation techniques can be accommodated within a multi-criteria analysis framework. Various multi-stakeholder approaches have been proposed within this framework to deal with the conflicts which arise because of the uneven distribution of costs and benefits and of the different objectives and weights attached to them [2, 4, 18].

Multi-criteria analysis is today widely used in transport project studies commissioned by public bodies. Guidance for its application exists in some countries in the form of manuals, an example is the manual issued by the UK government [9]. Multi-criteria analysis is also widely used as assessment tool in EU-funded research projects. A few projects which have dealt with the development of comprehensive frameworks for the modeling and assessment of transport policies, including investment policies,

P. Delle Site (✉) · F. Filippi
DITS, Dipartimento Idraulica Trasporti Strade,
University of Rome "La Sapienza",
Via Eudossiana 18,
00184 Rome, Italy
e-mail: paolo.dellesite@uniroma1.it

have adopted multi-criteria analysis: among these PROPO-LIS which has dealt with urban transport [16].

Typical decision problems addressed with multi-criteria analysis include prioritizing projects within a programme and choosing between alternative design solutions, often location-related, for an individual project. In principle, the projects cover the whole range of geographical levels which mark the boundaries of their impacts and determine the decision making bodies concerned: from international to national and local. Application examples taken from recent literature include the plans for the pan-European road and railway networks sponsored by the United Nations [27], the location of an intermodal barge terminal [17], the location of a high-speed railway station in a metropolitan area [19], the layout of a railway link to a city port [2], the plans for the development of the road network in an urban area [4], the plans for the development of the rail public transport network in a city [13].

Multi-attribute value theory (MAVT) is a well known and widely used multi-criteria methodology (a review of its mathematical foundation and of different versions implemented in commercial software is in Figueira et al. [12]). MAVT uses a functional approach, i.e. a multi-attribute value function is associated with each alternative. Most commonly the value function takes an additive form. Typically the additive value function is assessed using the weighted summation approach: single-attribute value functions are separately assessed, then attribute scaling constants are assessed to weight and combine the single-attribute functions.

In MAVT the scales of the single-attribute value functions are set up with reference to well-defined anchor points: this makes it possible for both the scores of the alternatives on the criteria and the weights of the criteria to be normalized with respect to the same range over which the alternatives vary on each criterion. The consequence of this property is that the elicitation of weights of the criteria is anchored to clearly defined alternatives. This helps the evaluators in their judgment task because it gives them reference points.

Conversely, the other functional methodology widely used, the Analytic Hierarchy Process (AHP) proposed by Saaty, suffers from the limitation that consistency with the mathematical structure of the model would require anchoring to alternatives which can be ill-defined. Weights would need to be derived from pair comparisons on the relative importance of a change from the value 0 for the score to the value 1 for the score on each criterion: however, given the ratio scale used for the scores, it might not always be possible to define the values taken by the attributes for the alternatives with score 0 and score 1. An example is when the attribute is defined by a quantitative indicator but is unfavourable like construction costs or when it is defined

only qualitatively. In these instances the evaluator does not know what the reference points with score 0 and score 1 are. This issue is discussed in Dyer [10] who also shows how this limitation of the AHP can be corrected, using ideas from MAVT, by rescaling the scores. Ferrari [11] proposes a similar modification to the AHP methodology and illustrates this with an application to transport projects.

There are different weight elicitation methods which are consistent with the classical additive model used in MAVT. Theoretical rigour requires all methods be implemented with explicit reference to the attribute ranges defined by the anchor points. With ratio weights (proposed by von Winterfeldt and Edwards [29]) and semantic scales, such as the AHP scale (proposed by Saaty [22]) and the MACBETH scale (proposed by Bana e Costa and Vasnick [1]), the weight eliciting questions are derived from an interpretation of weights as swings and the respondent is asked to compare swings among different attributes. The trade-off method (proposed by Keeney and Raiffa [15]) uses questions where the respondent is asked to adjust the outcome of one attribute to achieve indifference among alternatives.

