Transport Policy 23 (2012) 15–24

Contents lists available at SciVerse ScienceDirect

Transport Policy

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

Combining cost-benefit and multi-criteria analysis to prioritise a national road infrastructure programme

Astrid Gühnemann [a,*], James J. Laird [a], Alan D. Pearman [b]

[a] University of Leeds, Institute for Transport Studies, 36-40 University Road, Leeds LS2 9JT, United Kingdom
[b] University of Leeds, Centre for Decision Research, Maurice Keyworth Building, Leeds LS2 9JT, United Kingdom

ARTICLE INFO

Available online 4 July 2012

Keywords:
Cost-benefit analysis
Multi-criteria analysis
Investment
Roads
Sensitivity testing

ABSTRACT

This paper develops and then applies a novel approach of combining cost-benefit analysis (CBA) and multi-criteria analysis (MCA) within a road infrastructure development programme with the aim to support the effective implementation of transport policy when prioritising projects. By incorporating CBA results into an MCA framework this approach retains the strengths of each appraisal method and provides a procedure for decision makers to create an initial ranking of projects which is consistent between all candidate investments and has a clear link to policy goals. We further develop an approach for an incremental analysis that eliminates mutually exclusive projects and allows decision makers to develop a cost-effective investment programme in compliance with their strategic goals. Stakeholder confidence in the outcome of any infrastructure investment ranking exercise is important and can be enhanced by an understanding of the robustness of the ranking to variations in key inputs to the assessment exercise. Two complementary perspectives on sensitivity testing are outlined which between them facilitate an assessment of the robustness of the project ranking obtained. The applicability of the approach has been successfully demonstrated for the National Secondary Road Network in Ireland.

1. Introduction

Prioritising competing transport infrastructure investment schemes is a common problem at all levels of transport policy. Often the number of schemes to be considered is substantial. With pressure on resources, growing awareness of the range of scheme impacts to be taken into account and public pressure to demonstrate a transparent and consistent ranking process, decision making bodies require comprehensive and reliable appraisal methods.

Cost-benefit analysis (CBA), multi-criteria analysis (MCA), or a combination of both is typically applied in the appraisal processes which support the implementation of transport policy (see, e.g., Bristow and Nellthorp, 2000; Grant-Muller et al., 2001; Odgaard et al., 2005). Each brings its own perspective. However, while the principle that the two might usefully be used in tandem is long-established (see, e.g., Manheim et al., 1975) in practice effective implementation has proven elusive. In those few practical applications where they are used together, results can be formally combined, e.g. in the decision support system COSIMA, see Leleur (2012), or they may be presented separately for decision makers to make their own tradeoffs. Research shows that even in countries where CBA is the main tool of assessment, investment decisions are often strongly influenced by other political preferences (Eliasson and Lundberg, 2012; Quinet, 2011) or decision makers take account of criteria in a non-monetised way despite those criteria being included in the CBA (Odeck, 2010). Hence, providing a clear procedure for combining cost-benefit analysis and multi-criteria assessment results can help to make decision-makers' preferences more transparent by the explicit inclusion of all decision factors and thus establish a strong link between policy objectives and appraisal results while still providing information for cost-efficient investment decisions.

When looking in practice to develop methods to apply CBA and MCA in combination, commonly four problems must be addressed:

- There are often already procedures in place (CBA, MCA or both) that cannot simply be abandoned. The analyst must seek to accommodate both while ensuring adequate compatibility between the two is maintained so that the overall outcome is meaningful.
- Assessment criteria have to a greater or lesser extent already been established, measurement scales chosen and monetary values and/or multi-criteria weights determined. Often comparability with previous evaluations (even if relating to a different range of projects) has to be maintained.
- There is significant uncertainty about impact levels and weights, but not the time or resources formally to seek to model the full extent of the uncertainty.

* Corresponding author.
E-mail address: a.guehnemann@its.leeds.ac.uk (A. Gühnemann).

0967-070X/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.tranpol.2012.05.005

- Nonetheless the decision making body is keen to understand the robustness of project rankings, to have some insurance against criticisms of the rankings put forward.

This paper responds to the relatively limited number of real-life applications that combine CBA and MCA that directly addresses the above problems in a novel way and presents an application to a ‘real’ data set. An innovative procedure is developed for expressing the cost-benefit results in a way that facilitates the combination. The procedure then works directly to combine MCA and CBA measures, using already established measurement procedures, recognising the uncertainty that surrounds the parameterisation of the overall evaluation procedure and offering insights about the robustness of the ranking obtained.

2. Development of an MCA scoring and weighting framework incorporating CBA elements

2.1. Multi-criteria analysis in transport appraisal

In supporting the implementation of transport policy multi-criteria analysis has a number of advantages over conventional cost benefit analysis. van Wee (2011) provides a discussion of limitations of CBA from an ethical perspective. MCA can overcome several of these: it facilitates a stronger alignment with espoused transport policy by allowing impacts that cannot be expressed on a monetary scale or easily be quantified, but which are recognised as important by policy makers, such as distributional impacts, environmental effects or the achievement of strategic policy goals, to be formally included in an appraisal. It also allows the explicit exploration of different viewpoints from transport sector stakeholders (Thomopoulos et al., 2009; Macharis et al., 2010) which is increasingly a standard political expectation.

Different MCA techniques vary in how initial performance results against policy objectives are processed and aggregated (overviews of MCA techniques are provided, e.g., by Nijkamp et al. (1990), DCLG (2009), and for the transport sector in the PROPOLIS project Lautso et al. (2004)). A common element is a valuation step which is applied to transform impacts from their original units, e.g. qualitative descriptions, into scores on a preference scale. These results are generally presented to decision makers, together with further information in the form of tables, e.g. the Appraisal Summary Table in the transport appraisal guidance in the UK (DfT, 2011a).

Weights can then be applied to the impact scores in order to aggregate them into a single overall value and produce a ranking. These weights reflect the relative importance of the policy objectives against which the performance of projects is measured. They need to be obtained from the decision makers since socially accepted sets of weights generally do not exist. In transport appraisal practice in many countries, such as UK or Switzerland, no weightings are applied and decision makers are deliberately left to make their own trade-offs between criteria. Reasons for this include that the aggregation of impacts into one measure is seen as inadequate because it implies a potential compensation of impacts and reduces transparency of the results and, secondly, that it is difficult to establish accepted weights (Walter et al., 2006). However, if as is often the case, project rankings for a large number of projects within a transport investment programme are to be developed, a formal method is required to make results between projects comparable and thus to support evidence-based policy implementation.

One possibility is to define cut-off or aggregations rules such as was applied in the appraisal procedure for the last federal transport plan for Germany. Here the core of the appraisal was a CBA and the results from complementary non-monetary environmental and spatial assessments could lead to an up- or down-grading of projects in the priority rankings (BMVBW, 2003). This required the definition of thresholds which in itself can be arbitrary and debatable. Moreover, results below the threshold do not have any impact on rankings, which might lead to projects being over- or under-rated.