If different weight elicitation methods can be used within the same modelling framework, the research question arises whether different methods yield significantly (in statistical sense) different results in terms of weights. This in turn would have an effect on the resulting ranking of the alternatives. This type of investigation is helpful to detect behavioural biases. If we set up an experiment and find that the differences in the results obtained with the different methods are statistically significant, we are in a position to predict how the results are likely to differ using different weighting methods if we apply the multi-criteria analysis to new problems.

This type of analysis has already been carried out. Literature offers a range of experimental studies which, however, have considered decision contexts different from transport. Bell et al. [5] have considered climate change policies, Borcherding et al. [6] the siting of nuclear waste repositories, Pöyhönen and Hämäläinen [21] the evaluation of job alternatives and Schoemaker and Waid [24] the choice of candidates for college admission.

The results found by these authors are in agreement: the resulting weights do depend on the weighting methodology. Correlation was used as a descriptive measure of the convergence degree between data sets of results. Wherever one method is found to be more correlated with a second one than with a third one the statistical significance of these differences in correlations between pair of methods has been assessed by using p-values as in conventional hypothesis testing practice. In words, the p-values are a measure of the probability that the differences occur only by chance. In particular, it

was found that trade-off is an outlier, i.e. it shows low correlations with other methods.

The paper considers MAVT models and presents the results of an experiment aimed at assessing if the weights obtained with different weight elicitation methods are different in a statistically significant sense. Methods compared include ratio with swings, AHP scale with swings, trade-off. The experiment is carried out for a decision between layout alternatives of a metro line. The paper also reports on the insight which the experiment has provided on how the attributes of the alternatives should be described by indicators so that individuals can attach a value to them. Thus the paper contributes both to suggesting how MAVT should be applied to transport projects and to highlighting the differences in weights that can be expected when different weighting methods are used.

2 Attribute weighting

Weight elicitation methods must be consistent with the underlying mathematical model. The model here is the classical additive model of MAVT:

$$v(\mathbf{x}) = \sum_{i=1}^{n} w_i \cdot v_i(x_i)$$

where:

$v(\mathbf{x})$ multi-attribute value function of the alternative
$\mathbf{x}$ attribute vector of the alternative
w_i weight, or scaling constant, of attribute x_i
$v_i(x_i)$ single-attribute value function (score) on attribute x_i
n number of attributes

The single-attribute value functions are normalized so that (this choice is referred to as local scaling):

$$v_i(x_i^-) = 0 \quad i = 1, ..n$$
$$v_i(x_i^+) = 1 \quad i = 1, ..n$$

where:

x_i^- worst outcome of attribute x_i (anchor point) among the alternatives to be assessed

x_i^+ best outcome of attribute x_i (anchor point).

The multi-attribute value function is also normalized between 0 and 1

$$v(\mathbf{x}^-) = 0$$
$$v(\mathbf{x}^+) = 1$$

With this normalization of the single-attribute value functions and of the multi-attribute value function the weights must satisfy the restriction:

$$\sum_{i=1}^{n} w_i = 1$$

The interpretation of weights as swings is based on the following. Given the two alternatives:

$$\mathbf{x^1} = \left[x_i = x_i^-; \quad x_k = x_k^* \quad k \neq i \right]$$
$$\mathbf{x^2} = \left[x_i = x_i^+; \quad x_k = x_k^* \quad k \neq i \right]$$

we get:

$$v(\mathbf{x^2}) - v(\mathbf{x^1}) = w_i \cdot v_i(x_i^+) - w_i \cdot v_i(x_i^-) = w_i$$

Ratio weighting and weighting with AHP scale are elicitation methods which can be used consistently with this interpretation. In ratio weighting 100 points are assigned to the attribute with the highest value swing. The other swings are valued by judgment in terms of percentage of the highest value swing. Weights then follow.

In weighting with the AHP scale, for each pair of attributes the dominance of the swing in the first attribute on the swing in the second attribute is assessed by judgment according to the Saaty semantic scale. Weights are then derived based on the principal eigenvector of the pairwise comparison matrix as proposed by Saaty in his classical paper [22].