An alternative approach is to convert MCA results into monetary values, e.g. based on a weighted summation of aggregate CBA and MCA results as in the decision support system for transport investments, COSIMA (Barfod et al., 2011), or partly monetising impacts for which valuation approaches exist, albeit highly uncertain (DfT, 2011b) for a sensitivity analysis. Monetary valuation, however, implies economic rationality in people’s trade-off decisions between impacts. This represents a simplified model of human behaviour which might lead to biased results (see e.g. Loomes, 2006) and requires careful exploration of the robustness to some of the unit values employed (see e.g. Holz-Rau and Scheiner, 2011). However, due to the perceived simplicity of the single monetary criterion this is to our knowledge rarely carried out in the political process of transport planning. In contrast, the explicit use of judgements in MCA encourages decision makers, when rankings are presented derived from an MCA model, to explore the rankings and understand the robustness of the proposals for project selection that the rankings suggest against variation in the trade-off weightings.

Hence, in this paper, we develop a weighted MCA approach to extract a transparent formulation of decision makers’ priorities, alongside methodologies for sensitivity testing and robustness analysis. However, this approach posed the challenge to develop a scoring system that transforms all impacts, including CBA results, on to a common preference scale.

The use of CBA data in the weighted MCA is unusual and, aside from giving a conceptual link between the CBA and the MCA, also facilitates the definition of a value for money threshold in determining the optimum size of the investment programme.

The novelty of the method to be described resides, therefore, in the way it facilitates a significantly stronger alignment with stated transport policy goals. This is especially so when ranking a large number of relatively small projects, where impacts outside those usually measured by conventional economic means are often strong if not dominant. It responds to the recognised limitations of CBA in this respect with an MCA-based CBA/MCA combination model which aims to capture the relative strengths of each perspective on evaluation but to do so in a way that explicitly recognises the importance to policy makers working in the transport sector of

- encouraging, in interchanges with stakeholder groups, an understanding of the robustness or otherwise of choices by readily allowing sensitivity testing,
- recognising that developments in appraisal procedure are implemented for policy makers in a situation where certain procedures and other policy pronouncements already exist and so are best achieved as evolution to existing processes rather than as radically new ones which may seem to abandon procedures previously represented as appropriate.

2.2. Outline of the NSR network case study

The multi-criteria framework in this case study has been developed for the appraisal of the National Secondary Road (NSR) network in Ireland. This consists of 2700 km of primarily rural single carriageway ‘A-roads’ that complement the national primary routes. The aim of the appraisal process was to identify

NSR routes or sections of a route suitable for investment to upgrade to a higher design standard and to develop a prioritised list of routes for investment. The appraisal methodology followed the Department of Transport's Common Appraisal Framework (DoT, 2009) and the National Road Authority's (NRA) Project Appraisal Guidelines (NRA, 2008).

The objectives for upgrading the NSR network had already been determined, based on national policy documents. At the top level these objectives are the improvement of environment, safety, economy, accessibility and integration. The top-level objectives have been broken down further into sub-criteria describing the main impacts contributing to the achievement of objectives. For the aggregation of results, a linear weighted summation MCA approach was applied, the advantages of which include reliability, robustness, effectiveness and lower complexity compared to other approaches (DCLG, 2009; Hajkowicz, 2007; Keeney and Raiffa, 1993; Tsamboulas, 2007). By this means, a direct and evidenced link is established between stated policy objectives and their implementation on the ground through the model to be developed.

For this approach, the project impacts (see Grady et al. (2011) for the impact modelling approach) are firstly transformed from their original units into numerical values (scores) on a preference scale. The appraisal results are presented as a one-page tabular summary for each option, based on the Project Appraisal Balance Sheet from the NRA Project Appraisal Guidelines. Weights are then applied to each criterion to combine the different impacts into a single 'score', using a linear additive model.

2.3. Weighting

Ideally the weight allocated to each criterion and sub-criterion in the framework should reflect the preferences of the decision-makers, in this case the board of the NRA—who act as agents of the government. The NRA board had set out relative weights for each of the five main criteria: Environment has a weight of 10%, safety 10%, economy 35%, accessibility 10% and integration 35%. No preferences were made available for the eighteen sub-criteria and it was not possible to interview the NRA board to determine any such weights. The weights for each of the sub-criteria were therefore defined by reviewing policy statements and similar sources and allocating relative importance weights consistently with that information.

2.4. Scoring

The scoring system in an MCA is fundamental to the integrity of the appraisal. The system used must be consistent between schemes; consistent between sub-criteria; objective; and not biased by scheme size. Furthermore, for transparency and ease of understanding, the scoring system is more straightforward if it is symmetrical (e.g., an accident reduction of 1 gives a positive score while an equivalent increase is scored equal in absolute terms but opposite in sign), and is linear. In Ireland the scoring of each of the sub-criteria also has to lie in the range 1–7 (NRA, 2008) in order to be consistent with other published road scheme appraisals. A score of 4 is defined as neutral, whilst 1 is defined as highly negative and 7 highly positive.

To meet these requirements and to make best possible use of the data gathered during the appraisal process different approaches were adopted for the monetised and non-monetised sub-criteria. Additionally the scoring process also differed between the quantitatively assessed and the qualitatively assessed non-monetised sub-criteria.

2.4.1. Monetised impacts

An unusual aspect of this research is the fusion of cost benefit analysis data in the MCA. This has been achieved by basing the scoring system for the sub-criteria included in a CBA (air, noise, accident reduction, transport efficiency and effectiveness and wider economic impacts) on the monetised value rather than the actual impact (e.g. 10 accidents saved). The underlying assumption is that where reliable estimates for the monetary value of impacts exist these are a good reflection of the societal preferences given to these impacts. The score allocated to each sub-criterion is given by the following formula:

$$\text{Score}_j = 4 + 3\left[\frac{(\text{PV}_j^{\text{Do Something}} - \text{PV}_j^{\text{Do Minimum}})/\text{PVC}}{\theta\alpha_j}\right]$$
$$\text{Scores} < 1.0 \text{ are rounded up to } 1.0$$
$$\text{Scores} > 7.0 \text{ are rounded down to } 7.0 \qquad (1)$$

where Score_j is the score for sub-criterion j (e.g. accident reduction), $\text{PV}_j^{\text{Do Something}}$ is the present value of sub-criterion j in the do something, $\text{PV}_j^{\text{Do Minimum}}$ is the present value of sub-criterion j in the do minimum, PVC is the present value of costs, θ is a benefit cost ratio threshold that is deemed highly positive (see later for details), α_j is the average proportion that sub-criterion j contributes to an average scheme PVB (present value of benefits).

The scoring function is therefore linear between scores 1.0 and 7.0, symmetrical around the neutral score (4.0) and truncated at 1.0 and 7.0. This is by design and for the reasons outlined earlier.

There are two important aspects to the above scoring function. The first is that the scheme benefit by sub-criterion ($\text{PV}_j^{\text{Do Something}} - \text{PV}_j^{\text{Do Minimum}}$) is normalised by the PVC. This transformation prevents the scores being biased in favour of large projects. For the national secondary road projects the correlation coefficient between user benefits and the cost of construction is almost 0.6—that is all other things being equal we find projects with larger capital costs have larger net benefits. By applying this transformation the correlation reduces to -0.06—thus preventing any size bias in the programme evaluation.

The second important aspect of the scoring function is the $\theta\alpha_j$ term. This threshold term defines the level of impact that is highly positive and is therefore allocated a score of 7.0. Effectively it is an average 'highly positive' ratio of benefits for sub-criterion j to PVC. It is calculated as the product of a highly positive benefit cost ratio (BCR) and an expected proportion that j would contribute to total benefits. In this application θ was set to 2.5—as 2.5 is viewed as a strong BCR in Ireland—whilst α_j was determined by analysing the contribution of each of the sub-criteria to total benefits.

The values for α_j and $\theta\alpha_j$ are set out in Table 1. For the rural national secondary road projects analysed (310 projects) values for α_j are given in column A of Table 1. As the national secondary roads projects reflect a certain type of project (upgrading low quality single carriageway roads in rural areas) these 'median averages' were adjusted to be more indicative of projects in general (column B). This means that, e.g., safety impacts for the national secondary roads have a higher share than for an average project and the threshold for a highly positive impact is slightly reduced. This gives threshold values for each of the monetised sub-criteria as in Column C.

2.4.2. Non-monetised elements

Each of the non-monetised elements is assessed against the maximum impact possible under that element (including a negative maximum impact). For the qualitatively assessed impacts (biodiversity, water resources and cultural heritage) the maximum (negative) impact (score of 1.0) is defined as any permanent impact on an internationally important site or a permanent impact on a large part of a nationally important site.

Table 1
Calculation of 'highly positive' thresholds for monetised impacts.

Sub-criteria	Median contribution of each element to PVB (α_j)		
	National secondary road projects (%) (A)	General (%) (B)	Threshold value for maximum score ($\theta\alpha_j$) (C)
Air and climate	−0.2	1	0.03
Noise	−0.5	1	0.03
Transport efficiency and effectiveness	70	80	2.00
Wider economic impacts	3	3	0.08
Accident reduction	28	15	0.38

Note: θ=2.5.

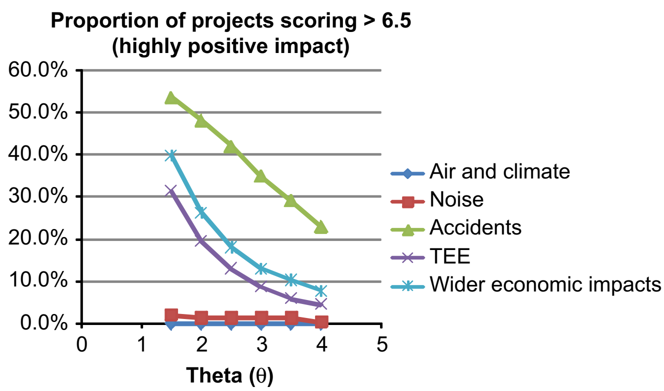


Fig. 1. Proportion of projects with highly positive scores as a function of theta (θ).

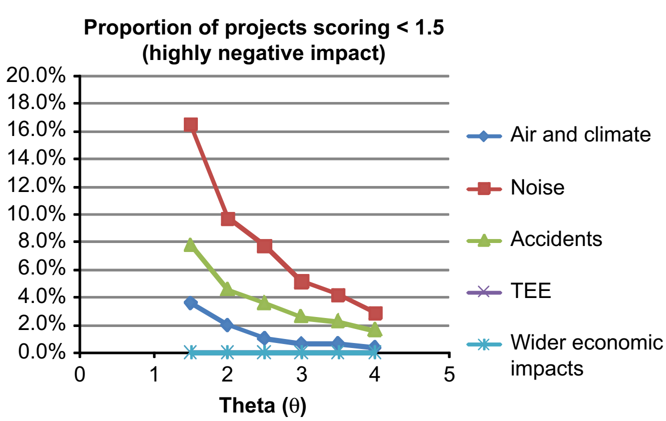


Fig. 2. Proportion of projects with highly negative scores as a function of theta (θ).

For the quantified impacts (accessibility, safety and integration) the maximum positive impact (score of 7.0) can be defined. Linear interpolation then gives a function for intermediate values between neutral (4.0) and highly positive (7.0).

2.4.3. Sensitivity of scores

Whilst θ has been determined by experience its value is relatively arbitrary. It is therefore worth exploring how sensitive the scores are to the value of θ. As is clear from Figs. 1 and 2, the score attributed to each monetised impact is very sensitive to θ. The higher is θ, the smaller the proportion of projects that score highly positive (score of 7) or highly negative (score of 1). The effect of increasing the value of θ therefore is to reduce the variance of the distribution of scores around the neutral mark (score = 4.0).

2.4.4. Compensatory impacts and red flags

With scores truncated at minimum and maximum bounds (1 and 7) no matter how adverse an impact is, with a weighted MCA it can always be compensated by positive impacts on other sub-criteria. The result could be a high overall MCA score and a recommendation that the project progresses, even though there may be an extremely unacceptable impact in terms of one of the sub-criteria (typically an environmental one).

This highlights one of the main weaknesses of using a weighted MCA in conjunction with a truncated scoring system. The approach adopted here for these circumstances was to identify any project that had a severe environmental impact with a red flag. For the project to be progressed through to completion it would be necessary to ensure the engineering design was able to mitigate any adverse environmental consequences of the project.

3. Defining the investment programme

3.1. Definition of an investment worthiness threshold

The strength of MCA is that it uses a set of consistent preferences from decision-makers to prioritise alternatives. Unlike CBA though it does not give a direct indication as to whether a project offers value. All the project's MCA score tells us is whether the project delivers benefits (score > 4.0) or dis-benefits (score < 4.0). In this application this is a weakness as there is a need to define the overall size of the investment programme. An investment worthiness threshold therefore needs to be identified.

We assume that the scores are all cardinal. That is, project scores do not just inform us whether one project is better than another, they inform us *how much* better it is. Thus a project with a score of 6.0 has twice the benefit (per Euro of investment) of a project with a score of 5.0 (a project that delivers no benefit has a score of 4.0). By making this assumption we can convert the PVC of a project into a 'score' thereby allowing us to define an investment worthiness threshold. This is done by making use of the relationship that a project that delivers a score of 7.0 with a given value of θ will have a BCR equal to θ (by definition). With a score of 7.0 project benefits equal 3 MCA 'points'. This then gives the following relationship for the PVC in MCA 'points':

$$\text{PVC score} = 4 + 3/\theta \tag{2}$$

For the national secondary road dataset where θ was set to 2.5 this meant that a worthiness threshold is 5.2 points. As can be seen from Table 2, with this threshold 31% of the 310 rural schemes appraised passed the worthiness test.

Table 2 also illustrates the sensitivity of both the scores and the number of projects that pass the worthiness test to θ. It arises because altering the value of θ alters the investment threshold value and the scores of the monetised sub-criteria. It does not, however, affect the scores of the non-monetised sub-criteria. As a consequence, as θ increases the rate of decrease in benefit scores (Column B) is slower than the rate of decrease in the PVC score (Column A). The result is that more projects achieve the value for money threshold for high levels of θ.

3.2. Capital budgeting

For many national secondary road corridors several mutually exclusive projects were appraised. When funds are limited, ranking on BCR is appropriate in a CBA for non-mutually exclusive projects, whilst an incremental analysis is appropriate for mutually exclusive projects (i.e. where several alternatives exist) (see e.g. Boardman and Greenberg, 2006). We have adapted these criteria to an MCA as follows.