The interpretation of weights as trade-offs is based on the following. Given the alternative:

$$\mathbf{x^a} = \left[x_i = x_i^+; \quad x_j = x_j^-; \quad x_k = x_k^* \quad k \neq i,j \right]$$

we search the outcome x_i^j of the attribute x_i that makes indifferent to $\mathbf{x^a}$ the following alternative:

$$\mathbf{x^b} = \left[x_i = x_i^j; \quad x_j = x_j^+; \quad x_k = x_k^* \quad k \neq i,j \right]$$

From $v(\mathbf{x^a}) = v(\mathbf{x^b})$ we get

$$\frac{w_i}{w_j} = \frac{1}{1 - v_i(x_i^j)}$$

In trade-off weighting the outcome of the attribute x_i is adjusted by judgment n-1 times in order to assess the n-1 ratios w_i/w_j. Weights then follow. This method requires that the attribute to be adjusted is described with a single quantitative indicator.

3 The experiment

3.1 The hypothesis

The experiment considers three weight elicitation methods: ratio with swings, AHP scale with swings, and trade-off. To formulate the hypothesis of the experiment we define convergent validity as the correlation shown by a pair of data sets obtained with different elicitation methods. The hypothesis of the experiment is that the convergent validity

between weights elicited with different methods depends on the methods.

3.2 The decision problem

The decision problem relates to the layout alternatives for the planned fourth line of the metro network in Rome. Two layout alternatives only are considered, one of minimum length (base alternative) and one of maximum length (complete alternative). Criteria include one cost attribute: construction costs, and three benefit attributes: travel time, road safety and air quality.

The experiment was carried out a first time based on quantitative descriptors for the benefit attributes that are directly obtainable from conventional transport planning models. Descriptors of benefits were yearly reductions, on the without-project case, of respectively, hours spent travelling, accidents resulting in injury or death, and particulate matter (PM) emissions.

The individuals participating in the experiment who were asked the questions to elicit weights reported a difficulty in attaching a value to the attributes described by total hours saved in a year and PM emissions saved in a year.

Thus the experiment was repeated based on a different choice of the descriptors for the travel time and air quality benefits. For the former the description was given in terms of number of yearly users having a time saving for one trip of a given value. For the latter the description was given in terms of consequences on health of the reduction of PM emissions: premature deaths and days of restricted working activity saved.

The descriptions of the four attributes for the two layout alternatives are in Table 1.

3.3 The multi-criteria model

The additive MAVT model is adopted. Single-attribute value functions are set up using local scaling (i.e. either 0 and 1 scores are assigned to the two alternatives). For the

trade-off method which requires the evaluation of the single-attribute function for the adjusted attribute a linearity assumption is made with respect to the quantitative descriptor of the attribute.

3.4 The questionnaire

Attribute weights are elicited in the experiment with a questionnaire. Fourteen experts in transport planning, including academics and practitioners, participated in the experiment. The questionnaire includes an introduction where the decision problem is explained and information on the alternatives are provided. These include descriptors for the four attributes used in the multi-criteria model as in Table 1. The questionnaire includes then three blocks of questions aimed at eliciting weights. Each block uses one elicitation method. The sequence in which blocks are presented to respondents was randomized to avoid that the order of questions could bias the results.

In the ratio and AHP methods the subjects are confronted with the change from the worst to the best outcome of each attribute (where worst and best are identified based on the two layout alternatives under examination).

In the ratio block the questions are as follows.

"Given an alternative with the worst outcome for each attribute, which attribute would you change first from worst to best? Which second? ... Assign 100 to the first. Assign between 0 and 100 to the second. … "

In the AHP scale block the questions are as follows.

"Compare the swing for worst to best of the attribute construction costs with the swing from worst to best of the attribute travel time benefits. Use the Saaty scale (1 to 9: 1 equal importance, 9 extreme importance) to judge the dominance of the first swing on the second (use the reciprocals 1 to 1/9 if the second swing dominates the first)". The question is repeated for each pair of attributes.