Our MCA scores are already normalised by PVC to prevent size of project bias. Therefore for non-mutually exclusive projects the project that gives the best value for money is the one with the highest MCA score.

For mutually exclusive projects, to our knowledge an established incremental method for the application of MCA to transport investments does not exist. We therefore borrowed the incremental analysis method from cost benefit analysis, comparing the incremental cost to derive an MCA score for the 'notional' project that upgrades the lower cost project to the higher cost project. If this score is greater than a defined value for money threshold then the higher cost project enters the programme instead of the lower cost option. If it is lower, the lower cost project remains in the programme (providing it provides value for money). The incremental analysis MCA score is therefore calculated as follows:

$$\text{Incremental MCA Score} = ((MCA_B \times PVC_B) - (MCA_A \times PVC_A))/(PVC_B - PVC_A) \quad (3)$$

where project A is the lower cost project; project B is the higher cost project; MCA_A is the MCA score of project A and PVC_A is the PVC of project A.

The appraisal of the national secondary road network projects used an incremental MCA threshold of 5.5. This was based on an analysis of the MCA score that was necessary to appear in the top half of schemes that represented value for money (when $\theta = 2.5$). Applying this decision criterion to our dataset gives an investment programme of 65 rural projects with a combined capital value of €1558 million.

As can be seen from Figs. 3 and 4 the size and content of the investment programme is sensitive to the incremental MCA threshold used. Two effects are noticeable. Firstly, the higher the incremental threshold the greater the proportion of lower cost projects that enter the programme (e.g. Type 3s). Secondly the higher the threshold the smaller and cheaper the programme is. The size of the programme shrinks as the lower cost schemes (e.g. the Type 3s) are less likely to exceed the investment worthiness threshold (see Section 3.1). The total cost of the programme

Table 2
PVC score and number of projects offering value for money for different values of θ (core appraisal weights).

Threshold θ	PVC score (A)	Average score (B)	Projects offering value for money	
			Absolute	Proportion of projects appraised (%)
1.5	6.00	5.17	32	10
2.0	5.50	5.07	72	23
2.5	5.20	4.99	97	31
3.0	5.00	4.93	115	37
3.5	4.86	4.88	147	47
4.0	4.75	4.83	170	55

Note: Based on analysis of 310 rural NSR projects (some of which are mutually exclusive).

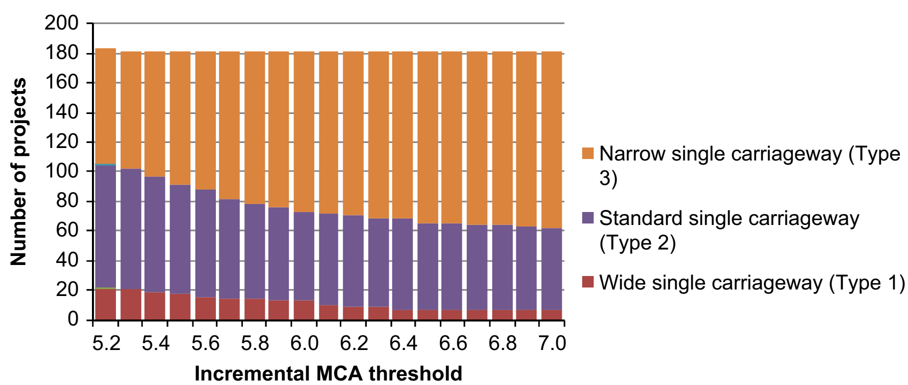


Fig. 3. Design standard of preferred projects by incremental MCA threshold (θ=2.5).

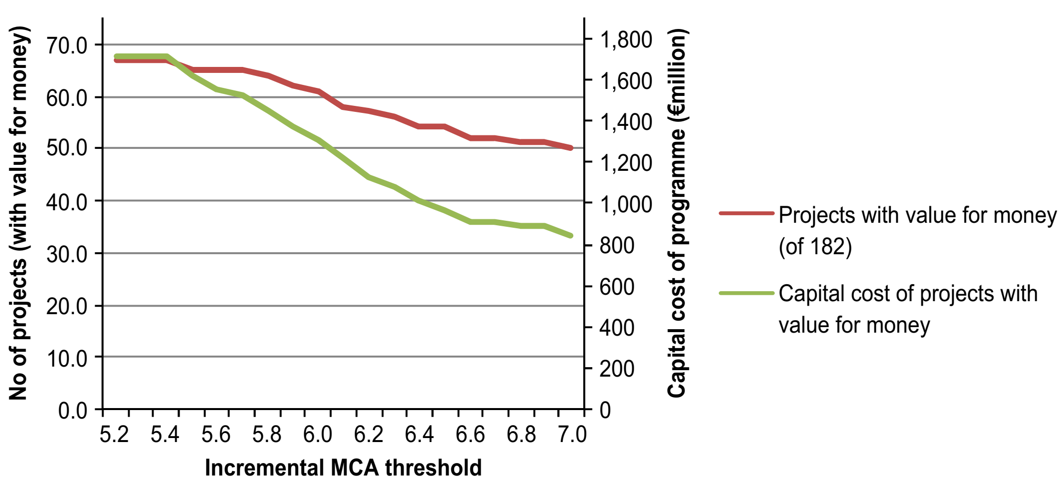


Fig. 4. Size of investment programme and capital cost by incremental MCA threshold (θ=2.5).

shrinks due to a combination of a shrinking programme and a shift towards a lower, cheaper design standard.

4. Sensitivity testing

Multi-criteria analysis is exposed to the same general sources of error as cost benefit analysis, relating to the uncertainty of data and models or the potential biased use of appraisal methods in the decision process (see, e.g., Mackie and Preston, 1998; Quinet, 2000). However, further variability is introduced through the scoring and weighting functions. The focus of our sensitivity tests is on the robustness of results to changes in weights—which are typically seen as being the less objective input of the two.

4.1. Sensitivity analysis of rankings to moderate changes in weights

If the overall rankings are relatively insensitive to changes, there can be greater confidence that they are a robust reflection of the relative importance given to the different criteria. Alternatively, a high level of sensitivity points towards the importance of being clear that the weights are indeed an accurate reflection of social preferences between the types of impact and to being confident that the assessments of impact for the relevant criteria have been accurately made.

Sensitivity testing initially focused on particular areas of the overall value tree:

- Sensitivity to changes in the weights between the five decision criteria.
- Sensitivity within the environment and integration criteria to changes in weights for the sub-criteria elements.

4.1.1. Sensitivity of ranking to changes in criteria weights

In the first instance, sensitivity was tested through exploring changes to the top 30 schemes identified by the original model (Fig. 5) as weights for the top-level criteria changed. The diagram confirms that the bulk of the contribution to their overall performance comes from the economy and integration impacts (the two most highly weighted) and that most schemes are recording broadly similar levels of impact on each criterion. This, combined with there being three further criteria involved, suggests there will not be high levels of sensitivity to change in the top-level weights.