Table 1 Alternatives and attributes

Attributes	Layout alternative	
	Base	Complete
1. Construction costs	1.7 billion €	2.4 billion €
2. Travel time benefits	360,000 daily users of the new line have an average saving of 12 min per trip	500,000 daily users of the new line have an average saving of 12 min per trip
3. Road safety benefits	in 1 year: 67 accidents, fatal and with injured, are saved, of them 3% are fatal	in 1 year: 101 accidents, fatal and with injured, are saved, of them 3% are fatal
4. Benefits relating to the effects on citizens' health of particulate matter emissions	in 1 year: 90 premature deaths due to cardio-respiratory diseases are saved, 4,100 days where activity is restricted due to health conditions are saved	in 1 year: 158 premature deaths due to cardio-respiratory diseases are saved, 7,200 days where activity is restricted due to health conditions are saved

Eur. Transp. Res. Rev. (2009) 1:199–206

203

In the trade-off block the attribute that subjects are asked to adjust is construction costs and the questions are formulated by asking the willingness to pay of the society to achieve improvements in each of the benefit attributes. Trade-off is equivalent in this instance to a pricing out method.

In the trade-off block the questions are as follows.

"Given an alternative with lowest construction costs and worst benefits, how much would you be willing to increase construction costs to change travel time benefits from worst to best?" The question is repeated for each benefit attribute.

3.5 The statistical analysis

The weights, derived from the three elicitation methods, are calculated for each subject. Within-subject inter-method Pearson correlation coefficients of weights are calculated and then averaged over subjects. This is done first for all weights, and, second, individually for the weight of each attribute.

The within-subject inter-method correlation for all weights is:

$$r_{ab}^s = \frac{\sum_{i=1}^{n} \left(w_{ia}^s - \frac{1}{n}\right)\left(w_{ib}^s - \frac{1}{n}\right)}{\sqrt{\sum_{i=1}^{n}\left(w_{ia}^s - \frac{1}{n}\right)^2 \cdot \sum_{i=1}^{n}\left(w_{ib}^s - \frac{1}{n}\right)^2}}$$

Where

s	index of individual
a,b	index of weighting method
i	index of attribute

This reduces to the following correlation for the weights of an individual attribute:

$$r_{ab}^s = \frac{\left(w_a^s - \frac{1}{n}\right)\left(w_b^s - \frac{1}{n}\right)}{\sqrt{\left(w_a^s - \frac{1}{n}\right)^2 \cdot \left(w_b^s - \frac{1}{n}\right)^2}}$$

Statistical significance is assessed using the p-values obtained from the Wilcoxon matched pairs signed-rank test. The logic is that of hypothesis testing. We have a sample of subjects and for each subject a pair of measurements. The null hypothesis H_0 states that the two series of measurements are drawn from the same population, i.e. they differ only by chance. The test chooses a particular statistic which is based on the measurements. We calculate then the a-priori probability (p-value) that if H_0 is true the statistic takes values as extreme as the value actually obtained from the sample. If this probability is low we can reject H_0 and

say that the differences in the two series of measurements are statistically significant.

In the Wilcoxon test the difference of the two measurements is calculated for each subject, then the absolute values of the differences are ordered in ranks. The statistic of the test is the sum of the ranks corresponding to positive differences. If H_0 is true the sum of the ranks corresponding to positive differences must be of the same order of magnitude, in probability terms, of the sum of the ranks corresponding to negative differences. Tables provide the probability (p-values) that if H_0 is true the statistic takes values greater or equal than the one obtained from the sample. Siegel and Castellan [25] provide these tables in the case of small samples (less than fifteen subjects).

Finally, the ranking of the two layout alternatives, resulting from the different weight vectors, are calculated, according to the additive MAVT model, for each subject.

4 Results

The hypothesis of the experiment is supported by the results, i.e. we can state that the convergent validity depends on the methods. Table 2 shows average correlations of weights for pairs of methods. Ratio with swings is well correlated with AHP scale with swings (0.789). The correlation of ratio and of AHP scale with trade-off is weak (respectively, 0.113 and 0.081). The differences in correlations are statistically significant: the correlation of ratio with AHP scale is significantly higher than the correlations of each of these two methods with trade-off (respectively $p<0.0020$ and $p<0.0002$). These results are in agreement with those of previous experiments where inter-method correlations of weights were calculated [5, 6, 24].

Table 3 shows average correlations of weights for individual attributes. For all attributes the correlation between ratio and AHP scale are higher than the correlations between these two methods and trade-off. The construction cost attribute shows the most marked differences in inter-method correlation. The correlation of the weight of this attribute is positive between ratio and AHP scale (0.571), it is even negative for the pair ratio with trade-off (−0.571) and for the pair AHP scale with trade-off (−0.429). This result suggests that high differences can be expected in the values of the weight of this attribute when moving from ratio and AHP scale to trade-off.

Table 2 Average within-subject inter-method correlation of weights

Ratio and AHP scale	0.789
Ratio and trade-off	0.113
AHP scale and trade-off	0.081

204

Eur. Transp. Res. Rev. (2009) 1:199–206

Table 3 Average within-subject inter-method correlation of weights of individual attributes

	1. Construction costs	2. Travel time benefits	3. Road safety benefits	4. Air quality benefits
Ratio and AHP scale	0.571	0.857	0.385	0.714
Ratio and trade-off	−0.571	0.571	0.333	0.333
AHP scale and trade-off	−0.429	0.714	0.385	0.385

And in fact, the analysis of weight values shows that the average weight of the cost attribute in the ratio method (average 0.209) and that in the AHP scale method (average 0.148) are lower than the average weight of the cost attribute in the trade-off method (average 0.452). These differences in weight values are significant at $p<0.0001$ in both cases. Thus individuals attach a value to the range of cost variation to varying extent according to the elicitation method.

The finding that the cost attribute has a higher weight in the trade-off method can be explained in terms of the compatibility principle (Slovic et al. [26] and Tversky et al. [28] discuss the principle and support it with experimental evidence). This states that the weight of any stimulus element is enhanced by its compatibility with the response mode. Individuals tend to focus on stimulus elements that are compatible with their response. In the case here, the trade-off method is implemented with pricing out questions. Therefore, in trade-off the construction cost attribute matches the response scale, i.e. both the attribute and the response of the individuals is in monetary units. In ratio and AHP scale the response scale is different from money. Thus in trade-off the attribute construction costs tends to be accentuated.

The finding on the weight of the cost attribute can also be seen as a manifestation of a loss aversion effect, well known in behavioural economics, when it is recognised that money is given lower value if the question is about reducing costs borne (i.e. a gain) as in ratio and AHP scale, it is given higher value if the question is about bearing higher costs and willingness to pay (i.e. a loss) as in trade-off.

The different weight assigned, according to the elicitation method, to the construction cost attribute translates into different values of the money equivalents for the three benefit attributes. The ratio of the benefit attribute weight to the construction cost weight equals the change in construc-

tion costs equivalent to bringing the benefit attribute from its worst to its best outcome. Table 4 shows that the differences of the money equivalents between the first two methods, ratio and AHP scale, and trade-off are substantial. In trade-off money equivalents are much lower as a consequence of the higher weight for the construction cost attribute. A similar result had been obtained by Borcherding et al. [6].

It is relevant at this point to recall that different measures of the equivalents between money and another attribute can be defined according to whether a loss or a gain is incurred. The theory of reference-dependent preferences, which applies the ideas of prospect theory to riskless choice, define four valuation measures (found in Bateman et al. [3] and in Munro and Sugden [20]): willingness to pay (money one would pay in return for a given gain in the attribute), willingness to accept (money one would accept in return for a given loss in the attribute), equivalent loss (money loss one would accept in place of a given loss in the attribute), and equivalent gain (money gain one would accept in place of a given gain in the attribute). A central assumption in the theory is loss aversion: losses are valued more heavily than gains. This implies that the four valuation measures need to be different and satisfy certain inequalities.