To examine this further, we explored what happens to the ranking of schemes as each one of the five criteria weights is varied from its initial level either down to zero or up to 100%, with proportionate changes in the weight allocated to the other four criteria. For clarity, the illustration is restricted to the top 15 performing schemes in Fig. 6. The vertical line in the diagram corresponds to the current weight level (e.g., 0.10 for environment in Fig. 6a).

Most schemes' performance score sensitivity lines remain broadly parallel to each other. There are few cross-overs, especially for smaller weight changes, suggesting relative insensitivity of the ranking to weight changes. There are, however, a limited number of cases where a relatively modest change in weight would lead to some change in ranking and more extreme weight changes would induce others. Overall, however, the rankings appear quite robust to reasonably sized deviations from the initial weight estimates.

4.1.2. Sensitivity of ranking to changes in sub-criteria weights

Variation in the top-level weights has the potential to induce greater changes in ranking for any given absolute amount of

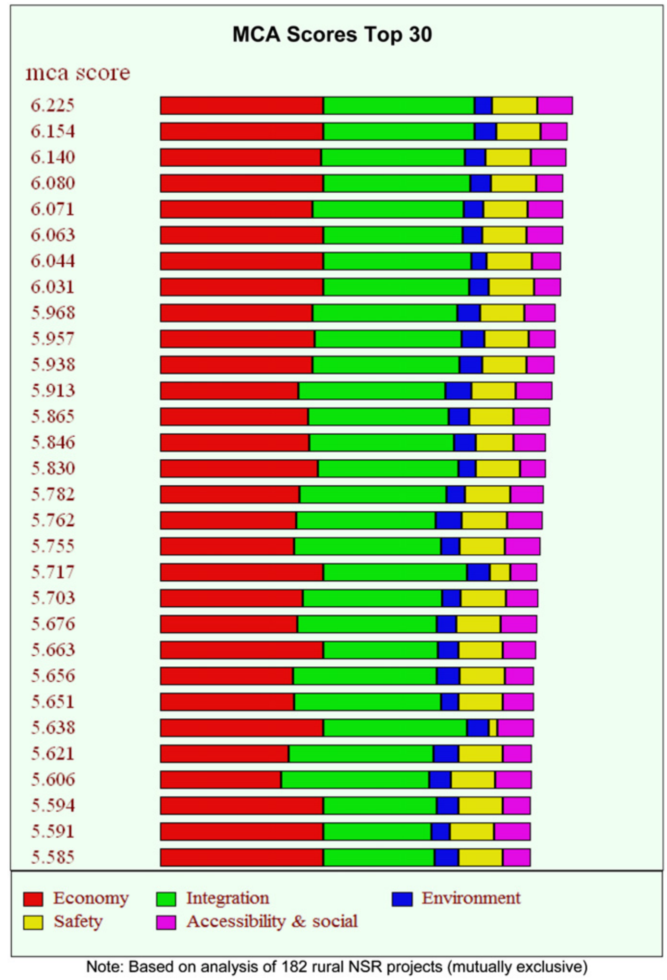


Fig. 5. Top 30 projects using initial weights.

weight variation. However, arguably, extreme amounts of variation are relatively unlikely, as most stakeholders may accept that each of the five criteria should have a degree of influence. This is less certain for lower level criteria, where the relevance of some of the individual sub-criteria may be contested. Although a given absolute amount of weight change will have less influence at the lower level, there might be a case for exploring more extreme levels of variation.

To explore possibilities of this type, we examined two cases. Within environment we assumed that all the weight was allocated to air quality (with the other six sub-criteria ignored) and within integration we separately assumed that all the weight was allocated to the geographical criterion. The results of these two experiments are shown in Tables 3 and 4. In the case of environment, judging by ranking changes in the top 30 schemes, there were only limited consequences. This is in part due to the fact that the top level criterion in this case has a relatively low weight. For integration, the potential overall ranking changes are greater; integration is one of the more heavily weighted top-level criteria. Other influences are that scores on integration are not strongly correlated with scores on the other four top-level criteria and that within the four sub-criteria that contribute to the Integration weighted score, there are few high correlations between impact levels.

It should be noted that what is being considered are very substantial changes in the sub-criteria weights. Generally, as reinforced by the intermediate weight change column in Table 4, the degree of sensitivity is quite slight.

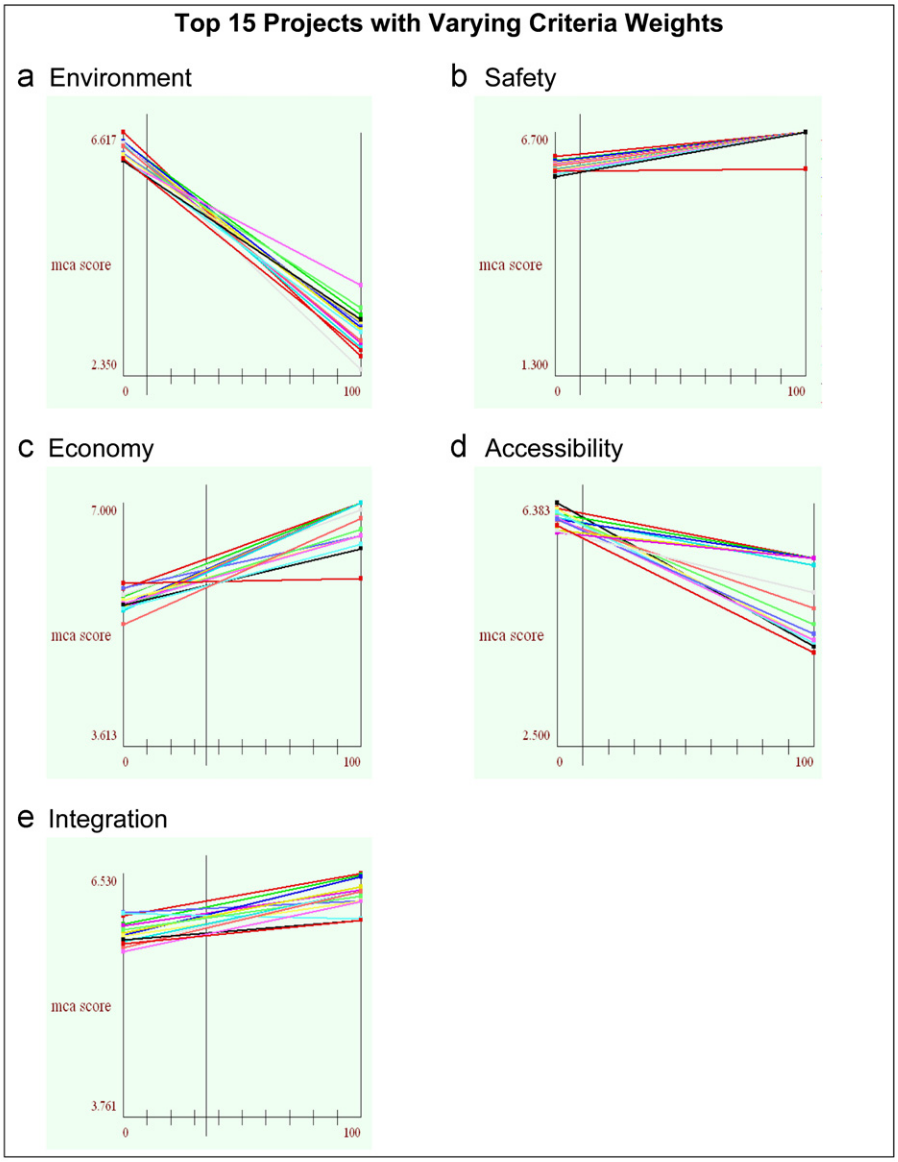


Fig. 6. Top 15 projects with varying criteria weights.