The result obtained here can be explained in terms of difference between equivalent gain and willingness to pay. In ratio and AHP scale the questions were about gains of both money and benefits (equivalent gain). In trade-off the question was about a loss in money and a gain in benefits (willingness to pay). According to the theory of reference-dependent preferences the equivalent gain needs to be higher than the willingness to pay and this expectation is confirmed by the results here.

Finally we note that the weighting method affects as well the resulting ranking of the two alternatives. When moving from ratio to AHP scale nobody in the sample changes the ranking. Conversely, when moving from ratio to trade-off

Table 4 Money equivalents (billion €) of the benefit attributes

	Travel time benefits	Road safety benefits	Air quality benefits
Ratio	1.67	1.62	2.04
AHP scale	3.17	4.61	5.27
Trade-off	0.35	0.39	0.60

and from AHP scale to trade-off 21% of the individuals in the sample (both cases) shows a reversal in the ranking.

5 Conclusions

The paper has presented the results of a weight elicitation experiment within a decision problem relating to transport projects. The experiment has made evident that the attributes of the transport projects need to be described in terms of impacts to which individuals are able to attach a value. The usual outputs of transport planning models are not sufficient if the aim is to conduct multi-criteria analyses. This translates into the rejection of descriptors such as yearly passenger-hours and pollutant emissions, and adoption of descriptors based on yearly users gaining a time saving and average time saving per trip, and on health consequences of pollutant emissions.

The paper has compared the weights obtained with the following three weight elicitation methods: ratio with swings, AHP scale with swings and trade-off. Results are in agreement with those found in the literature for other decision problems. The correlation of weights between a pair of methods depends on the methods. In particular, trade-off turns out to be an outlier. The main difference is shown by the weight of the construction cost attribute. The weight for this attribute obtained with the trade-off method is higher than with the other two methods. As the trade-off method has been implemented with pricing out questions this result can be explained in terms of the compatibility principle. An alternative explanation is loss aversion. When moving from the first two methods to trade-off the ranking of alternatives also shows to be affected.

Results suggest that selection of the weight elicitation method is of relevance in multi-attribute models as this affects both the weights and the resulting ranking of the alternatives. The findings here add to the body of experimental research documenting that weights are "constructed" in the elicitation process rather than "uncovered".

Further research could be carried out by implementing the trade-off method according to the four valuation measures which are defined in the theory of reference-dependent preferences. The experiment here has considered only a willingness to pay question but the trade-off method could be implemented by asking also willingness to accept, equivalent gain and equivalent loss questions. In case the weights obtained from the four variants of the trade-off method were highly correlated with each other one might conclude that weights depend principally on the methods (ratio and AHP scale on the one hand, trade-off on the other) and that differences can be explained on the basis of the different response scales used by the methods. In case they were not one might relate the differences in weights to valuation of losses and gains.