4.2. Robustness of results under high uncertainty on weights (Monte-Carlo simulation)

In the previous section we tested the impact of weight changes on the ranking of projects under the assumption that the structure of weight sets stayed the same as the original one. However, quite often not only will the exact values of weights be unknown but there might even be disagreement between different stakeholders about priority rankings between objectives. The question then is how stable would be the outcome of the MCA in terms of top schemes, size and structure of the investment programme under a significant change of weights. If neither size nor programme structure changes significantly even when assuming a completely random distribution of weights, there can be greater confidence in the MCA results for decision makers. To test this stability, a Monte Carlo simulation on random sets of weights was carried out. This analysis assumes complete uncertainty about the preferences of decision-makers. To this end, 5000 random weight vectors for the five main criteria were created following the method described in Tervonen and Lahdelma (2007), and the MCA score was calculated for all 182 rural schemes that remained after the removal of mutually exclusive projects.

4.2.1. Top schemes based on random weights

Table 5 gives the results from this analysis, showing the top 30 rural schemes with the highest probability of being ranked in the top 15 and in the top 30 schemes respectively, assuming random weights.

It shows that the top 10 projects from the original weighting rank among the most acceptable projects with a high probability of being among the best independently of the weights applied. This confirms the results given in Section 4.1 and underlines the stability of the programme ranking in terms of the inclusion of the top projects. However, there are a few projects that move up in priority (e.g., the now second-ranked R19b.1.T1 D, which moves from 12th, Table 4). If budgets were constrained to the top ten a more detailed comparison of these schemes with the original set would be advisable.

Table 3
Change in ranking of the top 30 projects if all the environment assessment is based on the air quality sub-criterion.

Alternative	Original rank	Rank if AQ&C weight=0.1
R10c.1.1.T1	1	2
R2g.1.T1	2	1
R19c.1.T1	3	3
R27a.1.T1	4	5
R2j.1.T2	5	6
R18b.1.T2	6	4
R2e.1.T1	7	9
R2f.1.T1	8	7
R19e.2.T2	9	8
R27b.2.T1	10	10
R27b.1.T1	11	11
R19b.1.T1 D	12	13
R32a.1.T2	13	14
R6f.1.2.T1	14	17
R18a.2.T2	15	12
R8e.3.T1	16	15
R16a.2.T2	17	20
R8e.2.T2	18	16
R19f.2.T2	19	18
R18a.1.T1	20	19
R32a.2.T2	21	27
R25a.1.T1	22	21
R19e.1.T2	23	23
R8e.4.T1	24	22
R6a.1.T2	25	25
R6f.2.T1	26	30
R19b.2.T2 D	27	29
R25a.2.T2	28	26
R20c.4.T2	29	31
R5c.2.T3	30	24

Note: Route names made anonymous

Table 4
Change in ranking of the top 30 projects if all the integration assessment is based on the geographical sub-criterion.

Alternative	Original rank	Intermediate case (geog weight=0.1925)	Rank if geog weight=0.35
R10c.1.1.T1	1	4	9
R2g.1.T1	2	3	8
R19c.1.T1	3	5	13
R27a.1.T1	4	8	17
R2j.1.T2	5	1	2
R18b.1.T2	6	6	14
R2e.1.T1	7	10	18
R2f.1.T1	8	11	19
R19e.2.T2	9	13	21
R27b.2.T1	10	14	22
R27b.1.T1	11	15	24
R19b.1.T1 D	12	16	28
R32a.1.T2	13	17	27
R6f.1.2.T1	14	20	32
R18a.2.T2	15	19	30
R8e.3.T1	16	24	33
R16a.2.T2	17	22	31
R8e.2.T2	18	25	36
R19f.2.T2	19	28	40
R18a.1.T1	20	27	37
R32a.2.T2	21	29	39
R25a.1.T1	22	9	5
R19e.1.T2	23	32	44
R8e.4.T1	24	33	46
R6a.1.T2	25	36	47
R6f.2.T1	26	39	53
R19b.2.T2 D	27	40	56
R25a.2.T2	28	12	4
R20c.4.T2	29	18	11
R5c.2.T3	30	2	1

Note: Route names made anonymous.

Table 5
Top 30 rural schemes by probability index based on random weights.

Rank	Scheme	Probability top 15 (%)	Original rank	Scheme	Probability top 30 (%)	Original rank
1	R19c.1.T1	97	3	R19c.1.T1	100	3
2	R19b.1.T1 D	87	12	R19b.1.T1 D	99	12
3	R10c.1.1.T1	87	1	R19e.2.T2	98	9
4	R2j.1.T2	84	6	R32a.1.T2	98	13
5	R18b.1.T2	81	5	R2j.1.T2	98	6
6	R19e.2.T2	78	9	R16a.2.T2	97	17
7	R16a.2.T2	73	17	R10c.1.1.T1	95	1
8	R2g.1.T1	71	2	R18b.1.T2	94	5
9	R32a.1.T2	68	13	R2g.1.T1	88	2
10	R27a.1.T1	55	4	R27b.1.T1	86	11
11	R19b.2.T2 D	50	26	R27a.1.T1	84	4
12	R8e.2.T2	47	18	R8e.2.T2	82	18
13	R27b.1.T1	43	11	R8e.3.T1	80	16
14	R20c.4.T2	41	29	R27b.2.T1	79	10
15	R2f.1.T1	38	8	R19b.2.T2 D	77	26
16	R27b.2.T1	37	10	R32a.2.T2	77	21
17	R2e.1.T1	37	7	R6f.1.2.T1	73	14
18	R32a.2.T2	30	21	R18a.1.T1	72	20
19	R25a.1.T1	30	22	R2f.1.T1	72	8
20	R8e.3.T1	28	16	R2e.1.T1	67	7
21	R9a.1.T2	27	47	R20c.4.T2	66	29
22	R6f.2.T1	27	27	R25a.1.T1	65	22
23	R20e.1.T2	24	49	R6f.2.T1	64	27
24	R5c.2.T3	21	30	R19e.1.T2	64	25
25	R6f.1.2.T1	20	14	R9a.1.T2	60	47
26	R6a.1.T2	19	24	R19d.1.T2	57	38
27	R19d.1.T2	17	38	R20e.1.T2	56	49
28	R27f.1.T2	16	45	R5c.2.T3	47	30
29	R24a.3.T2	14	34	R24a.3.T2	46	34
30	R19e.1.T2	14	25	R18a.2.T2	45	15

Note: Route names made anonymous.