References

1. Bana e Costa CA, Vasnick JC (1994) MACBETH—an interactive path toward the construction of cardinal value functions. Int Trans Oper Res 1:489–500
2. Bana e Costa CA, Nunes da Silva F, Vansnick J-C (2001) Conflict dissolution in the public sector: a case-study. Eur J Oper Res 130 (2):388–401
3. Bateman I, Munro A, Rhodes B, Starmer C, Sugden R (1997) A test of the theory of reference-dependent preferences. Q J Econ 112(2):479–505
4. Beinat E (ed) (1998) A methodology for policy analysis and spatial conflicts in transport policies. Final Report of the DTCS (Spatial decision support for negotiation and conflict resolution of environmental and economic effects of transport policies) Project, European Commission DG12
5. Bell ML, Hobbs BF, Elliott EM, Ellis H, Robinson Z (2001) An evaluation of multi-criteria methods in integrated assessment of climate policy. J Multi-Criteria Decis Anal 10(5):229–256
6. Borcherding K, Eppel T, von Winterfeldt D (1991) Comparison of weighting judgements in multiattribute utility measurement. Manage Sci 37(12):1603–1619
7. Bristow AL, Nellthorp J (2000) Transport project appraisal in the European Union. Transp Policy 7(1):51–60
8. DETR (2000) Guidance on the Methodology for Multi-Modal Studies (GOMMMS). Department of the Environment, Transport and the Regions, London
9. Dodgson J, Spackman M, Pearman AD, Philips LD (2000) Multi-criteria Analysis: a Manual. Department of the Environment, Transport and the Regions, London
10. Dyer JS (1990) Remarks on the analytic hierarchy process. Manage Sci 36(3):249–258
11. Ferrari P (2003) A method for choosing from among alternative transportation projects. Eur J Oper Res 150(1):194–203
12. Figueira J, Greco S., Ehrgott M (2004) Multiple Criteria Decision Analysis: State of the Art Surveys. Springer
13. Gerçek H, Karpak B, Kilinçaslan T (2004) A multiple criteria approach for the evaluation of the rail transit network in Istanbul. Transportation 31(2):203–228
14. Haezendonck E (ed) (2007) Transport Policy Evaluation: Extending the Social Cost-benefit Approach. Edward Elgar, Cheltenham
15. Keeney RL, Raiffa H (1976) Decision with Multiple Objectives: Preferences and Value Trade-offs. Wiley, New York
16. Lautso K, Wegener M, Spiekermann K, Shepperd I, Steadman P, Martino A, Domingo R, Gayda S (2004) PROPOLIS: Planning and Research of Policy for Land Use and Transport for Increasing Urban Sustainability. Final Report, Fifth Framework Programme, European Commission
17. Macharis C (2004) A methodology to evaluate potential locations for intermodal barge terminals: a policy decision support tools. In: Beuthe M, Himanen V, Reggiani A, Zamparini L (eds) Transport Developments and Innovations in an Evolving World. Springer, Berlin, pp 211–234
18. Macharis C (2004) Multi-criteria anlysis as a tool to include stakeholders in project evaluation: the MAMCA method. In: Haezendonck E (ed) Transport Policy Evaluation: Extending the

Social Cost-benefit Approach. Edward Elgar, Cheltenham, pp 115–131

19. Mateus R, Ferreira JA, Carreira J (2008) Multicriteria decision analysis (MCDA): Central Porto high-speed railway station. Eur J Oper Res 187(1):1–18

20. Munro A, Sugden R (2003) On the theory of reference-dependent preferences. J Econ Behav Organ 50(4):407–428

21. Pöyhönen M, Hämäläinen RP (2001) On the convergence of multiattribute weighting methods. Eur J Oper Res 129(3):569–585

22. Saaty TL (1977) Scaling method for priorities in hierarchical structures. J Math Psychol 15(3):234–281

23. Sayers TM, Jessop AT, Hills PJ (2003) Multi-criteria evaluation of transport options—flexible, transparent and user-friendly? Transp Policy 10(2):95–105

24. Schoemaker PJ, Waid CC (1982) An experimental comparison of different approaches to determining weights in additive value models. Manage Sci 28(2):182–196

25. Siegel S, Castellan NJ (1988) Nonparametric Statistics for the Behavioral Sciences. McGraw-Hill

26. Slovic P, Griffin D, Tversky A (2002) Compatibility effects in judgment and choice. In: Gilovich T, Griffin D, Kahneman D (eds) Heuristics and Biases. The Psychology of Intuitive Judgment. Cambridge University Press, pp 217–229

27. Tsamboulas DA (2007) A tool for prioritizing multinational transport infrastructure investments. Transp Policy 14(1):11–26

28. Tversky A, Sattah S, Slovic P (1988) Contingent weighting in judgment and choice. Psychol Rev 95(3):371–384

29. Von Winterfeldt D, Edwards W (1986) Decision Analysis and Behavioural Research. Cambridge University, Cambridge