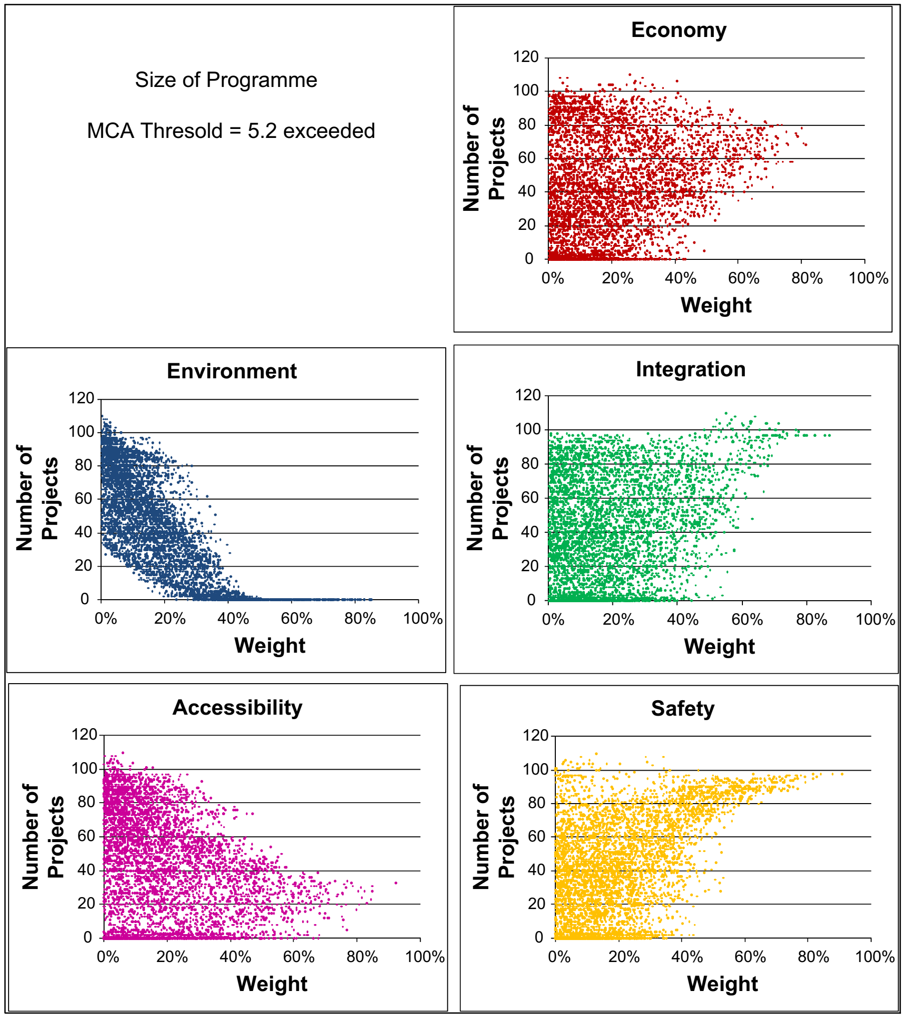
4.2.2. Impact of weight changes on the size of the programme

A policy relevant question is how many projects would be undertaken and what investment volume would result if the weights for the five criteria were changed? As described earlier, in the NSR network study a threshold score of 5.2 was applied as a cut-off value for deciding whether a project is worthwhile or not (see Table 2). With the original weighting scheme 65 (=36%) of the 182 rural schemes pass this cut-off score. Fig. 7 illustrates how changing weights for the five criteria would change the number of projects that are deemed worthwhile.

The figure shows that economy, integration and safety influence the size of the programme significantly only if weights exceed 40%. For accessibility and in particular environment, however, with a change from their current weights of 10% by value, the impact on the size of the programme can be significant (Table 6). The average number of projects passing the threshold over all weight sets is 42, showing that the relatively low weight for environment in the original weight set leads to a considerable increase in the size of the programme.

These results are largely to be expected given that most projects show positive results for safety and integration and negative results on environment. The impact of accessibility weights can be explained by the fact that a high accessibility score will be achieved by relatively few projects—those that have a high importance for the connectivity closer to agglomerations. In contrast, a high score for integration favours projects that have positive impacts for more remote regions.

One of the reasons for the very low environment scores is the early stage of planning of the projects. This means mitigation measures have not yet been taken into account and even avoidable local conflicts could lead to a highly negative impact score. Thus the current set of scores may give an overall too pessimistic



Note: Based on analysis of 182 rural NSR projects (non-mutually exclusive)

Fig. 7. Number of rural projects exceeding threshold depending on criteria weights.

indication of programme size if a high, but not unrealistic, weight is allocated to environment. This could be avoided if the scoring system for environment allowed for a more differentiated valuation between impacts that could potentially be mitigated or those that cannot.

5. Conclusions

Applying a multi-criteria method for the appraisal of the NSR network has facilitated the inclusion of political objectives in the assessment that cannot be monetised with sufficient confidence, such as integration of remote regions or environmental quality criteria. In this way a closer alignment is achieved between transport policy and the tools used to support the application of that policy. While a similar alignment is desirable across the full spectrum of public expenditure, transport involves particularly large levels of infrastructure investment with a disparate and dispersed set of impacts. The sector has posed, and continues to pose, substantial issues for politicians and others who must justify how public money is spent. The procedure developed and illustrated here allows a clearer evidencing that decision makers on the ground have followed through on stated policy objectives. This is of particular importance in the common situation where transport infrastructure funding decisions are distributed between different levels of policy making, e.g. when local authorities compete for central funding and demand a clear justification of prioritisation of funding decisions.

A specific innovation of this study was to develop a methodology that incorporates CBA elements into the MCA framework in a consistent and transparent manner. This has been accomplished by transforming the monetary results from the partial CBA into MCA scores through the definition of a benefit-cost ratio threshold for projects that are considered highly positive. The resulting programme of worthwhile projects has been further condensed by applying an incremental analysis to decide between mutually exclusive projects. This novel approach provides a conceptual link between CBA and MCA and facilitates the definition of a value for money threshold in determining the size of the investment programme. This latter is of importance to the transport sector, as to other major sectors of the economy, all of which typically are competing for their share of a limited public funding pot and wish to demonstrate that expenditure in their sector provides appropriate relative and absolute returns.

The process that transforms monetised impacts into scores relies on the definition of a high value BCR. This threshold has a strong influence on the scores of the monetised impacts and the number of projects that pass the value for money test. This presents the same conceptual problem as the definition of a BCR threshold based on CBA results alone. Often such BCR thresholds are derived from experience with what is realistic under budgetary constraints (as well as political considerations) rather than from methodological ones and factors

Table 6
Size of programme depending on weights for environment and accessibility.

Weight (%)	Average number of projects passing MCA threshold	
	Environment	Accessibility
0–5	72	52
5–10	63	49
10–15	54	46
15–20	45	42
20–25	34	41
25–30	23	36
30–35	14	32
35–40	6	31

outside the CBA have a strong influence (Eliasson and Lundberg, 2012). Further research across the transport sector would be required to establish a corresponding value for application in MCA frameworks and to establish evidence what constitutes a high BCR, ideally including ex-post evaluations of projects. Similarly, established scoring systems for other, e.g. environmental impacts, do not exist, which leads to a higher uncertainty and potentially too precautionary downgrading of projects in the process.

From the point of view of the transport policy decision maker it is important to have an understanding of the robustness of the size of, and rankings within, investment programmes created. For this reason, a variety of sensitivity tests has been devised to explore the extent to which the initial rankings may be affected by changes in view about the relative importance of different impacts. Based on these results, there is no strong evidence of very high sensitivities in the rankings. While the size of the programme is sensitive to changes in weights, schemes that come towards the top of the ranking are good performers in terms of the criteria and the relative importance of those criteria that have been assessed in collaboration with the decision makers, and reasonably sized variations in the data do not induce drastic changes in the rankings obtained. Even under more radical changes in the structure of weight sets, the best performing schemes stay towards the top of the rankings. One reason for this is that we are dealing with a relatively homogenous set of investment projects. Further research is necessary to analyse whether such a methodology works even if very different types of project were to be appraised, or if an upfront division of projects into different budgets, as suggested for the transport sector in OECD/ITF (2011), is more advisable.

Acknowledgements

This research has been possible as part of the National Secondary Road Needs Study, funded by the NRA in Ireland. We would also like to thank our project partners RPS, PMS and Fitzpatrick Associates for their input appraisal results and constructive discussions throughout the process.

References

Barfod, M.B., Salling, K.B., Leleur, S., 2011. Composite decision support by combining cost-benefit and multi-criteria decision analysis. Decision Support Systems 51 (2011), 167–175.
BMVBW (Bundesministerium für Verkehr, Bauen und Wohnen), 2003. Federal Infrastructure Plan 2003. Federal Ministry of Transport, Building and Housing, Berlin.
Boardman, A.E., Greenberg, D.H., Vining, A.R., Weimer, D.L., 2006. Cost-benefit Analysis: Concepts and Practice. Prentice Hall, Upper Saddle River, N.J.
Bristow, A.L., Nellthorp, J., 2000. Transport project appraisal in the European Union. Transport Policy 7 (1), 51–60.
DCLG (Department for Communities and Local Government), 2009. Multi-Criteria Analysis—A Manual. Communities and Local Government Publications, London.
DfT (Department for Transport), 2011a. The Appraisal Process. TAG Unit 2.5. Transport Analysis Guidance (TAG), April 2011. 〈www.dft.gov.uk/webtag〉.
DfT (Department for Transport), 2011b. Value for Money Assessments—Guidance. Published 4/5/2011. 〈http://www2.dft.gov.uk/about/howthedftworks/vfm/vfmassessments/index.html〉.
DoT (Department of Transport), 2009. Guidelines on a Common Appraisal Framework for Transport Projects and Programmes, June 2009, Dublin.
Eliasson, J., Lundberg, M., 2012. Do cost-benefit analyses influence transport investment decisions? Experiences from the Swedish transport investment plan 2010–2021. Transport Reviews 32 (1), 29–48.
Grady, A., Harris, R., Shortt, R., Noonan, M., Crosbie, N., Laird, J., Gaughran, A., 2011. National Secondary Roads Needs Study—Network Options Report. 〈http://www.nra.ie/NetworkManagement/NationalSecondaryRoadsNeedsStudy/〉.
Grant-Muller, S.M., Mackie, P., Nellthorp., J., Pearman, A., 2001. Economic appraisal of European transport projects: the state-of-art revisited. Transport Reviews 21 (2), 237–261.
Hajkowicz, S., 2007. A comparison of multiple criteria analysis and unaided approaches to environmental decision making. Environmental Science & Policy 10 (3), 177–184.
Holz-Rau, C., Scheiner, J., 2011. Safety and travel time in cost-benefit analysis: a sensitivity analysis for North Rhine-Westphalia. Transport Policy 18 (2), 336–346.
Keeney, R.L., Raiffa, H., 1993. Decisions with Multiple Objectives: Preferences and Value Tradeoffs. Cambridge University Press.
Lautso, K., Wegener, M., Spiekermann, K., Shepperd, I., Steadman, P., Martino, A., Domingo, R., Gayda, S., 2004. Planning and Research of Policy for Land Use and Transport for Increasing Urban Sustainability (PROPOLIS). Final Report to the European Commission, Version 2.0. 〈http://www.ltcon.fi/Propolis/〉.
Leleur, S., 2012. Complex Strategic Choices: Applying Systematic Planning for Strategic Decision Making. Springer Series in Decision Engineering.
Loomes, G., 2006. (How) can we value health, safety and the environment? Journal of Economic Psychology 27 (6), 713–736.
Macharis, C., De Witte, A., Turcksin, L., 2010. The multi-actor multi-criteria analysis (MAMCA) application in the Flemish long-term decision making process on mobility and logistics. Transport Policy 17 (2010), 303–311.
Mackie, P., Preston, J., 1998. Twenty-one sources of error and bias in transport project appraisal. Transport Policy 5 (1998), 1–7.
Manheim, M.L., Suhrbier, J.H., Bennett, E.D., Neumann, L.A., Colcord Jr., F.C., Reno Jr., A.T., 1975. Transportation Decision-making: A Guide to Social and Environmental Considerations. National Cooperative Highway Research Program Report 156. Transportation Research Board, National Research Council, Washington, D.C.
Nijkamp, P., Rietveld, P., Voogd, H., 1990. Multicriteria Evaluation in Physical Planning. Elsevier Science, Amsterdam.
NRA, 2008. National Roads Authority Project Appraisal Guidelines. Report Dated March 2008, Dublin.
Odeck, J., 2010. What determines decision-makers preferences for road investments? evidence from the Norwegian road sector. Transport Reviews 30 (4), 473–494.
Odgaard, T., Kelly, C., Laird, J., 2005. HEATCO Work Package 3: Current Practice in Project Appraisal in Europe, Deliverable 1/Volume 1 (Main Text). 〈http://heatco.ier.uni-stuttgart.de/〉.
OECD/ITF, 2011. Improving the Practice of Transport Project Appraisal, Summary of Discussions. ITF Round Tables, No. 149. OECD Publishing.
Quinet, E., 2000. Evaluation methodologies of transportation projects in France. Transport Policy 7 (2000), 27–34.
Quinet, E., 2011. The Practice of Cost-Benefit Analysis in Transport: The Case of France. OECD/ITF 'Improving the Practice of Transport Project Appraisal, Summary of Discussions'. ITF Round Tables, No. 149. OECD Publishing, pp. 25–54.
Tervonen, T., Lahdelma, R., 2007. Implementing stochastic multicriteria acceptability analysis. European Journal of Operational Research 178 (2007), 500–513.
Thomopoulos, N., Grant-Muller, S., Tight, M.R., 2009. Incorporating equity considerations in transport infrastructure evaluation: current practice and a proposed methodology. Evaluation and Program Planning 32 (4), 351–359.
Tsamboulas, D.A., 2007. A tool for prioritizing multinational transport infrastructure investments. Transport Policy 14 (1), 11–26.
van Wee, B., 2011. How suitable is CBA for the ex-ante evaluation of transport projects and policies? A discussion from the perspective of ethics. Transport Policy 19 (1), 1–7.
Walter, F., Lieb, S., Simmen, H., 2006. Nachhaltigkeitsbeurteilung des Bundes bei Strassenprojekten: NISTRA. Umweltrecht in der Praxis (URP), Vereinigung für Umweltrecht (VUR), vol. 20 (5), pp. 571–586.